Appendix to "Specification Testing in Models with Many Instruments"

Stanislav Anatolyev
New Economic School

Nikolay Gospodinov
Concordia University and CIREQ

1 Simulation Results with Heterogeneous Instruments

This simulation experiment evaluates the sensitivity of the test statistics to possibly heterogeneous instruments. The data for the Monte Carlo experiment are generated from the model

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

where $(e_i, v_i)' \sim iid\, N(0, \Sigma)$, $\Sigma = \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix}$, $\beta_0 = 0$, $\beta_1 = 1$, and $\gamma_j = 1/\sqrt{\ell}$ for $j = 1, ..., \ell$. The matrix of instruments Z is generated as described below. The local-to-zero γ_j's 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. The number of Monte Carlo replications is 5,000, the sample size is $n = 100$, 200 and 500 and the values of λ are 0.2, 0.5 and 0.8.

1.1 Heterogeneity across columns

First, we generate a matrix of instruments Z with heterogeneity across columns. In particular, we verify the robustness of our results in the paper to the distributional specification of instruments, dependence among instruments and the presence of deterministic and discrete instruments. The matrix of instruments Z $(n \times \ell)$ is comprised of five blocks each of dimension $n \times \ell/5$. The first three blocks of instruments are generated as standard uniform, standard normal and $\chi^2(1)$ random variables that are independent of each other. The final two blocks of instruments are dummy variables summing up to a constant term and interaction terms between instruments from the first four blocks half of which are interactions between randomly drawn (without replacement) dummy and continuous instruments and another half are interactions between randomly drawn continuous instruments. Finally, the columns of the resulting instrument matrix are randomly reshuffled.

The empirical rejection tests of the J and AR tests considered in the paper are reported in Tables A1 and A2, respectively. Overall, the results are very similar to the case of homogeneous regressors considered in the paper. The empirical size of the corrected J test is very close to its nominal level across all sample sizes and values of λ. The corrected AR test tends to overreject for small sample sizes and large values of λ but it approaches its nominal level as the sample size increases.

1.2 Heterogeneity across rows

While Assumption 3 in the paper rules out possible heterogeneity across rows in the instrument matrix, it is interesting to see how the results in Theorems 1 and 2 are affected if Assumption 3 is violated. For this reason, we generated an instrument matrix that is heterogeneous across i ($i = 1, ..., n$). In particular, the first half of the instrument matrix is comprised of standard normal random variables and the second half of the sample are $\chi^2(1)$ random variables. We also add a constant to the vector of instruments.

The results for the J and AR tests are reported in Tables A3 and A4 and show that the test statistics appear to be rather insensitive to the presence of heterogeneity of instruments across observations. Very similar results (not reported here) are obtained when the second half of the sample are normal random variables with mean zero and standard deviation 10. These simulation findings suggest that the corrected test statistics exhibit very little sensitivity to deviations from homogeneity in the instrument matrix.

TABLE A1. Empirical rejection rates at 5% and 10% nominal level of the J tests.

	5%			10%		
	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$
$n = 100$						
J	3.01%	0.48%	0.00%	7.50%	3.32%	0.00%
J_{DIN}	4.18%	0.76%	0.00%	8.28%	3.74%	0.00%
J_{corr}	4.84%	4.98%	4.26%	10.2%	11.1%	10.9%
$n = 200$						
J	2.80%	0.66%	0.00%	7.26%	3.28%	0.00%
J_{DIN}	3.66%	1.00%	0.00%	7.96%	3.42%	0.00%
J_{corr}	4.46%	4.82%	4.34%	10.0%	9.94%	10.7%
$n = 500$						
J	3.16%	0.96%	0.00%	7.22%	3.56%	0.01%
J_{DIN}	3.76%	1.08%	0.00%	7.54%	3.72%	0.00%
J_{corr}	4.72%	5.06%	4.76%	9.60%	10.5%	10.6%

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

TABLE A2. Empirical rejection rates at 5% and 10% nominal level of the AR tests.

	5%			10%		
	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$
$n = 100$						
AR	7.66%	15.1%	26.8%	13.3%	21.2%	31.5%
AR_{AS}	9.54%	16.3%	27.7%	13.8%	21.6%	32.0%
AR_{corr}	5.78%	7.44%	9.34%	11.0%	12.9%	14.7%
$n = 200$						
AR	7.40%	13.2%	26.9%	13.0%	19.0%	32.0%
AR_{AS}	8.60%	14.1%	27.7%	13.8%	19.4%	32.3%
AR_{corr}	5.32%	6.32%	7.92%	10.5%	11.1%	13.5%
$n = 500$						
AR	6.90%	13.5%	25.0%	12.4%	19.3%	30.4%
AR_{AS}	7.88%	14.1%	25.7%	12.9%	19.6%	30.6%
AR_{corr}	5.20%	5.90%	7.10%	9.92%	11.4%	12.2%

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

TABLE A3. Empirical rejection rates at 5% and 10% nominal level of the J tests.

	5%			10%		
	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$
$n = 100$						
J	3.24%	0.44%	0.00%	7.48%	3.00%	0.00%
J_{DIN}	4.38%	0.74%	0.00%	7.90%	3.38%	0.00%
J_{corr}	4.98%	4.70%	3.98%	9.66%	10.3%	11.0%
$n = 200$						
J	2.74%	0.68%	0.00%	7.44%	2.68%	0.00%
J_{DIN}	3.54%	0.84%	0.00%	8.06%	2.92%	0.00%
J_{corr}	4.50%	4.44%	4.52%	10.4%	9.88%	9.86%
$n = 500$						
J	2.96%	0.94%	0.00%	7.12%	3.22%	0.00%
J_{DIN}	3.52%	1.06%	0.00%	7.62%	3.36%	0.00%
J_{corr}	4.54%	4.96%	5.04%	9.66%	9.78%	10.8%

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

TABLE A4. Empirical rejection rates at 5% and 10% nominal level of the AR tests.

	5%			10%		
	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$	$\lambda = .2$	$\lambda = .5$	$\lambda = .8$
$n = 100$						
AR	7.62%	14.3%	28.8%	12.9%	20.0%	33.8%
AR_{AS}	9.24%	15.7%	29.5%	13.5%	20.5%	34.2%
AR_{corr}	5.72%	6.96%	9.56%	10.3%	12.0%	15.6%
$n = 200$						
AR	7.88%	13.7%	26.5%	14.4%	19.8%	3.17%
AR_{AS}	9.00%	14.8%	27.3%	14.9%	20.4%	3.20%
AR_{corr}	5.38%	6.18%	7.44%	10.9%	11.1%	12.7%
$n = 500$						
AR	6.90%	12.9%	25.9%	12.3%	19.1%	31.0%
AR_{AS}	8.04%	13.5%	26.4%	12.8%	19.4%	31.3%
AR_{corr}	5.06%	5.82%	7.24%	9.94%	10.3%	12.8%

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