Centre for
Economic
and Financial
Research
at
New Economic
School

September 2008

Specification Testing in Models with Many Instruments

Stanislav Anatolyev
Nikolay Gospodinov

Working Paper No 124

CEFIR / NES Working Paper series

Specification Testing in Models with Many Instruments*

Stanislav Anatolyev† — New Economic School
Nikolay Gospodinov‡ — Concordia University and CIREQ

September 2008

Abstract

This paper studies the asymptotic validity of the Anderson–Rubin (AR) test and the J test of overidentifying restrictions in linear models with many instruments. When the number of instruments increases at the same rate as the sample size, we establish that the conventional AR and J tests are asymptotically incorrect. Some versions of these tests, that are developed for situations with moderately many instruments, are also shown to be asymptotically invalid in this framework. We propose modifications of the AR and J tests that deliver asymptotically correct sizes. Importantly, the corrected tests are robust to the numerosity of the moment conditions in the sense that they are valid for both few and many instruments. The simulation results illustrate the excellent properties of the proposed tests.

Keywords: Instrumental variables, many instruments, Bekker's asymptotics, Anderson–Rubin test, test for overidentifying restrictions.

JEL codes: C12, C21.

*We would like to thank the co-editor Richard Smith, two referees and Bruce Hansen for very useful comments and suggestions. The second author gratefully acknowledges financial support from FQRSC and SSHRC.

†Address: New Economic School, Nakhimovsky Prospekt, 47, room 1721, Moscow, 117418, Russia. Web: sanatoly@nes.ru; http://www.nes.ru/~sanatoly/

‡Address: Nikolay Gospodinov, Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada. Web: nikolay.gospodinov@concordia.ca; http://alcor.concordia.ca/~gospodin/

1 Introduction

In the pursuit of improved precision of the instrumental variable (IV) estimator, researchers often face situations in which the number of instruments represents a nontrivial fraction of the sample observations available for estimation. For example, a large number of instruments can be constructed by interacting different variables (Angrist and Krueger, 1991) or using lagged dependent variables in panel data models (Arellano and Bond, 1991). While the conventional asymptotic setup implies that the increased dimensionality of the instrument matrix should lead to efficiency gains, the finite-sample behavior of the IV estimator and various test statistics is markedly deteriorated (Andersen and Sorensen, 1996; Burnside and Eichenbaum, 1996; among others).

Despite the voluminous recent literature on estimation in the presence of many (and possibly weak) instruments (Bekker, 1994; Chao and Swanson, 2005; Hansen, Hausman and Newey, 2006; among others), the asymptotic behavior of the tests for parameter and overidentifying restrictions has not been fully investigated. Andrews and Stock (2007) and Donald, Imbens and Newey (2003) derive the asymptotic distributions of some parameter and specification tests in models with *moderately many instruments*, i.e. when the number of instruments grows asymptotically but slowly relative to the sample size. We argue that in order to obtain a good asymptotic approximation for some of these tests one has to acknowledge the numerosity of instruments via a *many instruments* assumption of Bekker (1994).

It turns out that when the number of moment conditions is proportional to the sample size, the conventional J test for overidentifying restrictions tends to underreject and the size of the test is practically zero when the ratio of the number of moment conditions to sample size is close to one. Interestingly, despite its similar structure, the asymptotic size of the standard Anderson–Rubin (AR) test exceeds the nominal level when there are many instruments. Thus, the AR test tends to overreject and the size of the test is near 50% when the ratio is close to one. Similar conclusions apply to the asymptotically normal J and AR tests developed in Donald, Imbens and Newey (2003) and Andrews and Stock (2007). Intuitively, the asymptotic size distortions arise from the fact that there is a finite number of observations per moment condition, in contrast to their infinite number in the standard and moderately many instruments frameworks.

We propose modifications of the conventional J and AR tests that are based on critical values of a chi-squared distribution and are easy to implement. Importantly, the proposed "corrected" tests are robust to the numerosity of the moment conditions, in the sense that they do not require an *a priori* choice of asymptotic framework because they are valid under both fixed and many instrument asymptotics.

The rest of the paper is structured as follows. Section 2 introduces the model and the tests. The main theoretical results are established and discussed in Section 3. Section 4 presents Monte Carlo simulation results for the size properties of tests under consideration in finite samples. Section 5 concludes. All proofs are relegated to the Appendix.

2 Model, Assumptions and Tests

Consider the standard linear IV regression model

$$y_i = x_i'\beta + e_i, \quad E[x_i e_i] \neq 0,$$

where $\{y_i, x_i, z_i\}_{i=1}^n$ is a random sample and z_i denotes a vector of valid instruments.

The model can be written in matrix form as

$$y = X\beta + e, \tag{1}$$

where $y = (y_1, \cdots, y_n)'$ is $n \times 1$, $X = (x_1, \cdots, x_n)'$ is $n \times k$, $Z = (z_1, \cdots, z_n)'$ is $n \times \ell$, $e = (e_1, \cdots, e_n)'$ is $n \times 1$. In this paper, we consider the case when the dimension of β is small relative to n, but ℓ is large and comparable to n, although constrained to be smaller than n.

The model and the data are assumed to satisfy the following conditions.

Assumption 1 *The errors e_i satisfy $E[e|Z] = 0$, $E[ee'|Z] = \sigma^2 I_n$ and $E\left[|e_i|^4\right] < \infty$.*

Assumption 2 *As $n \to \infty$, $\ell/n = \lambda$, where $0 < \lambda < 1$.*

Assumption 1 imposes homoskedasticity and a finite fourth moment of the errors. Assumption 2 adopts the many instruments asymptotic framework of Bekker (1994) when the number of instruments is a nontrivial fraction of the sample size (see also Newey, 2004). If the number of instruments is fixed (conventional framework) or grows more slowly than the sample size (moderately many instruments framework), the noise that arises from the large dimensionality of Z vanishes in the limit which validates the use of conventional asymptotics for inference (Koenker and Machado, 1999). The advantage of the parameterization in Assumption 2 is that it explicitly recognizes the presence of this source of uncertainty and eventually leads to a better approximation to the exact distribution of the statistic of interest.

For convenience, the vector of instruments z_i will be treated as nonrandom.

Assumption 3 *Under the asymptotics of Assumption 2, $\max_{1\leq i\leq n} |z_i'(Z'Z)^{-1}z_i - \lambda| \to 0$.*

Assumption 3 requires that all diagonal elements of the projection matrix $P = Z(Z'Z)^{-1}Z'$ converge to λ (recall that under the standard or moderately many instruments asymptotics they converge to zero). When the instruments are generated in the random sampling framework or under stationarity, the expected value of $z_i'(Z'Z)^{-1}z_i$ is equal to λ. Indeed,

$$E\left[z_i'(Z'Z)^{-1}z_i\right] = \frac{1}{n}E\left[\operatorname{tr}\left((Z'Z)^{-1}\sum_i z_i z_i'\right)\right] = \frac{1}{n}E\left[\operatorname{tr}\left(I_\ell\right)\right] = \lambda,$$

where the first equality follows by symmetry over observations and properties of trace. In addition, Assumption 3 requires that the variance of each $z_i'(Z'Z)^{-1}z_i$ is zero, which is to be expected because the dimensionality of z_i linearly grows. The validity of Assumption 3 follows from the literature on large dimensional covariance matrices (Silverstein, 1995) in case the elements of Z are IID both across rows and columns, possibly after a rotating transformation, and have finite fourth moments (which, in particular, includes the case of normality of z_i). The IID requirement for the elements in z_i can be relaxed at the expense of existence of higher order moments (Ledoit and Wolf, 2004). Moreover, a limited amount of endogeneity is allowed; for example, lagged elements of x_i or y_i may be present among elements of z_i as long as they occupy only an asymptotically finite number of columns of Z.

Let $\hat{\beta}$ be an estimator of β. Later we will impose restrictions on the asymptotic behavior of $\hat{\beta}$. Also, let

$$\hat{e} = y - X\hat{\beta} \tag{2}$$

denote the vector of residuals and

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-k} \tag{3}$$

be the residual variance. Under Assumption 1, the standard J test for overidentifying restrictions is given by

$$J = \frac{\hat{e}'P\hat{e}}{\hat{\sigma}^2}, \tag{4}$$

and, under the null of correct moment restrictions $H_0 : E\left[e_i z_i\right] = 0$, is distributed as $\chi^2(\ell - k)$ in the conventional framework of fixed ℓ asymptotics. Alternatively, in the framework of moderately many instruments (more precisely, when $\ell^2/n \to 0$ as $\ell, n \to \infty$), Donald, Imbens, and Newey (2003) base their (right-sided) test on

$$J_{DIN} = \frac{J - \ell}{\sqrt{2\ell}} \xrightarrow{d} N\left(0, 1\right).$$

To construct the J statistic, a consistent estimator $\hat{\beta}$ is needed. It turns out that the choice of $\hat{\beta}$ is not important for the asymptotic behavior of J as long as the following conditions hold. Let Υ and V denote the matrices of observations and disturbances of the reduced form $X = \Upsilon + V$ with $E\left[V\right] = 0$.

Assumption 4 *(a) The estimator* $\hat{\beta}$ *satisfies* $\sqrt{n}(\hat{\beta}-\beta) = O_p(1)$*, (b)* $\Upsilon'\Upsilon/n \to Q$*, where* Q *is a positive definite matrix, (c)* $E\left[|v_{ij}|^4\right] < \infty$*.*

In our numerical work, we use the LIML estimator

$$\hat{\beta}_{LIML} = \left(X'(I_n - kM)X\right)^{-1} X'(I_n - kM)y,$$

where k is the smallest characteristic root of $(\overline{Y}'\overline{Y})(\overline{Y}'M\overline{Y})^{-1}$, $\overline{Y} = (y, X)$ and $M = I_n - P$. Note that part (a) of Assumption 4 permits the use of asymptotically inefficient and even non-normal estimators, as long as their rate of convergence is not slower than $\sqrt{n}$. Also, while part (b) of Assumption 4 rules out lack of identification ($\Upsilon = 0$), it allows for possibly weak instruments as the addition of new instruments does not provide additional information (Newey, 2004).

A popular test for $H_0 : \beta = \beta_0$ and the validity of the overidentifying restrictions is based on the Anderson–Rubin (AR) statistic

$$AR = (T - \ell)\frac{e_0'Pe_0}{e_0'Me_0}, \tag{5}$$

where $e_0 = y - X\beta_0$ is a vector of restricted errors. The Anderson–Rubin statistic possesses some appealing robustness properties, e.g. robustness to weak instruments, and is $\chi^2(\ell)$ distributed under fixed ℓ asymptotics. Alternatively, in the framework of moderately many instruments (more precisely, when $\ell^3/n \to 0$ as $\ell, n \to \infty$), Andrews and Stock (2007) show that

$$AR_{AS} = \sqrt{\ell}\left(\frac{AR}{\ell} - 1\right) \xrightarrow{d} N(0, 2).$$

3 Asymptotic Results

We first investigate the behavior of the conventional J and AR tests when one neglects the presence of many instruments, and carries out testing in the standard way, i.e. rejects when $J > q_\alpha^{\chi^2(\ell-k)}$ and $AR > q_\alpha^{\chi^2(\ell)}$. The following theorem describes the size of the conventional J and AR tests, along with the J_{DIN} and AR_{AS} tests, when the number of instruments grows at the same rate as the sample size.

Let $\Phi(x)$ be the standard normal cumulative distribution function, $\Phi^{-1}(x)$ be its quantile function, and $\alpha < 0.5$ be the target test size.

Theorem 1

(a) Suppose assumptions 1–4 hold. Then, the asymptotic size of the conventional and Donald, Imbens, and Newey (2003) J tests equals

$$\Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1-\lambda}}\right).$$

(b) *Suppose assumptions 1–3 hold. Then, the asymptotic size of the conventional and Andrews and Stock (2007) AR tests equals*

$$\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right).$$

Theorem 1 establishes that, under Bekker's asymptotics, the asymptotic size of the conventional J test is smaller than α and the asymptotic size of the conventional AR test exceeds α for all $\lambda > 0$. Consequently, the J test will underreject and the AR test will overreject in large samples. The same applies to the J_{DIN} and AR_{AS} tests. It turns out that the moderately many instruments framework cannot fully acknowledge the presence of many instruments, while Bekker's asymptotics can.

To visualize the effect of λ on the asymptotic behavior of the tests, Figure 1 plots the asymptotic p-value function of the J test at 1%, 5% and 10% nominal level which is identical to the asymptotic size of the AR test. Figure 1 shows that the over- (under-) rejection rates of the AR (J) test are not very large for $\lambda \leq 0.5$ but increase substantially as λ gets closer to one.

Note that, aside from α, only λ enters the asymptotic size formulas. Interestingly, some characteristics of the DGP that may potentially affect asymptotic sizes of the conventional tests are asymptotically negligible. In particular, the estimation uncertainty contained in $\sqrt{n}(\hat{\beta}-\beta)$ does participate in various parts of the stochastic expansion of the J statistic, but eventually cancels out, so the estimation uncertainty does not affect the asymptotic size. Another interesting feature of the asymptotic analysis is that the fourth moments of errors do not enter the asymptotic sizes, even though the formulas for the J and AR statistics do contain second powers of regression errors.

Given the results in Theorem 1, one approach to achieving asymptotically correct size in the presence of many instruments is to divide the J_{DIN} statistic and multiply the AR_{AS} statistic by $\sqrt{1-\lambda}$ (see Lemma 1 in the Appendix). However, we prefer, for a reason to be explained shortly, to correct the critical values of the conventional J and AR tests in such a way that their asymptotic size matches the target size. The *corrected J test* rejects when

$$J > q^{\chi^2(\ell-k)}_{\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right)}. \tag{6}$$

Similarly, the *corrected AR test* rejects when

$$AR > q^{\chi^2(\ell)}_{\Phi\left(\Phi^{-1}(\alpha)/\sqrt{1-\lambda}\right)}. \tag{7}$$

Below we state the asymptotic validity of the corrected J and AR tests under Bekker's asymptotics.

Theorem 2

(a) Suppose assumptions 1–4 hold. Then, the asymptotic size of the corrected J test equals α.

(b) Suppose assumptions 1–3 hold. Then, the asymptotic size of the corrected AR test equals α.

One appealing property of the corrected J and AR tests is that they are robust to numerosity of instruments. This follows from noticing that when ℓ is fixed, $\lambda \to 0$, and the corrected J and AR tests reduce to their conventional forms. By contrast, the corrected versions of the J_{DIN} and AR_{AS} tests based on asymptotic normality are not robust to numerosity of instruments and are invalid when ℓ is fixed.

Another important advantage of the corrected tests is their straightforward computation. The corrected tests are based on the J and AR statistics that are routinely produced by the standard statistical software packages and the χ^2 critical values. The only new input for the J test is $\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right)$ instead of α which can be computed easily (for example, `cdfn(sqrt(1-lambda)*cdfni(alpha))` in GAUSS and `norm(sqrt(1-lambda)*invnorm(alpha))` in STATA for prespecified values of `lambda` and `alpha`). Similar computation is required for $\Phi\left(\Phi^{-1}(\alpha)/\sqrt{1-\lambda}\right)$ to construct the corrected AR test.

4 Simulation Study

To evaluate the finite-sample performance of the proposed tests, we conduct a small simulation study. The data for the Monte Carlo experiment are generated from the model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i, \\ x_i &= \gamma_0 + \sum_{j=1}^{\ell-1} \gamma_j z_{ij} + v_i, \end{aligned} \tag{8}$$

where $\begin{pmatrix} e_i \\ v_i \end{pmatrix} = \text{chol}(\Sigma)\xi_i$, $\begin{pmatrix} \xi_i \\ z_i \end{pmatrix} \sim iid\, N\left(0, I_{\ell+1}\right)$, $\Sigma = \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix}$, $\beta_0 = 0$, $\beta_1 = 1$, $\gamma_0 = 0$ and $\gamma_j = 1/\sqrt{\ell}$ for $j = 1, ..., \ell - 1$. The local-to-zero γ_j'so allow for a drifting strength of each individual instrument but keep the information contained in all instruments fixed (see Assumption 4). The J statistic is used to test the validity of the $\ell - 2$ overidentifying restrictions and the AR statistic is used to test the joint hypothesis of $(\beta_0, \beta_1) = (0, 1)$ and validity of overidentifying restrictions.

Tables 1 and 2 present the empirical size at 5% and 10% nominal level of the conventional and corrected versions of the AR and J tests based on 5,000 Monte Carlo replications. We also include the tests proposed by Andrews and Stock (2007) and Donald, Imbens and Newey (2003) which are obtained under moderately many instruments. The purpose is to compare the quality of the three approximations corresponding to three different asymptotic frameworks (fixed, moderately large and large ℓ).

In order to assess the robustness of the tests to different degrees of overidentification, we consider values of $\lambda = \ell/n$ equal to 0.04, 0.2, 0.5 and 0.8. While $\lambda = 0.8$ may seem excessive, it bears some relevance to empirical applications since situations with similar ratios of number of moment conditions to sample size often arise in evaluating linear asset pricing models of large portfolios and estimating structural macroeconomic models by matching impulse response functions. The values of λ are used in combination with sample sizes of 100, 200 and 500.

Table 1 reports the results for the J test. For $\lambda = 0.04$ and 0.2, the size distortions of the standard J test are relatively small but the empirical rejection rate of this test quickly approaches zero as λ increases. The J_{DIN} test performs only slightly better than the conventional test for $\lambda \geq 0.2$ but it tends to overreject for $\lambda = 0.04$ since the asymptotic normality appears to require much larger values of ℓ. Our corrected J test has coverage very close to the nominal level for all values of λ and sample sizes.

The results for the AR tests are presented in Table 2. As part (b) of Theorem 1 suggests, the standard AR and AR_{AS} tests overreject and the rejection rates increase to 25–30% at 5% nominal level for $\lambda = 0.8$. Our corrected AR test performs much better although it slightly overrejects for large values of λ. As the sample size increases, the rejection rates approach the nominal level but this appears to be slower than in the case of testing for overidentifying restrictions.

5 Conclusions

This paper shows the asymptotic invalidity of the standard AR and J tests of parameter and overidentifying restrictions in the presence of many instruments. If the number of moment conditions is a nontrivial fraction of the sample size, the J test tends to underreject whereas the AR test tends to overreject even in large samples. The versions of the tests by Donald, Imbens and Newey (2003) and Andrews and Stock (2007), obtained under the assumption of moderately many instruments, exhibit an asymptotically equivalent behavior. By allowing the number of instruments to grow at the same rate as the sample size, we propose "corrected" J and AR tests that are chi-square distributed and are asymptotically valid for any number of moment conditions. Due to their simplicity and robustness, we recommend the use of these modified statistics in applied work. A future research agenda includes an extension to non-IID environments, in particular the cases of heteroskedasticity and serial correlation.

References

[1] Andersen, T. G. and B. E. Sorensen (1996): "GMM estimation of a stochastic volatility model: A Monte Carlo study," *Journal of Business and Economic Statistics*, 14, 328–352.

[2] Andrews, D. W. and J. H. Stock (2007): "Testing with many weak instruments," *Journal of Econometrics*, 138, 24–46.

[3] Angrist, J. D. and A. B. Krueger (1991): "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106, 979–1014.

[4] Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an applications to employment equations," *Review of Economic Studies*, 58, 277–297.

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

[6] Burnside, C. and M. Eichenbaum (1996): "Small-sample properties of GMM-based Wald tests," *Journal of Business and Economic Statistics*, 14, 294–308.

[7] Chao, J. and N. Swanson (2005): "Consistent estimation with a large number of weak instruments," *Econometrica*, 73, 1673–1692.

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

[9] Hansen, C., J. Hausman and W. K. Newey (2006): "Estimation with many instrumental variables," *Journal of Business and Economic Statistics*, forthcoming.

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

[11] Koenker, R. and J. A. F. Machado (1999): "GMM inference when the number of moment conditions is large," *Journal of Econometrics*, 93, 327–344.

[12] Ledoit, O. and M. Wolf (2004): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

[13] Newey, W. K. (2004): "Many instrument asymptotics," manuscript, MIT.

[14] Peiser, A. M. (1943): "Asymptotic formulas for significance levels of certain distributions," *Annals of Mathematical Statistics*, 14, 56–62.

[15] Silverstein, J. W. (1995): "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *Journal of Multivariate Analysis*, 55, 331–339.

A Appendix: Proofs

Lemma 1 *Under the conditions of Theorem 1,*

$$\sqrt{\ell}\left(\frac{J}{\ell}-1\right)\xrightarrow{d}N\left(0,2\left(1-\lambda\right)\right),\quad J_{DIN}\xrightarrow{d}N\left(0,1-\lambda\right)$$

and

$$\sqrt{\ell}\left(\frac{AR}{\ell}-1\right)\xrightarrow{d}N\left(0,2/\left(1-\lambda\right)\right),\quad AR_{AS}\xrightarrow{d}N\left(0,2/\left(1-\lambda\right)\right).$$

Proof. First, consider

$$\frac{J_0}{\ell}\equiv\frac{e'Pe}{\ell\sigma^2}=\frac{e'Z\left(Z'Z\right)^{-1}Z'e}{\ell\sigma^2}.$$

Now,

$$\begin{aligned}E\left[\frac{J_0}{\ell}-1\right] &= \frac{1}{\ell\sigma^2}E\left[\operatorname{tr}\left(e'Z\left(Z'Z\right)^{-1}Z'e\right)\right]-1\\ &= \frac{1}{\ell\sigma^2}\operatorname{tr}\left(\left(Z'Z\right)^{-1}Z'E\left[ee'\right]Z\right)-1=\frac{1}{\ell}\operatorname{tr}\left(I_\ell\right)-1=0,\end{aligned}$$

and

$$\begin{aligned}\frac{J_0}{\ell}-1 &= \frac{1}{\ell}\sum_{i=1}^{n}\sum_{j=1}^{n}z_i'\left(Z'Z\right)^{-1}z_j\frac{e_ie_j}{\sigma^2}-1\\ &= \frac{1}{\ell}\sum_{i=1}^{n}z_i'\left(Z'Z\right)^{-1}z_i\left(\frac{e_i^2}{\sigma^2}-1\right)+\frac{1}{\ell}\sum_{i\neq j}z_i'\left(Z'Z\right)^{-1}z_j\frac{e_ie_j}{\sigma^2}\\ &= A_1+A_2.\end{aligned}$$

By the *iid* and moment condition assumptions, A_1 and A_2 are uncorrelated. Let $\kappa=E\left[e_i^4\right]$. The variances of A_1 and A_2 are

$$\begin{aligned}var\left(A_1\right) &= \frac{n}{\ell^2}\left(z_i'\left(Z'Z\right)^{-1}z_i\right)^2\left(\kappa-1\right)=O\left(\frac{1}{\ell}\right),\\ var\left(A_2\right) &= \frac{1}{\ell^2}E\left[\sum_{i\neq j}\sum_{k\neq l}z_i'\left(Z'Z\right)^{-1}z_jz_k'\left(Z'Z\right)^{-1}z_l\frac{e_ie_j}{\sigma^2}\frac{e_ke_l}{\sigma^2}\right]\\ &= \frac{2}{\ell^2}\sum_{i\neq j}\left(z_i'\left(Z'Z\right)^{-1}z_j\right)^2=\frac{2}{\ell^2}\sum_{i=1}^{n}z_i'\left(Z'Z\right)^{-1}\left(\sum_{j=1,j\neq i}^{n}z_jz_j'\right)\left(Z'Z\right)^{-1}z_i\\ &= \frac{2}{\ell^2}\sum_{i=1}^{n}\left(z_i'\left(Z'Z\right)^{-1}z_i-\left(z_i'\left(Z'Z\right)^{-1}z_i\right)^2\right)\leq\frac{2}{\ell^2}n=O\left(\frac{1}{\ell}\right).\end{aligned}$$

Thus, the variance of A_1+A_2 is of order $O\left(1/\ell\right)$ and hence

$$\frac{J_0}{\ell}-1=O_p\left(\frac{1}{\sqrt{\ell}}\right).\tag{9}$$

Second,

$$\frac{\hat{e}'P\hat{e}}{\ell\sigma^2} = \frac{\left(e - X(\hat{\beta}-\beta)\right)' Z\left(Z'Z\right)^{-1} Z'\left(e - X(\hat{\beta}-\beta)\right)}{\ell\sigma^2}$$
$$= \frac{J_0}{\ell} - 2(\hat{\beta}-\beta)'\frac{X'Pe}{\ell\sigma^2} + o\left(\frac{1}{\sqrt{\ell}}\right). \tag{10}$$

Analogously,

$$\frac{\hat{\sigma}^2}{\sigma^2} - 1 = \frac{n}{n-k}\frac{\left(e - X(\hat{\beta}-\beta)\right)'\left(e - X(\hat{\beta}-\beta)\right)}{n\sigma^2} - 1$$
$$= \left(\frac{e'e}{n\sigma^2} - 1\right) - 2(\hat{\beta}-\beta)'\lambda\frac{X'e}{\ell\sigma^2} + o\left(\frac{1}{\sqrt{\ell}}\right).$$

Third,

$$\frac{J}{\ell} - 1 = \left(\frac{\hat{e}'P\hat{e}}{\ell\sigma^2} - 1\right)\frac{\sigma^2}{\hat{\sigma}^2} + \left(\frac{\sigma^2}{\hat{\sigma}^2} - 1\right)$$
$$= \left(\frac{J_0}{\ell} - 1\right) - \left(\frac{e'e}{n\sigma^2} - 1\right) - \frac{2}{\lambda\sigma^2}(\hat{\beta}-\beta)'\frac{X'\left(P-\lambda I\right)e}{n} + o_p\left(\frac{1}{\sqrt{\ell}}\right), \tag{11}$$

by (9) and (10) and because $\hat{\sigma}^2 = \sigma^2 + O\left(1/\sqrt{\ell}\right)$. Consider the third term

$$\frac{X'\left(P-\lambda I\right)e}{n} = \frac{\Upsilon'\left(P-\lambda I\right)e}{n} + \frac{V'\left(P-\lambda I\right)e}{n}. \tag{12}$$

The first term has mean zero and variance

$$\frac{\Upsilon'\left(P-\lambda I\right)\left(P-\lambda I\right)\Upsilon}{n^2} = (1-2\lambda)\frac{\Upsilon'P\Upsilon}{n^2} + \lambda^2\frac{\Upsilon'\Upsilon}{n^2} \to 0$$

because of Assumption 4 and the Cauchy–Schwarz inequality implying $\Upsilon'Z\left(Z'Z\right)^{-1}Z'\Upsilon \leq \Upsilon'\Upsilon$. Therefore, the first term in (12) is $o_p(1)$. Along the lines of Newey (2004, proof of Lemma 1) one can see that the second term in (12) has expected value

$$E\left[\frac{V'\left(P-\lambda I\right)e}{n}\right] = E\left[\frac{V'Pe}{n}\right] - \lambda E\left[\frac{V'e}{n}\right] = \frac{\ell}{n}E\left[v_i e_i\right] - \lambda E\left[v_i e_i\right] = 0$$

and variance that is $O(1/n)$. Therefore, the whole term (12) is $o_p(1)$. Thus, up to a $o_p(1)$ remainder,

$$\sqrt{\ell}\left(\frac{J}{\ell} - 1\right) \overset{A}{=} \frac{1}{\sqrt{\ell}}\sum_{i=1}^{n}\left(z_i'\left(Z'Z\right)^{-1}z_i - \lambda\right)\left(\frac{e_i^2}{\sigma^2} - 1\right) + \frac{1}{\sqrt{\ell}}\sum_{i\neq j} z_i'\left(Z'Z\right)^{-1}z_j\frac{e_ie_j}{\sigma^2} = B_1 + B_2.$$

Exactly as before, we compute the variance of the zero-mean term B_1 which yields

$$var\left(B_1\right) = \frac{n}{\ell}\left(z_i'\left(Z'Z\right)^{-1}z_i - \lambda\right)^2(\kappa - 1) \to 0.$$

Therefore, $B_1 = o_p(1)$. In order to derive the asymptotics for B_2, we check the conditions for the central limit theorem by Kelejian and Prucha (2001, Theorem 1) for linear quadratic forms where $b_{i,n} \equiv 0$. Assumption 1 of this CLT is satisfied for $\varepsilon_{i,n} \equiv e_i/\sigma$. Next, we verify Assumption 2 of this CLT for

$$a_{ij,n} \equiv \frac{1}{\sqrt{\ell}} z_i' \left(Z'Z\right)^{-1} z_j.$$

First, $a_{ij,n}$ is clearly symmetric. Second,

$$\begin{aligned}\sum_{i=1}^{n} |a_{ij,n}| &\leq \frac{1}{\sqrt{\ell}} \sum_{i=1}^{n} \left| z_i' \left(Z'Z\right)^{-1} z_j \right| \leq \sqrt{\frac{n}{\ell}} \left(\sum_{i=1}^{n} \left(z_i' \left(Z'Z\right)^{-1} z_j \right)^2 \right)^{1/2} \\ &= \sqrt{\frac{1}{\lambda}} \left(z_j' \left(Z'Z\right)^{-1} z_j \right)^{1/2} \leq \sqrt{\frac{1}{\lambda}}.\end{aligned}$$

Consequently, $\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^{n} |a_{ij,n}| < \infty$ in Assumption 2 of the CLT of Kelejian and Prucha (2001, Theorem 1) is satisfied. Finally, in their assumption 3(a) $\sup_{1 \leq i \leq n, n \geq 1} E\left[|\varepsilon_{i,n}|^{2+\eta} \right] < \infty$ holds by Assumption 1. Hence, the variance of B_2 is

$$var\left(B_2\right) = \frac{2}{\ell} \sum_{i=1}^{n} \left(z_i' \left(Z'Z\right)^{-1} z_i - \left(z_i' \left(Z'Z\right)^{-1} z_i \right)^2 \right) = \frac{2}{\ell} n \left(\lambda - \lambda^2\right) + o\left(1\right) \to 2\left(1 - \lambda\right)$$

and the limiting distribution of $\sqrt{\ell}\left(J/\ell - 1\right)$ is

$$\sqrt{\ell} \left(\frac{J}{\ell} - 1 \right) \xrightarrow{d} N\left(0, 2\left(1 - \lambda\right)\right).$$

For the AR test, note that

$$\frac{AR}{\ell} = \left(1 - \lambda\right) \left(\frac{e'e}{n\sigma^2} - \lambda \frac{J_0}{\ell} \right)^{-1} \frac{J_0}{\ell},$$

so

$$\left(1 - \lambda\right) \left(\frac{AR}{\ell} - 1 \right) = \left(\frac{J_0}{\ell} - 1 \right) - \left(\frac{e'e}{n\sigma^2} - 1 \right) + o_p\left(\frac{1}{\sqrt{\ell}} \right),$$

and, proceeding as before with (11), we get

$$\left(1 - \lambda\right) \sqrt{\ell} \left(\frac{AR}{\ell} - 1 \right) \xrightarrow{d} N\left(0, 2\left(1 - \lambda\right)\right).$$

Proof of Theorem 1. From Peiser (1943) it follows that

$$q_\alpha^{\chi^2(\ell - k)} = \ell - k + \Phi^{-1}\left(1 - \alpha\right) \sqrt{2\left(\ell - k\right)} + O\left(1\right) \tag{13}$$

or

$$\frac{q_\alpha^{\chi^2(\ell - k)}}{\ell} - 1 = \Phi^{-1}\left(1 - \alpha\right) \sqrt{\frac{2}{\ell}} + O\left(\frac{1}{\ell} \right).$$

Then, the size of the conventional J test is

$$\begin{aligned}
\Pr\left\{J > q_{\alpha}^{\chi^2(\ell-k)}\right\} &= \Pr\left\{\sqrt{\frac{\ell}{2(1-\lambda)}}\left(\frac{J}{\ell}-1\right) > \sqrt{\frac{\ell}{2(1-\lambda)}}\left(\frac{q_{\alpha}^{\chi^2(\ell-k)}}{\ell}-1\right)\right\} \\
&= \Pr\left\{N(0,1)+o_d(1) > \frac{\Phi^{-1}(1-\alpha)}{\sqrt{1-\lambda}} + O\left(\frac{1}{\sqrt{\ell}}\right)\right\} \\
&= 1-\Phi\left(\frac{\Phi^{-1}(1-\alpha)}{\sqrt{1-\lambda}}\right)+o(1) \to \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1-\lambda}}\right).
\end{aligned}$$

Using Lemma 1, the size of the Donald, Imbens, and Newey (2003) J test is

$$\Pr\left\{J_{DIN} > q_{1-\alpha}^{N(0,1)}\right\} = \Pr\left\{N(0,1-\lambda)+o_d(1) > \Phi^{-1}(1-\alpha)\right\} \to \Phi\left(\frac{\Phi^{-1}(\alpha)}{\sqrt{1-\lambda}}\right).$$

Similarly, the size of the conventional AR test is

$$\begin{aligned}
\Pr\left\{AR > q_{\alpha}^{\chi^2(\ell)}\right\} &= \Pr\left\{\sqrt{\frac{\ell(1-\lambda)}{2}}\left(\frac{AR}{\ell}-1\right) > \sqrt{\frac{\ell(1-\lambda)}{2}}\left(\frac{q_{\alpha}^{\chi^2(\ell-k)}}{\ell}-1\right)\right\} \\
&= \Pr\left\{N(0,1)+o_d(1) > \sqrt{1-\lambda}\Phi^{-1}(1-\alpha)+O\left(\frac{1}{\sqrt{\ell}}\right)\right\} \\
&= 1-\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(1-\alpha)\right)+o(1) \to \Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right).
\end{aligned}$$

Using Lemma 1, the size of the Andrews and Stock (2007) AR test is

$$\Pr\left\{AR_{AS} > q_{1-\alpha}^{N(0,2)}\right\} = \Pr\left\{N\left(0,\frac{2}{1-\lambda}\right)+o_d(1) > \sqrt{2}\Phi^{-1}(1-\alpha)\right\} \to \Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right).$$

Proof of Theorem 2. Using expansion (13), the actual size of the corrected J test (6) is

$$\begin{aligned}
&\Pr\left\{J > q_{\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right)}^{\chi^2(\ell-k)}\right\} \\
&= \Pr\left\{\sqrt{\frac{\ell}{2(1-\lambda)}}\left(\frac{J}{\ell}-1\right) > \frac{\Phi^{-1}\left(1-\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right)\right)}{\sqrt{1-\lambda}}+O\left(\frac{1}{\sqrt{\ell}}\right)\right\} \\
&= \Pr\left\{N(0,1)+o_d(1) > -\Phi^{-1}(\alpha)\right\} = 1-\Phi\left(-\Phi^{-1}(\alpha)\right)+o(1) \to \alpha.
\end{aligned}$$

Similarly, the actual size of the corrected AR test (7) is

$$\begin{aligned}
&\Pr\left\{AR > q_{\Phi\left(\Phi^{-1}(\alpha)/\sqrt{1-\lambda}\right)}^{\chi^2(\ell-k)}\right\} \\
&= \Pr\left\{\sqrt{\frac{\ell(1-\lambda)}{2}}\left(\frac{AR}{\ell}-1\right) > \sqrt{1-\lambda}\Phi^{-1}\left(1-\Phi\left(\Phi^{-1}(\alpha)/\sqrt{1-\lambda}\right)\right)+O\left(\frac{1}{\sqrt{\ell}}\right)\right\} \\
&= \Pr\left\{N(0,1)+o_d(1) > -\Phi^{-1}(\alpha)\right\} \to \alpha.
\end{aligned}$$

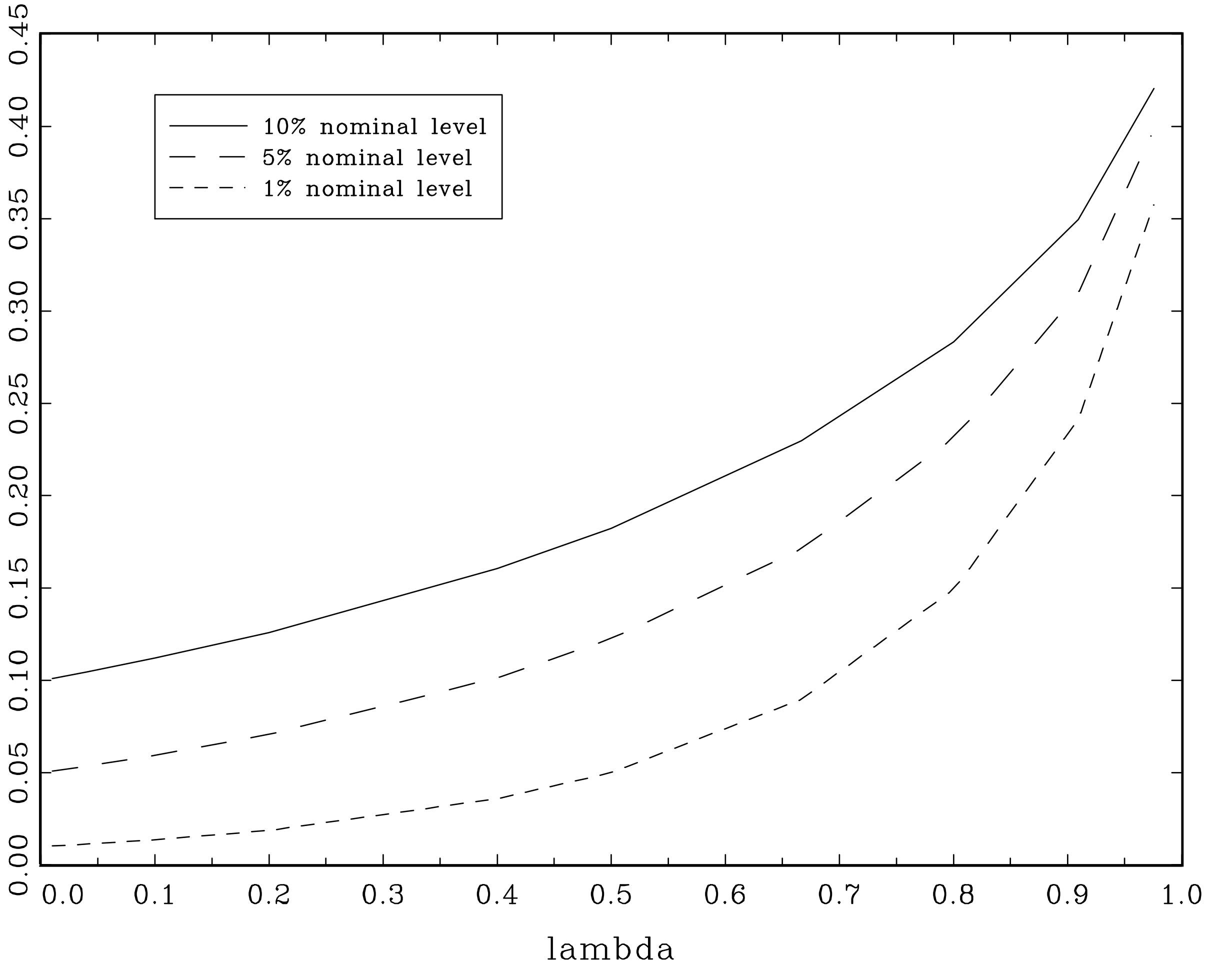

Figure 1: Asymptotic size (p-value) of the conventional AR (J) test as a function of $\lambda \in [0, 1)$.

Table 1. Empirical rejection rates at 5% and 10% nominal level of the J tests.

	5%				10%			
	$\lambda = .04$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$	$\lambda = .04$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$
$n = 100$								
J	5.06%	2.66%	0.52%	0.00%	10.38%	7.40%	3.08%	0.02%
J_{DIN}	7.12%	4.08%	0.92%	0.00%	10.66%	8.08%	3.52%	0.02%
$J_{corrected}$	5.50%	4.54%	4.76%	4.52%	10.88%	9.96%	10.30%	10.54%
$n = 200$								
J	4.92%	3.00%	0.84%	0.00%	10.14%	7.40%	3.02%	0.02%
J_{DIN}	7.00%	3.84%	1.02%	0.00%	10.84%	7.92%	3.22%	0.02%
$J_{corrected}$	5.24%	4.94%	4.44%	4.96%	10.56%	9.98%	10.00%	10.54%
$n = 500$								
J	5.44%	3.28%	0.62%	0.00%	10.50%	8.02%	2.82%	0.01%
J_{DIN}	6.90%	4.04%	0.82%	0.00%	11.14%	8.42%	2.92%	0.01%
$J_{corrected}$	5.94%	5.20%	4.20%	4.62%	10.98%	10.44%	9.54%	10.44%

Notes: J, J_{DIN} and $J_{corrected}$ denote the conventional J test, the J statistic of Donald, Imbens and Newey (2003) and the test proposed in this paper, respectively.

Table 2. Empirical rejection rates at 5% and 10% nominal level of the AR tests.

	5%				10%			
	$\lambda = .04$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$	$\lambda = .04$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$
$n = 100$								
AR	6.28%	7.40%	14.52%	29.04%	11.58%	12.96%	20.40%	33.97%
AR_{AS}	8.58%	8.80%	15.68%	29.97%	12.22%	13.94%	20.86%	34.36%
$AR_{corrected}$	5.94%	5.22%	6.96%	9.36%	11.08%	10.10%	12.28%	14.86%
$n = 200$								
AR	5.26%	7.90%	13.34%	27.03%	10.80%	13.56%	19.46%	31.95%
AR_{AS}	7.32%	9.12%	14.46%	27.79%	11.52%	14.06%	19.76%	32.29%
$AR_{corrected}$	4.96%	5.78%	5.98%	8.40%	10.36%	10.78%	11.34%	13.52%
$n = 500$								
AR	6.12%	8.00%	13.34%	25.15%	11.46%	13.94%	19.26%	29.67%
AR_{AS}	7.36%	8.94%	13.92%	25.67%	12.16%	14.44%	19.68%	29.90%
$AR_{corrected}$	5.78%	5.86%	4.98%	6.80%	10.76%	11.10%	10.52%	12.34%

Notes: AR, AR_{AS} and $AR_{corrected}$ denote the conventional Anderson–Rubin test, the AR statistic of Andrews and Stock (2007) and the test proposed in this paper, respectively.