Finite sample properties of the Anderson and Rubin (1949) test

Maurice Bun[a], Helmut Farbmacher[b,c*], Rutger Poldermans[a]

[a]Amsterdam School of Economics, University of Amsterdam, The Netherlands

[b]Department of Economics, University of Mannheim, Germany

[c]Munich Center for the Economics of Aging, Max Planck Society, Germany

July 26, 2017

Abstract

Most studies nowadays use uncentered (as opposed to centered) moment conditions to form the weighting matrix for the GMM version of the Anderson and Rubin (AR) test statistic. Remarkably, both versions of the GMM-AR statistic do not cover the usual definition of the AR statistic under homoskedasticity (IV-AR). We propose a finite sample correction for the GMM-AR test statistic, which nests the usual IV-AR statistic and performs distinctly better in finite samples. Moreover, we derive an asymptotic distribution of the IV-AR under homoskedasticity but non-normal errors which has correct size even if the number of instruments is as large as the sample size.

Key Words: Many instruments, Anderson and Rubin test, Asymptotic distribution, Non-normality

*farbmacher@mpisoc.mpg.de

1 INTRODUCTION

Most studies nowadays use uncentered (as opposed to centered) moment conditions to form the weighting matrix for the GMM version of the Anderson and Rubin (1949) test statistic (GMM-AR). When the number of moment conditions is fixed or increases at a slower rate than the number of observations, this choice does not matter asymptotically. However, in finite samples it affects inference. We show that the GMM-AR statistic based on the centered or uncentered weighting matrix can analogously be interpreted as basic Wald or Lagrange multiplier (LM) test statistics. Having established these links, we use the knowledge about the finite-sample behaviors of the Wald and LM test to improve the small-sample properties of the GMM-AR test.

Most closely related to our study is the work by Anatolyev and Gospodinov (2011) on the validity of the AR test in a linear model under homoskedasticity (IV-AR). They propose a many instruments modification for the IV-AR statistic, which has asymptotically correct size under few and many instruments. In addition to their result, we show that the IV-AR statistic is asymptotically F-distributed—even under non-normal errors. Moreover, we show that using a centered or uncentered definition of the weighting matrix does not matter for inference in the homoskedastic linear model.

The picture is different, however, when we allow for heteroskedasticity. While our asymptotic results for the homoskedastic linear model cannot be generalized to GMM, we discuss a degrees-of-freedom correction which markedly improves the finite sample properties of the GMM-AR test. Moreover, it nests the IV-AR statistic as a special case, and can thus be seen as a generalization of the IV-AR test to the GMM setup.

We obtain a more accurate understanding of the extent to which the centering of the weighting matrix influences the size of the GMM-AR test through Monte Carlo simulations. As long as the number of moment conditions (m) is small compared to the number of observations (n) (as it is often the case in microeconomic applications), the difference between a centered and uncentered definition of the weighting matrix is negligible but in macroeconomic applications m/n can be considerable, and here the choice between these two definitions is essential. In small samples (relative to the number of moment conditions) we see a strong difference in the actual size of asymptotically (if $m/n \to 0$) equivalent test statistics, which points to conflicting inferences in practice.

2 MODEL AND INFERENCE

2.1 Model and Test Statistics

To describe the model, let w_i ($i = 1, ..., n$) be independent and identically distributed observations of a data vector w. $g(w, \beta) = (g_1(w, \beta), ..., g_m(w, \beta))'$ is an $m \times 1$ vector of functions of w and a $p \times 1$ vector of parameters, β, where $m \geq p$. β_0 is a $p \times 1$ vector of true parameters satisfying the moment conditions

$$E[g(w_i, \beta_0)] = 0. \tag{1}$$

We want to test the null hypothesis $H_0 : \beta_0 = \beta$ using the Anderson and Rubin (1949) test. We start with the important special case of a linear model, $g(w_i, \beta_0) = z_i(y_i - x_i'\beta_0)$ imposing additionally homoskedasticity and relax these assumptions later. The IV-AR test statistic is defined as

$$\frac{u'P_Z u}{u'M_Z u/(n-m)}\,. \tag{2}$$

To the best of our knowledge, there is only this definition in the IV setup, which has been used for instance by Anderson and Rubin (1949), Staiger and Stock (1997) or Bekker and Kleibergen (2003).
On the other hand, there are two versions of the GMM-AR test in the literature (see Table 1). The first version uses a weighting matrix which is based on uncentered moments, i.e.,

$$\widehat{\Omega}(\beta) = n^{-1} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)' \,, \tag{3}$$

where $g_i(\beta) = g(w_i, \beta)$. The second definition of the weighting matrix is based on centered moments, i.e.,

$$\widetilde{\Omega}(\beta) = n^{-1} \sum_{i=1}^{n} [g_i(\beta) - \widehat{g}(\beta)][g_i(\beta) - \widehat{g}(\beta)]' = \widehat{\Omega}(\beta) - \widehat{g}(\beta)\widehat{g}(\beta)' \,. \tag{4}$$

where $\widehat{g}(\beta) = n^{-1} \sum_{i=1}^{n} g_i(\beta)$. The advantage of this definition is that it is also valid when $\beta \neq \beta_0$ (Hall, 2000). The corresponding GMM-AR test statistics are defined as

$$\begin{aligned} \widehat{AR} &= n\,\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}\widehat{g}(\beta) \,, \\ \widetilde{AR} &= n\,\widehat{g}(\beta)'\widetilde{\Omega}(\beta)^{-1}\widehat{g}(\beta) \,, \end{aligned} \tag{5}$$

Remarkably, both versions of the GMM-AR statistic do not cover the usual IV-AR statistic from Equation (2) as a special case. Therefore, we propose a third definition of the GMM-AR statistic, which can be seen as a generalization of Equation (2) to the GMM case:

$$\widetilde{AR}_{df} = n\,\widehat{g}(\beta)'\widetilde{\Omega}_{df}(\beta)^{-1}\widehat{g}(\beta) \,, \tag{6}$$

with

$$\widetilde{\Omega}_{df}(\beta) = (n-m)^{-1} \sum_{i=1}^{n} [g_i(\beta) - \widehat{g}(\beta)][g_i(\beta) - \widehat{g}(\beta)]' = \frac{n}{n-m} (\widehat{\Omega}(\beta) - \widehat{g}(\beta)\widehat{g}(\beta)') \,. \tag{7}$$

Compared to the commonly used definition of the centered weighting matrix, our definition incorporates a degrees-of-freedom correction, which clearly matters in finite samples. To the best of our knowledge, this definition has not been used in the GMM literature so far. To show that the well known IV-AR statistic as defined in Equation (2) is a special case of our definition of the GMM-AR statistic in (6), we will make use of the following relations:

$$\begin{aligned} \widetilde{AR}_{df} &= \frac{n-m}{n} \widetilde{AR} \\ \widetilde{AR} &= \frac{\widehat{AR}}{1 - \widehat{AR}/n} \,. \end{aligned} \tag{8}$$

The latter link has already been established in Newey and Smith (2004) and Antoine *et al.* (2007). Then we can write:

$$\begin{aligned} \frac{u'P_Z u}{u'M_Z u/(n-m)} &= (n-m) \frac{\frac{u'P_Z u}{u'u}}{1 - \frac{u'P_Z u}{u'u}} \\ &= \frac{n-m}{n} \frac{n\widehat{Q}_L}{1 - \widehat{Q}_L} \\ &= \widetilde{AR}_{df,L} \,, \end{aligned} \tag{9}$$

where $\widehat{Q}_L$ is the criterion function of the Liml estimator and $\widetilde{AR}_{df,L}$ the corresponding AR test under conditional homoskedasticity.

2.2 IV Inference

Interpreting the IV-AR test as an F-test from an auxiliary regression of

$$u = Z\gamma + w \tag{10}$$

and testing $H_0 : \gamma = 0$, we have

$$F = \frac{\widehat{\gamma}' Z' Z \widehat{\gamma}}{\widehat{\sigma}^2_w} / m = \frac{u' P_Z u}{u' M_Z u/(n-m)} / m = \frac{\widetilde{AR}_{df,L}}{m} \,. \tag{11}$$

Adapting Anatolyev (2012)'s Theorem 3 about the asymptotic distribution of the F test with many restrictions to the IV-AR test, we can show that

$$Pr\left(\frac{\widetilde{AR}_{df,L}}{m} > q_\alpha^{F(m,n-m)} \right) \to \alpha \,. \tag{12}$$

This holds not only under many instrument asymptotics (i.e., $m/n \to \lambda > 0$) but also under non-normal errors. The latter result is new in the literature. While it is well known that the F-distribution is appropriate under normal errors, there is so far no result in the literature showing that it is also appropriate under non-normal errors. Moreover, from a standard link between the F-distribution and the Beta-distribution it follows that if $\widetilde{AR}_{df,L}/m = \frac{(n-m)}{m} \frac{\widehat{Q}_L}{1-\widehat{Q}_L} \sim F(m, n-m)$, then $\widehat{Q}_L = \widetilde{AR}_L/n \sim \text{Beta}\left(\frac{m}{2}, \frac{n-m}{2}\right)$. This result implies that the definition of the weighting matrix does not matter for the IV-AR test as long as we use the appropriate asymptotic approximation.
Anatolyev and Gospodinov (2011) show that $\widetilde{AR}_{df,L} \sim N(m, 2m/(1-\lambda))$. Expectation and variance derived from our results are asymptotically equivalent to the results derived in Anatolyev and Gospodinov (2011). However, the normal approximation performs poorly at both ends of $\lambda \in [0, 1]$. They successfully solve this issue if λ is small by using a χ^2 approximation with corrected critical values. If λ is large, their approximation, however, overrejects as they report in their simulations. The Beta-distribution—and hence also the F-distribution—solves this issue at both ends of λ as it can change its skewness over the parameter space of λ. The Beta-distribution is right-skewed, just like the χ^2-distribution, when λ is small and is left-skewed when λ is close to 1.

2.3 GMM Inference

In the following we give an analogue interpretation for the GMM-AR test, which illustrate the finite-sample differences between all three definitions and provides some guidance on how to choose between them. Let $G(\beta) = [g_1(\beta), \cdots, g_n(\beta)]'$ and i be an $n \times 1$ vector of ones. Thus, $\widehat{g}(\beta) = G(\beta)' i/n$ and $\widehat{\Omega}(\beta) = G(\beta)' G(\beta)/n$. In the following we will suppress the dependence of G on β. Consider the following linear projection

$$i = G\gamma + e \tag{13}$$

with parameter $\widehat{\gamma} = (G'G)^{-1} G' i$, prediction $\widehat{i} = G(G'G)^{-1} G' i$ and $\widehat{e} = i - \widehat{i}$. The total sum of squares can be partitioned into the explained and the residual sum of squares as follows

$$\begin{aligned} i'i &= \widehat{i}'\widehat{i} + \widehat{e}'\widehat{e} \\ \Leftrightarrow 1 &= \widehat{i}'\widehat{i}/i'i + \widehat{e}'\widehat{e}/i'i \end{aligned} \tag{14}$$

where $\widehat{e}'\widehat{e}/i'i = \widehat{e}'\widehat{e}/n = \widehat{\sigma}_e^2$ and

$$
\begin{aligned}
n\,R^2 = \widehat{i}'\widehat{i} &= i'G(G'G)^{-1}G'G(G'G)^{-1}G'i \\
&= n\,i'G/n(G'G/n)^{-1}G'i/n \\
&= n\,\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}\widehat{g}(\beta) \\
&= n\,\widehat{Q}\,.
\end{aligned}
\tag{15}
$$

Thus, $\widehat{Q}$, which is the criterion function of the CUE estimator, is also the coefficient of determination (R^2) in Equation (13). The GMM-AR test, which is usually defined as $\widehat{AR} = n\widehat{Q}$, can thus be interpreted as a Lagrange Multiplier (LM) test of joint significance of γ. It is interesting to note that there is also an analogue interpretation of the GMM-AR statistic based on a centered weighting matrix. Namely, it can be interpreted as the Wald statistic of joint significance of γ, which is defined as $W = \sqrt{n}(\widehat{\gamma}-0)'\widehat{Var}(\sqrt{n}\widehat{\gamma})^{-1}\sqrt{n}(\widehat{\gamma}-0)$ with $\widehat{Var}(\sqrt{n}\widehat{\gamma}) = \widehat{\sigma}_e^2(G'G/n)^{-1}$. It follows that

$$
\begin{aligned}
W &= ni'G(G'G)^{-1}(G'G/n)(G'G)^{-1}G'i/\widehat{\sigma}_e^2 \\
&= n\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}\widehat{\Omega}(\beta)\widehat{\Omega}(\beta)^{-1}\widehat{g}(\beta)/\widehat{\sigma}_e^2 \\
&= \widehat{AR}/[1-\widehat{AR}/n] = \widetilde{AR}.
\end{aligned}
\tag{16}
$$

Both AR statistics are asymptotically $\chi^2(m)$ distributed under the null hypothesis but their finite sample behaviors differ in the same way that LM statistics differ from Wald tests in finite samples. $\widetilde{AR}$ has a larger rejection frequency in finite samples than $\widehat{AR}$. This is in line with the well-known fact that the Wald and LM statistics in linear models satisfy the inequality $W \geq LM$ (see Berndt and Savin, 1977; Breusch, 1979; Newey and West, 1987). The gap between $\widehat{AR}$ and $\widetilde{AR}$ increases with rising $\widehat{Q}$, which in turn—evaluated at β_0—increases in m and decreases in n (see Lemma A15 in Newey and Windmeijer, 2009b).
Having established the link between the centered AR statistic and the Wald statistic, we can use the knowledge about the finite-sample behavior of the Wald test to improve the small sample properties of the $\widetilde{AR}$ statistic. Usually, Wald statistics contain a degrees-of-freedom correction to correct the bias in the estimator of the inverse of σ_e^2. While $\widehat{\sigma}_e^2$ is an unbiased estimator of the population variance, $1/\widehat{\sigma}_e^2$ is a biased estimator of the inverse of the population variance. An unbiased estimator would be $(n-m)/n * 1/\widehat{\sigma}_e^2$ (see Lemma 7.7.1 in Anderson, 2003) and the corresponding AR statistic would thus be

$$
\widetilde{AR}_{df} = \frac{n-m}{n}\widetilde{AR}\,. \tag{17}
$$

This is our proposed definition of the GMM-AR test from Equation (6). While the degrees-of-freedom correction does not affect the asymptotic properties of the Wald test (if $m/n \to 0$), it is known to be important in small samples (see, e.g., Evans and Savin 1982). Therefore we expect $\widetilde{AR}_{df}$ to perform better in finite samples than $\widetilde{AR}$ in the same way we expect a Wald test with degrees-of-freedom correction to perform better than without correction. On the other hand, for the LM test (and $\widehat{AR}$, respectively) we do not need a degrees-of-freedom correction to get good size properties in small samples (see, e.g., Evans and Savin, 1982). Finally, we have three asymptotically (if $m/n \to 0$) equivalent AR test statistics which differ in their small sample behavior and thus may lead to conflicting inferences in practice. We obtain a more accurate understanding of the extent to which this affects the size and the power of the discussed tests through simulations.

3 SIMULATION RESULTS

We conducted a series of Monte Carlo simulations for both linear and nonlinear models, and with normal and non-normal errors. The design of the linear model is

$$y_i = \beta_0 x_i + u_{1i} + u_{2i}$$
$$x_i = z_i'\pi + v_i$$
$$u_{1i} = \rho v_i + \sqrt{1-\rho^2} w_i$$
$$v_i \sim N(0,1),\ \ w_i \sim N(0,1),\ z_i \sim N(0,I), \qquad \pi = \sqrt{\frac{CP}{mn}}\iota_m$$

where ι_m is an m-vector of ones. u_{2i} is either standard Cauchy or $\chi^2(2)$ distributed. x has no causal effect on y (i.e., $\beta_0 = 0$) and the constant is set to zero as well. The sample size n is 100; the degree of endogeneity ρ is set to 0.5. We hold the asymptotic F statistic fixed at 1, which implies that the set of instruments is equally weak with varying number of instruments (i.e., $CP = F^{\infty} * m$). For the nonlinear specification we replace the linear index by $exp(\beta_0 x_i)$, set $\beta_0 = 1$, drop u_{2i}, and let u_{1i} still be normally distributed.

Figure 1 depicts the rejection frequencies for the IV-AR ($\widetilde{AR}_{df,L}$) under various asymptotic approximations. The usual χ^2_m approximation performs very poorly once the number of instruments is large relative to the sample size. Anatolyev and Gospodinov's (2011) correction of the critical values performs distinctly better but still overrejects. This is in line with their reported Monte Carlo results. However, comparing $\widetilde{AR}_{df}/m$ with the critical value from the F-distribution as we suggest, gives the correct actual size even when the errors are non-normal. Note that the uncentered IV-AR rejects exactly the same set of simulation replications when we compare $\widehat{AR}_L/n$ with the critical value obtained from the Beta distribution.

While the averages of the GMM-AR statistics based on the uncentered weighting matrix ($\widehat{AR}$) approach the large sample mean of a $\chi^2(m)$-distributed random variable, their 95% percentiles are smaller than the corresponding asymptotic values (displayed below Table 2). Therefore, the uncentered GMM-AR statistic becomes more and more conservative with respect to the actual size. In very small samples the actual size is around 0.002 although the nominal significance level is set to 0.05.

On the other hand, the GMM-AR test based on a centered weighting matrix ($\widetilde{AR}$) and a $\chi^2(m)$ approximation severely overrejects. $\widetilde{AR}$ does not contain a degrees-of-freedom correction and therefore its actual size deteriorates when the number of moment conditions becomes large relative to the sample size—being even larger than the rejection frequency of 2SLS in very small samples, which here ranges from 0.094 to 0.668. In line with the well-known properties of GMM, the rejection frequencies of 2SLS (using Wald tests) increase with the number of moment conditions. The average bias of 2SLS in our simulations ranges from 0.201 to 0.248 approaching its theoretical value of 0.250 when the number of moment conditions is large (see Chao and Swanson, 2005).

Our proposed GMM-AR statistic ($\widetilde{AR}_{df}$), which uses the same degrees-of-freedom correction as the IV-AR, is on average slightly larger than the large sample mean of a $\chi^2(m)$-distributed random variable but its 95% percentile is very close to the asymptotic value. Reflecting this, the actual size of our proposed GMM-AR test is quite close to the nominal size—even in samples where the number of instruments is large relative to the number of observations (see upper panel of Table 2). The results are very similar in our nonlinear specification (see Table 3).

4 CONCLUSION

The finite sample properties of the GMM-AR test statistic based on the centered or uncentered weighting matrix follow the behavior of basic Wald or Lagrange multiplier test statistics without degrees-of-freedom correction. Analog to the poor performance of Wald tests without a degrees-of-freedom correction, the GMM-AR statistic based on the centered weighting matrix is severely size distorted in small samples. We propose a modification for the GMM-AR test based on the centered weighting matrix, which has better finite sample properties (almost no size distortions) than the usually used GMM-AR test statistics. Therefore, it should be the preferred definition of the AR statistic in a GMM setup.

For the special case of a linear homoskedastic model, we show that the usual definition of the AR test, which uses a centered weighting matrix, is asymptotically F-distributed—even when the errors are non-normally distributed. This result is new in the literature. Using a standard link between the F-distribution and the Beta distribution, we show that centering does actually not matter in the linear homoskedastic model.

A generalization of these results to the related J-test is an interesting field for future research. A paper that partly goes in this direction is Hayakawa (2016).

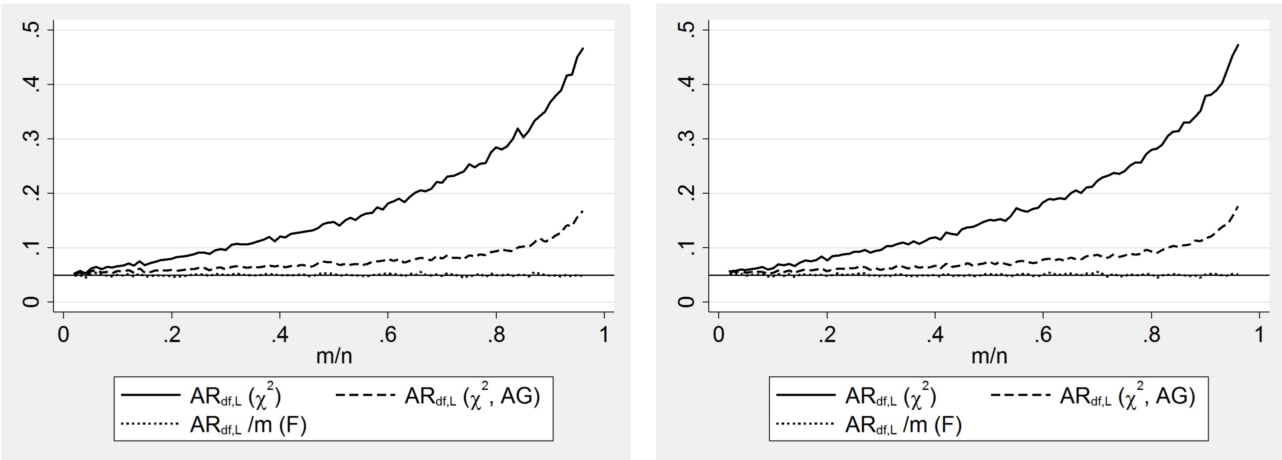
Table 1: Definition of the weighting matrix

Centered

Hansen et al. (1996); Stock and Wright (2000); Stock et al. (2002); Kleibergen (2005); Antoine *et al.* (2007); Kleibergen and Mavroeidis (2009); Caner (2010); Li and Xiao (2012)

Uncentered

Donald and Newey (2000); Donald *et al.* (2003); Newey and Smith (2004); Bond and Windmeijer (2005); Guggenberger and Smith (2005); Han and Phillips (2006); Antoine *et al.* (2007); Guggenberger (2008); Windmeijer (2008); Newey and Windmeijer (2009a); Wright (2010); Hausman et al. (2011); Caner and Yildiz (2012); Caner (2014)



(a) Cauchy errors

(b) $\chi^2(2)$ errors

Figure 1: Actual size for the IV-AR test ($\widetilde{AR}_{df,L}$).
(Note: AG denotes the modification proposed in Anatolyev and Gospodinov, 2011)

Table 2: Simulation results for GMM-AR test (linear model).

	$\widehat{AR}$			$\widetilde{AR}$			$\widetilde{AR}_{df}$		
m	mean	p95	RF	mean	p95	RF	mean	p95	RF
				$n = 100$					
3	2.98	7.45	0.042	3.13	8.05	0.055	3.04	7.80	0.050
5	4.98	10.41	0.038	5.35	11.62	0.062	5.08	11.04	0.049
10	9.96	17.00	0.029	11.28	20.49	0.091	10.15	18.44	0.052
20	20.05	28.74	0.015	25.58	40.32	0.216	20.46	32.26	0.061
30	30.04	39.31	0.007	43.84	64.78	0.465	30.69	45.35	0.067
40	40.00	49.00	0.002	68.12	96.07	0.776	40.87	57.64	0.070
				$n = 1000$					
3	3.06	8.04	0.054	3.07	8.11	0.056	3.06	8.08	0.055
5	5.03	11.02	0.049	5.07	11.14	0.052	5.04	11.08	0.050
10	10.02	18.23	0.049	10.14	18.56	0.054	10.04	18.38	0.051
20	19.95	30.89	0.044	20.39	31.87	0.055	19.99	31.23	0.048
30	30.18	43.18	0.043	31.18	45.13	0.070	30.24	43.77	0.050
40	39.90	54.54	0.039	41.63	57.69	0.072	39.97	55.38	0.047

$\rho = 0.5$; $F = 1$; 10,000 replications. Rejection frequencies for $H_0 : \beta_0 = 0$.
Nominal significance level 5%. The asymptotic critical values are $\chi^2(3) = 7.81$, $\chi^2(5) = 11.07$, $\chi^2(10) = 18.31$, $\chi^2(20) = 31.41$, $\chi^2(30) = 43.77$, $\chi^2(40) = 55.76$.

Table 3: Simulation results for GMM-AR test (nonlinear model).

	$\widehat{AR}$			$\widetilde{AR}$			$\widetilde{AR}_{df}$		
m	mean	p95	RF	mean	p95	RF	mean	p95	RF
3	2.95	7.39	0.041	3.10	7.98	0.053	3.01	7.74	0.048
5	4.93	10.29	0.035	5.29	11.47	0.059	5.02	10.90	0.046
10	9.97	17.08	0.031	11.29	20.60	0.091	10.16	18.54	0.053
20	19.98	28.74	0.015	25.47	40.33	0.214	20.38	32.26	0.059
30	30.00	39.21	0.007	43.75	64.50	0.464	30.62	45.15	0.065
40	39.91	48.78	0.001	67.85	95.23	0.777	40.71	57.14	0.065

$n = 100$; $\rho = 0.5$; $F = 1$; 5,000 replications. Rejection frequencies for $H_0 : \beta_0 = 1$.
Nominal significance level 5%. The asymptotic critical values are $\chi^2(3) = 7.81$, $\chi^2(5) = 11.07$, $\chi^2(10) = 18.31$, $\chi^2(20) = 31.41$, $\chi^2(30) = 43.77$, $\chi^2(40) = 55.76$.

5 References

Anatolyev, S. (2012). Inference in regression models with many regressors. *Journal of Econometrics*, *170*(2), 368–382.

Anatolyev, S., & Gospodinov, N. (2011). Specification testing in models with many instruments. *Econometric Theory*, *27*(02), 427–441.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. New York: Wiley.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63.

Antoine, B., Bonnal, H., & Renault, E. (2007). On the efficient use of the informational content of estimating equations: implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, *138*(2), 461 –487.

Bekker, P., & Kleibergen, F. (2003). Finite-sample instrumental variables inference using an asymptotically pivotal statistic. *Econometric Theory*, *19*(5), 744–753.

Berndt, E. R., & Savin, N. E. (1977). Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica*, *45*(5), 1263–1277.

Bond, S., & Windmeijer, F. (2005). Reliable inference for gmm estimators? finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews*, *24*(1), 1–37.

Breusch, T. S. (1979). Conflict among criteria for testing hypotheses: extensions and comments. *Econometrica*, *47*(1), 203–207.

Caner, M. (2010). Testing, estimation in GMM and CUE with nearly-weak identification. *Econometric Reviews*, *29*(3), 330–363.

Caner, M. (2014). Near exogeneity and weak identification in generalized empirical likelihood estimators: many moment asymptotics. *Journal of Econometrics*.

Caner, M., & Yildiz, N. (2012). CUE with many weak instruments and nearly singular design. *Journal of Econometrics*, *170*(2), 422–441.

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

Donald, S. G., Imbens, G. W., & Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, *117*(1), 55–93.

Donald, S. G., & Newey, W. K. (2000). A jackknife interpretation of the continuous updating estimator. *Economics Letters*, *67*(3), 239–243.

Evans, G. B. A., & Savin, N. E. (1982). Conflict among the criteria revisited; the W, LR and LM tests. *Econometrica*, *50*(3), 737–748.

Guggenberger, P. (2008). Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator. *Econometric Reviews*, *27*(4-6), 526–541.

Guggenberger, P., & Smith, R. J. (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory*, *21*(04), 667–709.

Hall, A. R. (2000). Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica*, *68*(6), 1517–1527.

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

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics*, *14*(3), 262–280.

Hausman, J., Lewis, R., Menzel, K., & Newey, W. (2011). Properties of the CUE estimator and a modification with moments. *Journal of Econometrics*, *165*(1), 45–57.

Hayakawa, K. (2016). On the effect of weighting matrix in GMM specification test. *Journal of Statistical Planning and Inference*, *178*, 84–98.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, *73*(4), 1103–1123.

Kleibergen, F., & Mavroeidis, S. (2009). Weak instrument robust tests in GMM and the new keynesian phillips curve. *Journal of Business & Economic Statistics*, *27*(3), 293–311.

Li, H., & Xiao, Z. (2012). Weak instrument inference in the presence of parameter instability. *The Econometrics Journal*, *15*(3), 395–419.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777–787.

Newey, W. K., & Windmeijer, F. (2009b). Supplement to "generalized method of moments with many weak moment conditions". *Econometrica Supplemental Material*, *77*(3).

Newey, W. K., & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, *72*(1), 219–255.

Newey, W. K., & Windmeijer, F. (2009a). Generalized method of moments with many weak moment conditions. *Econometrica*, *77*(3), 687–719.

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), 1055–1096.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529.

Windmeijer, F. (2008). GMM for panel data count models. In *The econometrics of panel data* (pp. 603–624). Springer.

Wright, J. H. (2010). Testing the adequacy of conventional asymptotics in GMM. *Econometrics Journal*, *13*(2), 205–217.