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

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

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

[b]Economics and Research Division, De Nederlandsche Bank, The Netherlands

[c]Department of Economics, University of Munich, Germany

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

November 20, 2018

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. Using an asymptotic expansion, we point to a missing degrees-of-freedom correction in the centered version of the GMM-AR test, which implicitly incorporates an Edgeworth correction. We further work out the relation between the centered and uncentered AR test in the simplified setting of a homoskedastic linear model. In a series of Monte-Carlo simulations we illustrate the validity of our arguments in finite samples.

Key Words: Many instruments, Anderson and Rubin test, Instrumental variables, GMM

*farbmacher@econ.lmu.de

1 Introduction

Most studies nowadays use uncentered (as opposed to centered) moment conditions to form the weighting matrix of the Anderson and Rubin (1949) test statistic (see Table 1). When the number of moment conditions is fixed, this choice does not matter asymptotically. However, in finite samples it affects inference. We discuss a degrees-of-freedom correction which markedly improves the finite sample properties of the centered GMM-AR test. 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 substantiate the degrees-of-freedom correction. Moreover, we use an asymptotic expansion of the GMM-AR test to obtain an Edgeworth approximation of its finite sample distribution. It turns out that our modified GMM-AR statistic implicitly incorporates an Edgeworth correction and therefore follows a χ^2 distribution more closely than does the centered AR statistic without degrees-of-freedom correction or the uncentered AR statistic. In this part of the study we focus on a setup in which the number of moments (m) is fixed as the sample size (n) grows.

In the homoskedastic linear model we, however, consider a setup in which the number of moments grows at the same rate as the sample size such that $m/n \to \mu$ with $0 \leq \mu < 1$ in the limit. In this setup our study is most closely related to 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 centered IV-AR statistic is asymptotically F-distributed even with non-normal errors. Moreover, we show that using a centered or uncentered definition of the weighting matrix does not matter for inference in their setting.

The assumption in Anatolyev and Gospodinov (2011) that the diagonal elements of the projection matrix associated with the instruments do not exhibit variation asymptotically is crucial when we study the homoskedastic linear model. This assumption occurs under certain conditions on the instruments. Our derived results thus allow researcher to shift distributional assumptions about the error term (normality), which are not verifiable, to distributional considerations about the instrumental variables, which can in principle be verified. Anatolyev and Yaskov (2017) discuss several settings under which this assumption does or does not hold. For instance, while it holds with Gaussian instruments, independent instruments drawn from different distributions, and indicator instruments with equal group sizes, it does not hold when the indicator instruments have unequal group sizes or for dummy instruments with disjoint supports. The latter case comprises the famous study of Angrist and Krueger (1991), which uses a set of dummy variables indicating quarter-of-birth and year-of-birth. On the other hand, in recent studies economists used genetic information as instruments (for example, Ding *et al.* 2009, van Hinke *et al.* 2013, Dixon *et al.* 2016, Willage 2018). These instruments are independent, which follows from the theoretical considerations behind Mendelian randomizaton (Davey Smith and Ebrahim, 2003). Thus we expect the assumption about the diagonal elements of the projection matrix to hold in these applications. In a series of Monte-Carlo simulations we show how our theoretical results perform under these different setups.

We obtain a more accurate understanding of the extent to which the centering of the weighting matrix influences the size of the Anderson and Rubin test through Monte Carlo simulations. As long as the number of moment conditions (m) is small compared to the number of observations (n), the difference between a centered and uncentered definition of the weighting matrix is negligible but in applications where m/n is considerable, the choice between these two definitions is essential. We see substantial differences in the actual size of asymptotically (if m is fixed) equivalent test statistics, which may point to conflicting inferences in practice.

In the next section we discuss the different definitions of the AR test. In Section 3 and 4 we provide theoretical results for the GMM-AR and the IV-AR test, respectively. Section 5 shows Monte Carlo simulations results to substantiate our theoretical findings. Section 6 concludes.

2 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 $H_0 : \beta = \beta_0$ 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

$$\widetilde{AR}_{df,L} = \frac{u'P_Z u}{u'M_Z u/(n-m)}\,, \tag{2}$$

where $P_Z = Z(Z'Z)^{-1}Z'$, $M_Z = I_n - P_Z$ and I_n the identity matrix of dimension n. 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

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

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

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

where $\widehat{g}(\beta) = n^{-1}\sum_{i=1}^{n} g_i(\beta)$. An unbiased estimator of the inverse of the covariance matrix of the moments needs a degrees-of-freedom correction, which has not been used in the literature so far. Therefore, we also consider a centered weighting matrix with degrees-of-freedom correction

$$\widetilde{\Omega}_{df}(\beta) = (n-m)^{-1}\sum_{i=1}^{n}[g_i(\beta) - \widehat{g}(\beta)][g_i(\beta) - \widehat{g}(\beta)]'\,. \tag{5}$$

This is also in line with the standard definition of the IV-AR in the literature (see equation 2), which also contains a degrees-of-freedom correction. The corresponding GMM-AR test statistics are defined as

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

The degrees-of-freedom correction clearly matters in finite samples. To the best of our knowledge, $\widetilde{AR}_{df}$ has not been used in the literature so far.

3 GMM Inference

In the following we give an interpretation for the GMM-AR test, which illustrates the finite-sample differences between all three definitions (see equation 6) and provides guidance on how to choose between them along with theoretical results. We assume throughout in this section that the following conditions hold:

Assumption GMM1. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\beta_0) \xrightarrow{d} N(0,\Omega(\beta_0))$.

Assumption GMM2. *As* $n\to\infty$, m *is finite.*

Assumption GMM3. $\widehat{\Omega}(\beta_0)^{-1}, \widetilde{\Omega}(\beta_0)^{-1}, \widetilde{\Omega}_{df}(\beta_0)^{-1} \xrightarrow{p} \Omega(\beta_0)^{-1}$, *with* $\Omega(\beta_0) = E\left[g_i(\beta_0)g_i(\beta_0)'\right]$.

We emphasize that in this section we consider a setup where $m/n \to 0$. Moreover, all three weighting matrices estimate the inverse of the covariance matrix of $g_i(\beta_0)$ consistently. Note that the true covariance matrix of $g_i(\beta_0)$ can be interpreted as both being uncentered and centered using $E[g_i(\beta_0)] = 0$.

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 auxiliary regression

$$i = G\gamma + e \tag{7}$$

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

$$\begin{aligned} i'i &= \widehat{i'i} + \widehat{e}'\widehat{e} \\ \Leftrightarrow 1 &= \widehat{i'i}/i'i + \widehat{e}'\widehat{e}/i'i \end{aligned} \tag{8}$$

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

$$\begin{aligned} n\,R^2 = \widehat{i'i} &= i'G(G'G)^{-1}G'G(G'G)^{-1}G'i \\ &= n\,i'G/n(G'G/n)^{-1}G'i/n \\ &= n\,\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}\widehat{g}(\beta) \\ &= n\,\widehat{Q} \end{aligned} \tag{9}$$

Thus, $\widehat{Q}$, which is the criterion function of the continuous updating estimator (CUE), is also the coefficient of determination (R^2) in Equation (7). The uncentered GMM-AR test, which is 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

$$\begin{aligned} W &= ni'G(G'G)^{-1}(G'G/n)(G'G)^{-1}G'i/\widehat{\sigma}_e^2 \\ &= n\widehat{g}(\beta)'\widehat{\Omega}(\beta)^{-1}\widehat{\Omega}(\beta)\widehat{\Omega}(\beta)^{-1}\widehat{g}(\beta)/\widehat{\sigma}_e^2 \\ &= \widehat{AR}/[1 - \widehat{AR}/n] = \widetilde{AR}. \end{aligned} \tag{10}$$

A similar link between the centered and uncentered criterion function of the CUE has already been established in Newey and Smith (2004) and Antoine *et al.* (2007). 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). Proposition 1 further quantifies the expected difference between centered and uncentered AR statistics.

Proposition 1. *Under Assumptions GMM1-GMM3, we have under* H_0

$$E\left(\widetilde{AR} - \widehat{AR}\right) = \frac{m^2 + 2m}{n} + \frac{m^3 + 6m^2 + 8m}{n^2} + o\left(n^{-2}\right).$$

Proof. See Appendix A.1. Q.E.D.

The first and second terms in the approximation are of order $O(n^{-1})$ and $O(n^{-2})$, respectively. For small m the second term in the approximation is relatively small compared to the first term. For example, when $n = 100$ and $m = 3$ we have $\frac{m^2+2m}{n} = 0.15$ and $\frac{m^3+6m^2+8m}{n^2} = 0.01$. For larger m the second term matters and improves on the approximation. For $m = 20$ we have $\frac{m^2+2m}{n} = 4.4$ and $\frac{m^3+6m^2+8m}{n^2} = 1.06$, which is almost 25% of the first term. We will see in the Monte Carlo simulations in Section 5 that the approximation of Proposition 1 is quite accurate for moderate m and closely matches the differences in mean of the centered and uncentered AR statistics.

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 resulting AR statistic would thus be

$$\widetilde{AR}_{df} = \frac{n-m}{n}\widetilde{AR}, \tag{11}$$

which equals our definition in 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.

Corollary 1. *Under the assumptions of Proposition 1, we have under H_0*

$$E\left(\widetilde{AR}_{df} - \widehat{AR}\right) = \frac{2m}{n} + \frac{4m^2 + 8m}{n^2} + o\left(n^{-2}\right).$$

Proof. See Appendix A.1. Q.E.D.

From Corollary 1 we infer that, compared to the difference between $\widetilde{AR}$ and $\widehat{AR}$ statistics, the degrees-of-freedom correction removes the $\frac{m^2}{n}$ and also $\frac{m^3}{n^2}$ bias terms. Especially when m is not very small compared to n this is going to matter for inference.

Finally, we have three asymptotically (if m is fixed) 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 finite-sample performance of the discussed tests from the following result derived by an asymptotic expansion. For this result we make the additional assumption:

Assumption GMM4. *$g_i(\beta_0)$ is symmetrically distributed.*

We then have the following:

Proposition 2. *Under Assumptions GMM1-GMM4, we have under H_0*

$$\Pr\left[\widetilde{AR} \leq x\right] = \Pr_{\chi^2_m}\left[x - \frac{m+2}{n}x\right] + o\left(n^{-1}\right),$$
$$\Pr\left[\frac{n-m-2}{n}\widetilde{AR} \leq x\right] = \Pr_{\chi^2_m}[x] + o\left(n^{-1}\right),$$
$$\Pr\left[\widehat{AR} \leq \frac{nx}{n-m-2+x}\right] = \Pr_{\chi^2_m}[x] + o\left(n^{-1}\right),$$

where $\Pr_{\chi^2_m}[x]$ is the distribution of a χ^2_m distributed random variable evaluated at x.

Proof. See Appendix A.2. Q.E.D.

Proposition 2 shows that the centered AR statistic is size distorted up to order $O(n^{-1})$ when using asymptotic critical values from the χ^2_m distribution. More importantly, the size distortion is increasing in the ratio m/n. Moreover, Proposition 2 shows that a degrees of freedom correction, which is straightforward to implement, solves the bias in the centered AR statistic. Our proposed degrees of freedom correction $\frac{n-m}{n}$ is slightly different, but will be equally effective even for rather small n. These theoretical results justify the use of a degrees-of-freedom correction of the centered AR as an implicit Edgeworth correction. This explains the superior behavior of $\widetilde{AR}_{df}$ reported in the Monte Carlo simulations of Section 5. Finally, Proposition 2 also shows that Edgeworth critical values can be constructed for the uncentered AR statistic.

We used the assumption of a symmetrical distribution to show the effectiveness of the proposed degrees of freedom correction. In principle Proposition 2 can be generalized to accommodate skewed distributions, but it will depend on the uneven moments of the distribution of the sample moments.

4 IV Inference

In this section we further investigate the important case of a linear model. We can interpret the IV-AR test as an F-test of joint significance of γ from the following auxiliary regression

$$u = Z\gamma + w\,. \tag{12}$$

In this section we assume throughout that the following conditions hold:

Assumption IV1. *The errors u_i satisfy $E(u|Z) = 0, E(uu'|Z) = \sigma_2 I_n$, and $E(|u_i|^4) < \infty$.*

Assumption IV2. *As $n \to \infty$, $m/n = \mu$, where $0 < \mu < 1$.*

Assumption IV3. *Under the asymptotics of Assumption IV2, $n^{-1}\sum_{i=1}^{n} |z_i'(Z'Z)^{-1}z_i - \mu| \to 0$.*

Assumption IV1 imposes homoskedasticity and a finite fourth moment of the structural errors. Assumption IV2 states that the number of instruments is a nontrivial fraction of the sample size and refers to the many instruments asymptotic framework of Bekker (1994). Hence, we now consider a setup in which the number of instruments can grow at the same rate as the sample size. This relaxes the framework considered in the previous section. However, we have to impose Assumption IV3, which requires that (almost) all diagonal elements of the projection matrix P_Z converge to μ. Anatolyev and Yaskov (2017) discuss several settings under which this assumption does or does not hold. In the Monte Carlo simulations we give further insights on the relevance of Assumption IV3.

Testing $H_0 : \gamma = 0$, we have

$$F = \frac{\widehat{\gamma}' Z' Z \widehat{\gamma}}{\widehat{\sigma}_w^2} / m = \frac{u' P_Z u}{u' M_Z u / (n-m)} / m = \frac{\widetilde{AR}_{df,L}}{m} . \qquad (13)$$

From this relation we derive the following propositions.

Proposition 3. *Under Assumptions IV1-IV3, we have under* H_0

$$\sqrt{m}\left(\frac{\widetilde{AR}_{df,L}}{m} - 1\right) \xrightarrow{d} N\left(0, \frac{2}{1-\mu}\right)$$
$$\sqrt{m}\left(\frac{\widehat{AR}_L}{n} - \frac{m}{n}\right) \xrightarrow{d} N\left(0, 2\mu^2(1-\mu)\right)$$

where $\widehat{AR}_L = n\frac{u' P_Z u}{u'u}$ *is based on the restricted variance estimate in (12).*

Proof. See Appendix A.3. Q.E.D.

Proposition 4. *Under Assumptions IV1-IV3, we have under* H_0

$$Pr\left(\frac{\widetilde{AR}_{df,L}}{m} > q_\alpha^{F(m,n-m)}\right) \to \alpha$$
$$Pr\left(\frac{\widehat{AR}_L}{n} > q_\alpha^{B(\frac{m}{2},\frac{n-m}{2})}\right) \to \alpha$$

Proof. See Appendix A.4. Q.E.D.

This result implies that the definition of the weighting matrix does not matter for the IV-AR test under Assumptions IV1-IV3 as long as we use the appropriate asymptotic approximation. Anatolyev and Gospodinov (2011) also show that $\widetilde{AR}_{df,L} \sim N(m, 2m/(1-\mu))$. Expectation and variance derived from the F and Beta distribution are asymptotically equivalent to the normal approximation derived in Proposition 3. However, the normal approximation performs poorly at both ends of $\mu \in [0,1]$. Anatolyev and Gospodinov (2011) successfully solve this issue if μ is small by using a χ^2 approximation with corrected critical values. If μ 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 μ as it can change its skewness over the parameter space of μ. The Beta-distribution is right-skewed, just like the χ^2-distribution, when μ is small and is left-skewed when μ is close to 1.

5 Simulation Results

We conducted a series of Monte Carlo simulations for both linear and nonlinear models, and with normal and non-normal errors. To shed more light on the relevance of Assumption IV3 for the IV-AR, we use different sets of instrumental variables in the linear model. The design of the linear model is

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

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 (F^∞) fixed at 1, which implies that the set of instruments is equally weak with varying number of instruments (i.e., $CP = F^\infty * m$). We compare three different sets of instrumental variables: Gaussian instruments, independent categorical instruments[1], and mutually exclusive dummy instruments. The choices of the instruments are motivated by the discussion in Anatolyev and Yaskov (2017).

Figure 1 depicts the rejection frequencies for the IV-AR ($\widetilde{AR}_{df,L}$) under various asymptotic approximations when the instruments are normally distributed. 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

[1] The values of the categorical instruments are $0, 1, 2$ with varying probabilities. This is the setup used in Mendelian randomization studies.

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. Using a centered or uncentered weighting matrix under Assumptions IV1-IV3 is thus irrelevant if we use the appropriate asymptotic distribution to obtain the critical values.

Figure 2 illustrates the importance of Assumption IV3 when it comes to the approximation derived in Proposition 3 and 4. We defined u_2 as standard Cauchy in these simulations. From the theoretical results in Anatolyev and Yaskov (2017) we expect Assumption IV3 to hold for the Gaussian instruments (Figure 1) and the independent categorical instruments, which is verified by the simulation results depicted in Figure 2a. If we know that Assumption IV3 holds, our approximation controls the nominal size even when the number of instruments is large compared to the sample size. On the other hand, if Assumption IV3 does not hold, all discussed approximations fail once the number of instruments becomes too large relative to the sample size (see Figure 2b for mutually exclusive dummy instruments).

In the following we analyze the performance of the degrees-of-freedom correction discussed for the GMM-AR statistic. We start with a discussion of the linear model with Gaussian errors and instruments and then discuss the simulation results from a nonlinear specification as well. Both settings are such that Assumptions GMM1-GMM4 hold. Therefore, we expect our correction to be effective. Table 2 shows the simulation results of the GMM-AR statistics for the linear model.[2] Note that the approximation in Proposition 1 is quite accurate for moderate m/n and matches the differences in mean as reported in Table 2. When $n = 100$ and $m = 10$, for example, we have $\frac{m^2+2m}{n} = 1.2$ and $\frac{m^3+6m^2+8m}{n^2} = 0.168$, and the observed mean difference is 1.32. For $m = 20$ this is $\frac{m^2+2m}{n} = 4.4$ and $\frac{m^3+6m^2+8m}{n^2} = 1.056$, and the observed mean difference is 5.53. 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 the small sample setting ($n = 100$) with many instruments ($m = 20$) the actual size is around 0.015 although the nominal significance level is set to 0.05. On the other hand,

[2] The average bias of 2SLS in our simulations ranges from 0.201 to 0.245 approaching its theoretical value of 0.250 when the number of moment conditions is large (see Chao and Swanson, 2005).

the GMM-AR test based on a centered weighting matrix ($\widetilde{AR}$) severely overrejects (actual size around 0.216 for $m = 20$ and $n = 100$). $\widetilde{AR}$ 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.

The GMM-AR statistic ($\widetilde{AR}_{df}$), which uses a degrees-of-freedom correction, 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 a test based on $\widetilde{AR}_{df}$ is quite close to the nominal size—even in samples where the number of instruments ($m = 20$) is relatively large compared to the number of observations ($n = 100$), the actual size is around 0.061.

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. The results are very similar in our nonlinear specification (see Table 3). The uncentered GMM-AR statistic tends to underreject and the centered GMM-AR statistic without degrees-of-freedom correction tends to overreject. This pattern becomes more distinct the larger the number of moment conditions is with respect to the sample size. The GMM-AR statistic ($\widetilde{AR}_{df}$), which uses a degrees-of-freedom correction, performs very well in this setting with actual size close to the nominal level.

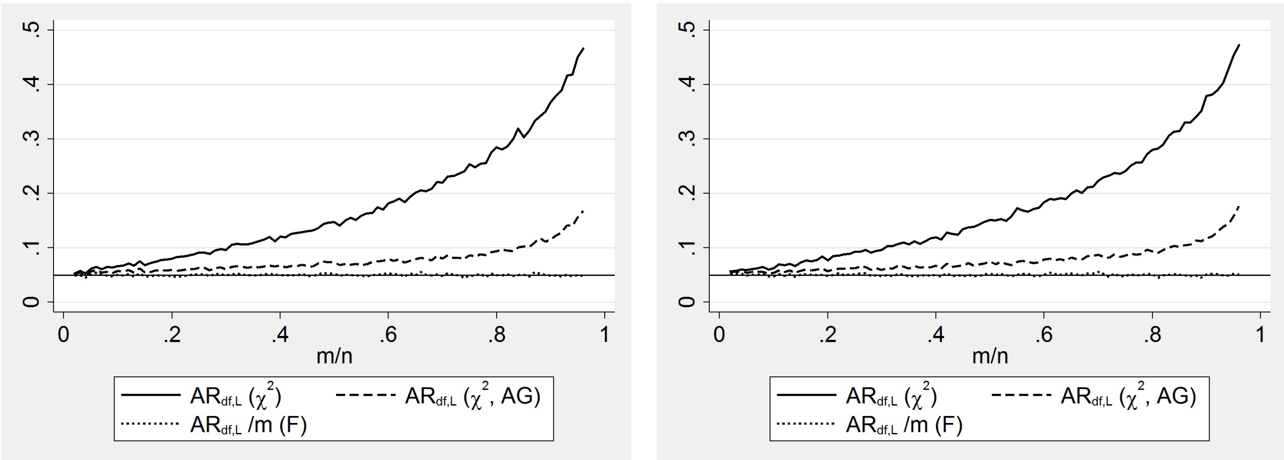
6 Conclusion

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. 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 degrees-of-freedom correction for the centered GMM-AR test, which has better finite sample properties than the usually used GMM-AR test statistics. We substantiate this modification using an asymptotic expansion of the AR statistic.

We further work out the relation between the centered and uncentered AR test in the simplified setting of a homoskedastic linear model. The assumption that the diagonal elements of the projection matrix associated with the instruments do not exhibit variation asymptotically is crucial in this part of our study. This assumption occurs under certain conditions on the instruments, which thus allows us to shift distributional assumptions about the error term (normality), which are not verifiable, to distributional considerations about the instrumental variables, which can in principle be verified. For the 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, while the uncentered version is asymptotically Beta-distributed. Both approximations work equally well which suggests that centering does not matter in this setup as long as we use the appropriate asymptotic approximation. 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

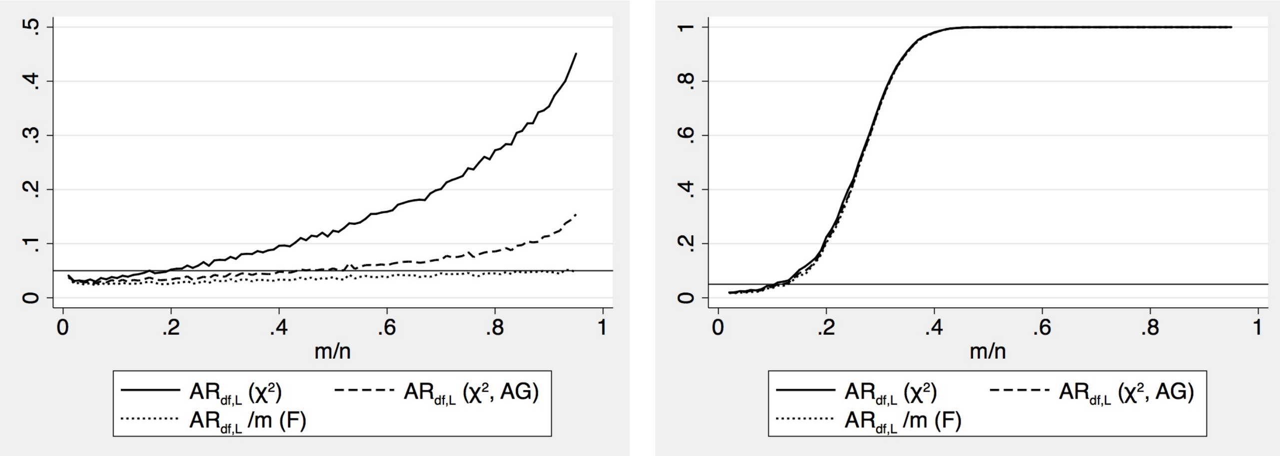
Centered	
	Hansen et al. (1996); Stock and Wright (2000); Stock et al. (2002); Kleibergen (2005); Antoine *et al.* (2007); Kleibergen and Mavroeidis (2009); Caner (2010); Li and Xiao (2012)
Uncentered	
Donald and Newey (2000); Donald *et al.* (2003); Newey and Smith (2004); Bond and Windmeijer (2005); 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 ($\widetilde{AR}_{df,L}$) with Gaussian instruments. (Note: AG denotes the modification proposed in Anatolyev and Gospodinov, 2011)



(a) Independent categorical IVs

(b) Mutually exclusive dummy IVs

Figure 2: Actual size for the IV-AR test ($\widetilde{AR}_{df,L}$) with Cauchy errors.
(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
					$n = 1000$				
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

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

7 References

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

Anatolyev, S., & Yaskov, P. (2017). Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory*, *33*(3), 717–738.

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.

Bao, Y., & Ullah, A. (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference*, *140*(5), 1193–1205.

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

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.

Davey Smith, G., & Ebrahim, S. (2003). Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22.

Ding, W., Lehrer, S. F., Rosenquist, J. N., & Audrain-McGovern, J. (2009). The impact of poor health on academic performance: new evidence using genetic markers. *Journal of Health Economics*, *28*(3), 578–597.

Dixon, P., Davey Smith, G., von Hinke, S., Davies, N., & W., H. (2016). Estimating marginal healthcare costs using genetic variants as instrumental variables: Mendelian Randomization in economic evaluation. *PharmacoEconomics*, *34*(11), 1075–1086.

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. (2018). Expansions of GMM statistics and the bootstrap. Econometrics Discussion Paper 2006/09, University of Amsterdam.

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. (n.d.). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, (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.

Phillips, P. C. B., & Park, J. Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica*, *56*(5), 1065–1083.

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.

von Hinke, S., Davey Smith, G., Lawlor, D. A., Propper, C., & Windmeijer, F. (2013). Child height, health and human capital: evidence using genetic markers. *European Economic Review*, *57*, 1–22.

Willage, B. (2018). The effecto of weight on mental health: new evidence using genetic IVs. *Journal of Health Economics*, *57*, 113–130.

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.

A Appendix

A.1 Proof of Proposition 1

Proof. We suppress the dependence of $\widetilde{AR}$, $\widehat{AR}$, Ω, $\widehat{\Omega}$ and $\widehat{g}$ on β_0. Note that $\frac{\widehat{AR}}{n}$ is $O_p(n^{-1})$. Under standard (finite m) asymptotics we can therefore expand (10) as follows:

$$\begin{aligned}
\widetilde{AR} &= \widehat{AR}\left(1 + \frac{\widehat{AR}}{n} + \left(\frac{\widehat{AR}}{n}\right)^2 + \ldots\right) \\
&= \widehat{AR} + \frac{1}{n}\widehat{AR}^2 + \frac{1}{n^2}\widehat{AR}^3 + o_p\left(n^{-2}\right) \\
&= \widehat{AR} + \frac{1}{n}\left(n\widehat{g}'\Omega^{-1}\widehat{g}\right)^2 + \frac{1}{n^2}\left(n\widehat{g}'\Omega^{-1}\widehat{g}\right)^3 + o_p\left(n^{-2}\right),
\end{aligned}$$

where the last line follows from the fact that $\widehat{\Omega} = \Omega + O_p(n^{-1/2})$, hence

$$\widehat{AR} = n\widehat{g}'\Omega^{-1}\widehat{g} + o_p(1).$$

Analyzing the difference $\widetilde{AR} - \widehat{AR}$, taking the expected value of the expansion we get:

$$E\left(\widetilde{AR} - \widehat{AR}\right) = \frac{1}{n}E\left(\left(n\widehat{g}'\Omega^{-1}\widehat{g}\right)^2\right) + \frac{1}{n^2}E\left(\left(n\widehat{g}'\Omega^{-1}\widehat{g}\right)^3\right) + o\left(n^{-2}\right),$$

Assumption GMM1 implies that $\sqrt{n}\widehat{g}$ is asymptotically $N(0,\Omega)$ distributed, hence $\sqrt{n}\Omega^{-1/2}\widehat{g}$ is approximately standard normal. We can therefore define

$$\sqrt{n}\Omega^{-1/2}\widehat{g} = \psi + O_p\left(n^{-1/2}\right),$$

with ψ a multivariate standard normal, hence $n\widehat{g}'\Omega^{-1}\widehat{g} = \psi'\psi + O_p\left(n^{-1/2}\right)$. We know from results on quadratic forms in normally distributed random variables (see e.g. Bao and Ullah, 2010) that:

$$\begin{aligned}
E\left((\psi'\psi)^2\right) &= m^2 + 2m, \\
E\left((\psi'\psi)^3\right) &= m^3 + 6m^2 + 8m.
\end{aligned}$$

Substituting these expressions we have:

$$E\left(\widetilde{AR} - \widehat{AR}\right) = \frac{m^2 + 2m}{n} + \frac{m^3 + 6m^2 + 8m}{n^2} + o\left(n^{-2}\right).$$

Q.E.D.

Proof of Corollary 1.

Proof. We have:

$$\begin{aligned}
\widetilde{AR}_{df} &= \frac{n-m}{n}\widetilde{AR} \\
&= \left(1 - \frac{m}{n}\right)\widehat{AR}\left(1 + \frac{\widehat{AR}}{n} + \left(\frac{\widehat{AR}}{n}\right)^2 + \ldots\right),
\end{aligned}$$

hence we have for the difference:

$$\begin{aligned}
\widetilde{AR}_{df} - \widehat{AR} &= -\frac{m}{n}\widehat{AR} + \left(1 - \frac{m}{n}\right)\frac{\widehat{AR}^2}{n} + \left(1 - \frac{m}{n}\right)\frac{\widehat{AR}^3}{n^2} + ... \\
&= -\frac{m}{n}\widehat{AR} + \left(1 - \frac{m}{n}\right)\frac{\widehat{AR}^2}{n} + \frac{\widehat{AR}^3}{n^2} + o_p\left(n^{-2}\right).
\end{aligned}$$

Taking expectations and exploiting earlier results from Proposition 1 we have:

$$\begin{aligned}
E\left(\widetilde{AR}_{df} - \widehat{AR}\right) &= -\frac{m}{n}m + \left(1 - \frac{m}{n}\right)\frac{m^2 + 2m}{n} + \frac{m^3 + 6m^2 + 8m}{n^2} + o\left(n^{-2}\right) \\
&= \frac{2m}{n} + \frac{4m^2 + 8m}{n^2} + o\left(n^{-2}\right).
\end{aligned}$$

Q.E.D.

A.2 Proof of Proposition 2

Proof. We follow the approach from Kleibergen (2018) to derive an Edgeworth expansion of the centered AR statistic by replacing the covariance matrix estimator $\widetilde{\Omega}(\beta)$ in the construction of $\widetilde{AR}(\beta)$ by the following Taylor expansion around the true value β_0:

$$\begin{aligned}\widetilde{\Omega}(\beta_0)^{-1} &= \Omega(\beta_0)^{-1} - \Omega(\beta_0)^{-1}\left(\widetilde{\Omega}(\beta_0) - \Omega(\beta_0)\right)\Omega(\beta_0)^{-1} \\ &\quad + \Omega(\beta_0)^{-1}\left(\widetilde{\Omega}(\beta_0) - \Omega(\beta_0)\right)\Omega(\beta_0)^{-1}\left(\widetilde{\Omega}(\beta_0) - \Omega(\beta_0)\right)\Omega(\beta_0)^{-1} + o_p\left(n^{-1}\right).\end{aligned}$$

Substituting this in the centered AR statistic and suppressing the dependence of Ω, $\widetilde{\Omega}$ and $\widehat{g}$ on β_0, we obtain under $H_0 : \beta = \beta_0$ the following higher-order expression:

$$\begin{aligned}\widetilde{AR}(\beta_0) &= n\widehat{g}'\Omega^{-1}\widehat{g} - n\widehat{g}'\Omega^{-1}\left(\widetilde{\Omega} - \Omega\right)\Omega^{-1}\widehat{g} \\ &\quad + n\widehat{g}'\Omega^{-1}\left(\widetilde{\Omega} - \Omega\right)\Omega^{-1}\left(\widetilde{\Omega} - \Omega\right)\Omega^{-1}\widehat{g} + o_p\left(n^{-1}\right).\end{aligned}$$

We use Theorem 3 of Kleibergen (2018), which is an adaptation of Theorem 2.4 from Phillips and Park (1988), to provide the following Edgeworth approximation for the finite sample distribution of the centered AR statistic:

$$\Pr\left[\widetilde{AR}(\beta_0) \leq x\right] = \Pr{}_{\chi^2_m}\left[x - \frac{1}{n}\frac{E\left(S\right)}{m}x\right] + o\left(n^{-1}\right),$$

where $\Pr_{\chi^2_m}[x]$ is the distribution of a χ^2_m distributed random variable evaluated at x. Furthermore, the quantity $E(S)$ is defined as the sum of the $O(1)$ terms from:

$$\begin{aligned}E\left(S_1\right) &= nE\left(-n\widehat{g}'\Omega^{-1}\left(\widetilde{\Omega} - \Omega\right)\Omega^{-1}\widehat{g}\right) \\ E\left(S_2\right) &= nE\left(n\widehat{g}'\Omega^{-1}\left(\widetilde{\Omega} - \Omega\right)\Omega^{-1}\left(\widetilde{\Omega} - \Omega\right)\Omega^{-1}\widehat{g}\right).\end{aligned}$$

Kleibergen (2018) provides analytical expressions of these higher-order terms in the expansion. It is shown that:

$$E\left(S\right) = m^2 + 2m + vec\left(\Omega^{-1}\right)' E\left[\left(g_i g_i' \otimes g_i\right)\right]\Omega^{-1}E\left[\left(g_i g_i' \otimes g_i\right)\right] vec\left(\Omega^{-1}\right).$$

Under Assumption GMM4 this simplifies to:

$$E\left(S\right) = m^2 + 2m,$$

resulting in:

$$\Pr\left[\widetilde{AR} \leq x\right] = \Pr{}_{\chi^2_m}\left[x - \frac{m+2}{n}x\right] + o\left(n^{-1}\right).$$

Using this result it is not difficult to see that:

$$\begin{aligned}\Pr\left[\frac{n-m-2}{n}\widetilde{AR} \leq x\right] &= \Pr\left[\widetilde{AR} \leq \frac{n}{n-m-2}x\right] \\ &= \Pr{}_{\chi^2_m}\left[x\right] + o\left(n^{-1}\right).\end{aligned}$$

Furthermore, combining Proposition 2 with (10) we have:

$$\begin{aligned}
\Pr\left[\frac{n-m-2}{n}\frac{\widehat{AR}}{1-\frac{\widehat{AR}}{n}} \leq x\right] &= \Pr\left[\frac{n-m-2}{n}\widehat{AR} + x\frac{\widehat{AR}}{n} \leq x\right] \\
&= \Pr\left[\frac{n-m-2+x}{n}\widehat{AR} \leq x\right] \\
&= \Pr\left[\widehat{AR} \leq \frac{nx}{n-m-2+x}\right].
\end{aligned}$$

Q.E.D.

A.3 Proof of Proposition 3

Define $\widehat{\sigma}^2 = \frac{u'M_Z u}{n-m}$ and $\widetilde{\sigma}^2 = \frac{u'u}{n}$. We use the following Lemmas in the proof of Proposition 3.

Lemma 1. *Under Assumptions IV1-IV3,*

$$\frac{\sigma^2}{\widehat{\sigma}^2} - 1 = \frac{\mu}{1-\mu}\left(\frac{u'P_Z u}{m\sigma^2} - 1\right) - \frac{1}{1-\mu}\left(\frac{u'u}{n\sigma^2} - 1\right) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

Proof. See proof of Anatolyev's Theorem 2. Q.E.D.

Lemma 2. *Under Assumptions IV1-IV3,*

$$\frac{\sigma^2}{\widetilde{\sigma}^2} - 1 = 1 - \frac{u'u}{n\sigma^2}$$

Proof.

$$\begin{aligned}
\frac{\sigma^2}{\widetilde{\sigma}^2} &= \left(1 + \frac{\widetilde{\sigma}^2}{\sigma^2} - 1\right)^{-1} \\
&= 1 - \left(\frac{\widetilde{\sigma}^2}{\sigma^2} - 1\right) + o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

Q.E.D.

Proof of Proposition 3:

Proof. We have the following representations for

$$\begin{aligned}
\sqrt{m}\left(\frac{\widetilde{AR}_{df,L}}{m} - 1\right) &= \sqrt{m}\left(\frac{u'P_Z u}{m\sigma^2} - 1\right) \\
&\quad + \sqrt{m}\left(\frac{\sigma^2}{\widehat{\sigma}^2} - 1\right) + \sqrt{m}\left(\frac{\sigma^2}{\widehat{\sigma}^2} - 1\right)\left(\frac{u'P_Z u}{m\sigma^2} - 1\right) \\
&= A + o_p\left(\frac{1}{\sqrt{m}}\right)
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{m}\left(\frac{\widehat{AR}_L}{n} - \frac{m}{n}\right) &= \sqrt{m}\left(\frac{u'P_Z u}{u'u} - \frac{m}{n}\right) \\
&= \sqrt{m}\left(\frac{\sigma^2}{\widetilde{\sigma}^2}\frac{u'P_Z u}{m\sigma^2}\frac{m}{n} - \frac{m}{n}\right) \\
&= \sqrt{m}\frac{m}{n}\left(\frac{\sigma^2}{\widetilde{\sigma}^2}\frac{u'P_Z u}{m\sigma^2} - 1 + \frac{u'P_Z u}{m\sigma^2} - \frac{u'P_Z u}{m\sigma^2} + \left(\frac{\sigma^2}{\widetilde{\sigma}^2} - 1\right) - \left(\frac{\sigma^2}{\widetilde{\sigma}^2} - 1\right)\right) \\
&= \frac{m}{n}\left(\sqrt{m}\left(\frac{u'P_Z u}{m\sigma^2} - 1\right) + \sqrt{m}\left(\frac{\sigma^2}{\widetilde{\sigma}^2} - 1\right) + \sqrt{m}\left(\frac{\sigma^2}{\widetilde{\sigma}^2} - 1\right)\left(\frac{u'P_Z u}{m\sigma^2} - 1\right)\right) \\
&= B + o_p\left(\frac{1}{\sqrt{m}}\right)
\end{aligned}$$

where

$$
\begin{aligned}
A &= \frac{1}{1-\mu}\frac{1}{\sqrt{m}}\left(\frac{u'P_Z u}{\sigma^2} - m - \left(\mu\frac{u'u}{\sigma^2} - m\right)\right) \\
&= A_1 + A_2 + o_p(1)
\end{aligned}
$$

and

$$
\begin{aligned}
B &= \mu\frac{1}{\sqrt{m}}\left(\frac{u'P_Z u}{\sigma^2} - m - m\left(\frac{\widetilde{\sigma}^2}{\sigma^2} - 1\right)\right) \\
&= \mu\frac{1}{\sqrt{m}}\left(\frac{u'P_Z u}{\sigma^2} - m - \left(\mu\frac{u'u}{\sigma^2} - m\right)\right) \\
&= \mu(1-\mu)A
\end{aligned}
$$

with

$$
\begin{aligned}
A_1 &= \frac{1}{\sqrt{m}}\frac{1}{1-\mu}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i\frac{u_i^2}{\sigma^2} - z_i'(Z'Z)^{-1}z_i - \mu\frac{u_i^2}{\sigma^2} + \mu\right) \\
&= \frac{1}{\sqrt{m}}\frac{1}{1-\mu}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i - \mu\right)\left(\frac{u_i^2}{\sigma^2} - 1\right) \\
A_2 &= \frac{1}{\sqrt{m}}\frac{1}{1-\mu}\sum_{i=1}^{n} z_i'(Z'Z)^{-1}z_j\frac{u_i u_j}{\sigma^2}\,.
\end{aligned}
$$

As we test all parameters in our auxiliary regression, our A_1 and A_2 are special cases of the same quantities in Anatolyev (2012), where we set $\Xi_R = \Xi_{I_m} = (Z'Z)^{-1}$. Using Assumption IV3 and $n^{-1}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i - \mu\right)^2 \leq n^{-1}\sum_{i=1}^{n}\left|z_i'(Z'Z)^{-1}z_i - \mu\right|$, we get $Var(A_1) \to 0$. Then, using the normality result of A_2 in the proof of Anatolyev's Theorem 2, we can derive the required results from $A_2 \overset{d}{=} N(0, 2(1-\mu)^{-1})$. Q.E.D.

A.4 Proof of Proposition 4

Proof. The proof for the exact test on the centered $\widetilde{AR}_{df,L}$ can easily be obtained by interpreting it as an F-test as proposed in Equation (13) and following the proof of Anatolyev's Theorem 3.

The proof for the exact test on the uncentered $\widehat{AR}_L$ starts with

$$
B\left(m/2, (n-m)/2\right) \overset{d}{=} \frac{\epsilon_1}{\epsilon_1 + \epsilon_2}\,,
$$

where $\epsilon_1 \sim \chi^2(m)$ and $\epsilon_2 \sim \chi^2(n-m)$ are independently distributed. Then,

$$\begin{aligned}
\frac{\epsilon_1}{\epsilon_1+\epsilon_2} &= \frac{1}{1+\frac{\epsilon_2}{\epsilon_1}} = \frac{\frac{m}{n-m}}{\frac{m}{n-m}+\frac{m}{n-m}\frac{\epsilon_2}{\epsilon_1}} \\
&\overset{d}{=} \frac{m}{n-m}\left(\frac{m}{n-m}+\left(1+\sqrt{\frac{2}{n-m}}\eta_2\right)\left(1+\sqrt{\frac{2}{m}}\eta_1\right)^{-1}\right)^{-1} \\
&= \frac{m}{n-m}\left(\frac{n}{n-m}+\sqrt{\frac{2}{n-m}}\eta_2-\sqrt{\frac{2}{m}}\eta_1+o_d\left(\frac{1}{\sqrt{m}}\right)\right)^{-1} \\
&= \frac{m}{n}\left(1+\frac{n-m}{n}\sqrt{\frac{2}{n-m}}\eta_2-\frac{n-m}{n}\sqrt{\frac{2}{m}}\eta_1+o_d\left(\frac{1}{\sqrt{m}}\right)\right)^{-1} \\
&= \frac{m}{n}-\frac{m}{n}\frac{n-m}{n}\sqrt{\frac{2}{n-m}}\eta_2+\frac{m}{n}\frac{n-m}{n}\sqrt{\frac{2}{m}}\eta_1+o_d\left(\frac{1}{\sqrt{m}}\right) \\
&\overset{d}{=} \frac{m}{n}+N\left(0,2\left(\frac{m}{n}\right)^2\left(\frac{n-m}{n}\right)^2\left(\frac{1}{n-m}+\frac{1}{m}\right)\right)
\end{aligned}$$

where we use

$$\frac{m}{n-m}\frac{\chi^2(n-m)}{\chi^2(m)} \overset{d}{=} \frac{1+\sqrt{2/(n-m)}\,\eta_2}{1+\sqrt{2/m}\,\eta_1}$$

with $m\to\infty$, $n-m\to\infty$, and η_1 and η_2 being independent standard normally distributed. Therefore, we have for the quantile of the $B\left(\frac{m}{2},\frac{n-m}{2}\right)$ distribution that

$$q_\alpha^{B(m/2,(n-m)/2)} = \frac{m}{n}+\Phi^{-1}(1-\alpha)\sqrt{2\frac{m}{n}\frac{n-m}{n^2}}+o\left(\frac{1}{\sqrt{m}}\right)$$

Then

$$\begin{aligned}
Pr\left\{\frac{\widehat{AR}_L}{n} > q_\alpha^{B\left(\frac{m}{2},\frac{n-m}{2}\right)}\right\} &= Pr\left\{\frac{\widehat{AR}_L}{n} > \frac{m}{n}+\Phi^{-1}(1-\alpha)\sqrt{2\frac{m}{n}\frac{n-m}{n^2}}+o\left(\frac{1}{\sqrt{m}}\right)\right\} \\
&= Pr\left\{\frac{\sqrt{m}\left(\frac{\widehat{AR}_L}{n}-\frac{m}{n}\right)}{\sqrt{2\mu^2(1-\mu)}} > \Phi^{-1}(1-\alpha)\sqrt{\frac{2\left(\frac{m}{n}\right)^2\frac{n-m}{n}}{2\mu^2(1-\mu)}}+o(1)\right\} \\
&= 1-Pr\left\{\frac{\sqrt{m}\left(\frac{\widehat{AR}_L}{n}-\frac{m}{n}\right)}{\sqrt{2\mu^2(1-\mu)}} \le \Phi^{-1}(1-\alpha)\sqrt{\frac{2\left(\frac{m}{n}\right)^2\frac{n-m}{n}}{2\mu^2(1-\mu)}}+o(1)\right\} \\
&\to 1-\Phi\left(\Phi^{-1}(1-\alpha)\right)=\alpha\,.
\end{aligned}$$

Q.E.D.