# Finite sample properties of the GMM Anderson-Rubin test

Maurice J.G. 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]Munich Center for the Economics of Aging, Max Planck Society, Germany

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

July 17, 2019

**Abstract**

In the construction of the GMM version of the Anderson and Rubin (AR) test statistic there is the choice to use either uncentered or centered moment conditions to form the weighting matrix. We show that the centered GMM-AR test becomes oversized if the number of moment conditions is moderately large, while the uncentered version becomes conservative at conventional significance levels. 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. Monte Carlo experiments corroborate our theoretical findings and illustrate the accuracy of the degrees-of-freedom corrected, centered GMM-AR statistic in finite samples.

**Key Words:** Anderson and Rubin test, centering, degrees-of-freedom, Generalized Method of Moments

*corresponding author: farbmacher@econ.lmu.de

# 1 Introduction

The Generalized Method of Moments (Hansen, 1982) is a commonly employed procedure to estimate and test the parameters of econometric models. A main reason for using GMM is that it provides asymptotically efficient inference exploiting a minimal set of statistical assumptions. Despite the optimal asymptotic properties of GMM estimators and corresponding Wald test statistics, their behavior in finite samples can be rather peculiar due to the weakness of moment conditions. In case of weak identification, GMM coefficient estimators are biased and corresponding Wald-type tests perform poorly (Stock and Wright, 2000). To overcome the aforementioned problems, identication-robust GMM statistics can be used (Stock and Wright, 2000; Kleibergen, 2005; Newey and Windmeijer, 2009), which are based on the continuous updating GMM objective function. The main advantage of these identification-robust statistics is that, unlike conventional Wald-type inference, their limiting ($\chi^2$) distributions do not rely on the identification strength.

In this study we analyze the finite sample properties of the GMM version of the Anderson and Rubin (1949) test statistic, henceforth labeled as GMM-AR. Proposed by Stock and Wright (2000), the identification-robust GMM-AR statistic is based on the continuous updating GMM objective function. In the construction of the GMM-AR test statistic there is the choice to use either uncentered or centered moment conditions to form the weighting matrix. Table 1 provides a non-exhaustive overview of the recent literature exploiting, among other things, either the uncentered or centered version of the GMM-AR statistic. When the number of moment conditions is fixed or moderately large, the choice for centering does not matter asymptotically. However, in finite samples it affects inference and this is the purpose of this study.

We show analytically that the centered GMM-AR test using common $\chi^2$ critical values becomes oversized when the number of moments increases, while the uncentered version becomes conservative at conventional significance levels. The former corroborates the simulation results of Kleibergen and Mavroeidis (2009) and Hayakawa and Pesaran (2015), who both use the centered GMM-AR statistic and report large size distortions when the number of moment conditions is relatively large compared to sample size. The latter theoretical finding has to our knowledge not been explicitly noticed before. Newey and Windmeijer (2009) show that, when the number of moment conditions is not too large compared to the sample size, the uncentered

GMM-AR statistic remains size correct asymptotically. However, their simulation results also show that the uncentered GMM-AR tends to underreject as the number of moments increases.

We propose a degrees-of-freedom correction for the centered GMM-AR test which markedly improves its finite sample properties. To substantiate our degrees-of-freedom correction, we derive an asymptotic expansion of the GMM-AR test and obtain an Edgeworth approximation of its finite sample distribution using results from Kleibergen (2019). It turns out that the modified centered GMM-AR statistic implicitly incorporates an Edgeworth correction and therefore follows a $\chi^2$ distribution more closely relative to the centered AR statistic without degrees-of-freedom correction or the uncentered AR statistic. We focus on a setup in which the number of moments ($m$) is either fixed or just moderately large with respect to the sample size ($n$) such that $m^3/n \to 0$, which corresponds to the many moments setup of Newey and Windmeijer (2009).

We obtain a more accurate understanding of the extent to which the centering of the weighting matrix and the degrees-of-freedom correction influence 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, the difference between a centered and uncentered definition of the weighting matrix is negligible. In applications where $m$ becomes moderately large, however, the choice between these definitions is essential. We observe substantial differences in the actual size of the test statistics, which may point to conflicting inferences in practice. Furthermore, we find that the degrees-of-freedom corrected, centered GMM-AR statistic is size correct for both small and moderately large number of moment conditions.

In the next section we discuss the different definitions of the GMM-AR test and propose our degrees-of-freedom correction. In Section 3 we provide theoretical results to substantiate our correction. Section 4 shows Monte Carlo simulation results to illustrate our theoretical findings. Section 5 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, $\beta$, where $m \geq p$. $\beta_0$ 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. In case of a linear model, i.e. $g(w_i, \beta_0) = z_i(y_i - x_i'\beta_0)$, and imposing homoskedasticity, the IV-AR test statistic is defined as

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

where $Z$ is an $n \times m$ matrix containing the instrumental variables and $u = y - X\beta_0$ with $X$ a $n \times p$ matrix containing the endogenous regressors. $P_Z = Z(Z'Z)^{-1}Z'$, $M_Z = I_n - P_Z$ 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).

When the number of instruments is small, the IV-AR statistic is asymptotically $\chi^2$ distributed. In case of many instruments, alternative asymptotic approximations are available. Andrews and Stock (2007) consider the linear IV model, but allow for moderately many instruments ($m^3/n \to 0$) and show that the IV-AR statistic is normally distributed in this setup. Anatolyev and Gospodinov (2011) consider a setup in which the number of instruments grows at the same rate as the sample size such that $m/n \to \mu$ with $0 \leq \mu < 1$, and propose a many instrument modification for the IV-AR statistic.

While there exists only one IV-AR statistic, there are two versions of the GMM-AR test in the literature. 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 version exploits a weighting matrix 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 $\widehat{g}(\beta) = n^{-1}\sum_{i=1}^{n} g_i(\beta)$. Table 1 provides a non-exhaustive overview of studies using, among other things, either uncentered or centered GMM-AR statistics.

In this study we also consider a third GMM-AR statistic, which is based on the centered weighting matrix, but including a degrees-of-freedom correction. More specifically, we consider

$$\widetilde{\Omega}_{df}(\beta) = (n-m-2)^{-1}\sum_{i=1}^{n}[g_i(\beta) - \widehat{g}(\beta)][g_i(\beta) - \widehat{g}(\beta)]' \,. \tag{5}$$

Apart from the "$-2$"-term, this definition is also in line with the IV-AR test statistic, see equation (2), which also contains a degrees-of-freedom correction. The corresponding GMM-AR test statistics are defined as

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

Although asymptotically negligible for a small or moderately large number of moments, the degrees-of-freedom correction matters in finite samples, as we shall see. To the best of our knowledge, $\widetilde{AR}_{df}$ has not been used in the literature so far. In the next section we provide the theoretical background for this implementation.

# 3 Theoretical results

In the following we motivate our degrees-of-freedom correction for the centered GMM-AR statistic in several ways. First, we provide an algebraic connection between centered and uncentered GMM-AR test statistics by drawing parallels with basic Wald and Lagrange Multiplier test statistics. Second, we provide guidance on how to choose between the different versions of the GMM-AR test using higher-order asymptotic results.

## 3.1 Algebraic connections

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 $\beta$. Consider the auxiliary regression

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

with $\widehat{\gamma} = (G'G)^{-1}G'i$, 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

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

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

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

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 therefore be interpreted as a Lagrange Multiplier (LM) test of joint significance of $\gamma$.

The centered GMM-AR test is linked to the uncentered one in the following way

$$\widetilde{AR} = \frac{\widehat{AR}}{1 - \widehat{AR}/n}. \tag{10}$$

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). Note that the centered GMM-AR statistic can be interpreted as a Wald statistic of joint significance of $\gamma$.[1] The finite sample behavior of uncentered and centered AR statistics thus differs in the same way that LM statistics differ from Wald tests in finite samples. In particular, $\widetilde{AR}$ has a larger rejection frequency in finite samples than $\widehat{AR}$ due to the well-known fact that the Wald and LM statistics

[1] Define $W = \sqrt{n}(\widehat{\gamma} - 0)'\widehat{Var}(\sqrt{n}\widehat{\gamma})^{-1}\sqrt{n}(\widehat{\gamma} - 0)$ with $\widehat{Var}(\sqrt{n}\widehat{\gamma}) = \widehat{\sigma}_e^2(G'G/n)^{-1}$. It follows that $W = ni'G(G'G)^{-1}(G'G/n)(G'G)^{-1}G'i/\widehat{\sigma}_e^2 = \widehat{AR}/(1 - \widehat{AR}/n)\,.$

in linear models satisfy the inequality $W \geq LM$ (see Berndt and Savin, 1977; Breusch, 1979; Newey and West, 1987).

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 $\sigma_e^2$. 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 $(n-m)/(n\,\widehat{\sigma}_e^2)$ (see Lemma 7.7.1 in Anderson, 2003) and the resulting AR statistic would be $((n-m)/n)\ \widetilde{AR}$, which is just slightly different from our definition of $\widetilde{AR}_{df}$ in equation (6). Since a degrees-of-freedom correction is known to be important in small samples (see, e.g., Evans and Savin, 1982), 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.

## 3.2 Asymptotic results

For the asymptotic analysis we assume throughout that the following conditions hold

**Assumption 1.**

*(i)* $g_i(\beta_0)$ *is independent across* $i$;

*(ii)* *the eight order moments of* $g_i(\beta_0)$ *are finite.*

**Assumption 2.a.** *as* $n \to \infty$, $m$ *is finite.*

**Assumption 2.b.** *as* $n \to \infty$, $mE\left[\|g_i(\beta_0)\|^4\right]/n \to 0$.

The existence of eight order moments in Assumption 1 is stronger than necessary to obtain the limiting $\chi^2$ distribution of the GMM-AR statistic, but is needed to derive the higher-order expansions later on (Kleibergen, 2019). Assumption 2.a is the standard case of a fixed number of moment conditions, while Assumption 2.b is equal to the many weak moments setup in Newey and Windmeijer (2009) in which the number of moment conditions grows slowly with the sample size. In particular, it allows for a moderately large number of moment conditions in the sense that $m^3/n \to 0$. Occasionally we will use this many moments setup, but most results in this section have been derived under fixed $m$ asymptotics.

In simulation studies typically the uncentered GMM-AR statistic is shown to have control over size, i.e. actual rejection frequencies do not exceed the nominal significance level (e.g. Newey and Windmeijer, 2009). In contrast, some studies (Kleibergen and Mavroeidis, 2009; Hayakawa and Pesaran, 2015) report large size distortions for the centered GMM-AR statistic. We therefore first analyze the difference between centered and uncentered GMM-AR statistics.

Proposition 1 quantifies, in expectation, the difference between centered and uncentered AR statistics. Part A analyzes the setting in which the number of moment conditions ($m$) is fixed, while Part B allows $m$ to grow but at a slower rate than the sample size.

**Proposition 1.**

***Part A:*** *Under Assumptions 1 and 2.a, we have under* $H_0$

$$E\left[\widetilde{AR} - \widehat{AR}\right] = \frac{m^2 + 2m}{n} + O\left(n^{-2}\right).$$

***Part B:*** *Under Assumptions 1 and 2.b, we have under* $H_0$

$$\widetilde{AR} - \widehat{AR} = \frac{1}{n}\left(n\hat{g}'\Omega^{-1}\hat{g}\right)^2 + o_p\left(\frac{m}{n}\right),$$

*where* $\frac{1}{n}\left(n\hat{g}'\Omega^{-1}\hat{g}\right)^2 = O_p\left(\frac{m}{n}\right)$ *with*

$$E\left[\frac{1}{n}\left(n\hat{g}'\Omega^{-1}\hat{g}\right)^2\right] = \frac{m^2 + 2m}{n} + O\left(\frac{m^2}{n^2}\right).$$

*Proof.* See Appendix A.1.

In both settings the leading term in the approximation of the difference has the same expected value equal to $\frac{m^2+2m}{n}$. For small $m$ the difference between uncentered and centered GMM-AR is relatively small. For example, when $n = 100$ and $m = 3$ we have $\frac{m^2+2m}{n} = 0.15$, whereas for $m = 20$ we have $\frac{m^2+2m}{n} = 4.4$, which is already roughly a quarter of the expectation of the uncentered GMM-AR statistic. The Monte Carlo simulations in Section 4 will demonstrate that, for moderately large $m$, the approximation of Proposition 1 captures a large fraction of the differences in mean between centered and uncentered AR statistics.

For the rest of the asymptotic analysis we focus on the case where the number of moment conditions ($m$) is fixed. Subsequently, the degrees-of-freedom correction of the centered GMM-AR statistic is compared to the uncentered GMM-AR statistic.

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

$$E\left[\widetilde{AR}_{df} - \widehat{AR}\right] = O\left(n^{-2}\right).$$

*Proof.* See Appendix A.1.

From Corollary 1 we infer that, compared to the difference between $\widetilde{AR}$ and $\widehat{AR}$ statistics, the degrees-of-freedom correction removes the $O\left(n^{-1}\right)$ term in Proposition 1. Especially when $m$ is not very small compared to $n$, this is going to matter for inference.

Although the three GMM-AR test statistics in (6) are asymptotically equivalent (if $m$ is fixed), it is clear that their finite-sample behavior differs and thus may lead to conflicting inferences in practice. Because $\widetilde{AR}_{df}$ is in expectation much closer to $\widehat{AR}$ than to $\widetilde{AR}$, it is expected that the degrees-of-freedom correction is effective in reducing the size distortion of the centered GMM-AR statistic. To obtain a more accurate understanding of the extent to which the finite-sample performance of the discussed tests is affected, a higher-order asymptotic expansion is derived under the following additional assumptions.

**Assumption 3.** *Cramèr condition: for a m-dimensional vector $t \in \mathbb{R}^m$, it holds that*

$$\lim_{\|t\|\to\infty} \sup \left|E\left[\exp\left(it'\psi\right)\right]\right| < 1,$$

*where $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i(\beta_0) \to_d \psi$.*

**Assumption 4.** *$g_i(\beta_0)$ is symmetrically distributed.*

Assumption 3 is necessary for the Edgeworth approximation to hold (Kleibergen, 2019), while Assumption 4 removes its dependence on odd moments. This yields the following results.

**Proposition 2.** *Under Assumptions 1, 2.a, 3 and 4, 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* $\chi^2_m$ *distributed random variable evaluated at* $x$.

*Proof.* See Appendix A.2.

Proposition 2 shows that the centered AR statistic is size distorted up to order $O(n^{-1})$ when using asymptotic critical values from the $\chi^2_m$ distribution. More importantly, the size distortion is increasing in the ratio $m/n$. We illustrate this result in Figure 1, which shows the size distortion for various values of $m$ with $n = 100$. The $x$-axis represents the number of moment conditions and the $y$-axis shows the rejection frequency of the centered GMM-AR statistic according to the Edgeworth approximation from Proposition 2. The nominal significance level $\alpha$ is set equal to 0.05. It is clearly seen that for all values of $m$ the centered GMM-AR test rejects the null hypothesis too often and that the size distortion increases in $m$.

Proposition 2 implies that, for given $m$, the uncentered AR statistic is likely to be conservative at conventional levels of significance. We illustrate this in Figure 2a for $m = 20$ using a p-value plot. The $x$-axis of the p-value plot represents the nominal significance level $\alpha$ and the $y$-axis represents the rejection frequency of the uncentered GMM-AR statistic derived from Proposition 2. Hence, dots below the diagonal indicate conservative test results. In this example the uncentered GMM-AR statistic is size correct only for a nominal significance level equal to 0.34, while the test underrejects at lower significance levels.[2] Additionally, Proposition 2 implies that, for a given significance level, the uncentered GMM-AR statistic is getting increasingly conservative for larger $m$. Figure 2b depicts this for the conventional significance level $\alpha = 0.05$.

Finally, Proposition 2 shows that our degrees-of-freedom correction, which is straightforward to implement, solves the size distortion of the centered AR statistic and does not suffer from the

[2] Let $x$ denote the $1-\alpha$ quantile of a $\chi^2_m$ distribution. According to Proposition 2 the uncentered GMM-AR test has correct size only if $nx/(n-m-2+x) = x$. Given that $x > 0$, this means that the uncentered GMM-AR statistic attains the nominal significance level only if $x = m+2$ and becomes conservative when $x > m+2$.

underrejection of the uncentered AR test. These theoretical results justify the use of a degrees-of-freedom correction of the centered GMM-AR statistic as an implicit Edgeworth correction. It explains the superior behavior of $\widetilde{AR}_{df}$ reported in the Monte Carlo simulations of Section 4, as we shall see.

Note that the assumption of a symmetrical distribution has been used 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 moment conditions.

# 4 Simulation Results

In the following we analyze the finite sample performance of the three versions of the GMM-AR statistic by conducting a series of Monte Carlo simulations for linear and nonlinear models. First, the linear model is discussed and subsequently the simulation results for a nonlinear specification are reported. Both settings are such that Assumptions 1-4 hold. Therefore, under the null hypothesis we expect the centered GMM-AR test to overreject, the uncentered GMM-AR test to underreject, and the degrees-of-freedom corrected, centered GMM-AR test to be size correct.

The design of the linear model is

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

where $\iota_m$ is an $m$-vector of ones. Without loss of generalization we assume that $x$ 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 set to either 100 or 1000, while the degree of endogeneity $\rho$ is set to 0.5. The asymptotic first-stage F statistic ($F^\infty$) is fixed at 1, which implies that the set of instruments is equally weak with varying number of moment conditions. The concentration parameter $CP = F^\infty \cdot m$ is then equal to $m$. The number of replications for each experiment is $10,000$.

Table 2 shows the simulation results of the three GMM-AR statistics for the linear model. For various values of $m$ and $n$ we report the mean and 95% percentile of the sampling distribution of each of the three test statistics. Furthermore, we calculate the actual rejection frequency (RF) of nominal 5% tests.

For small to moderate $m/n$ the approximation in Proposition 1 captures a large fraction of 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 the observed mean difference is 1.32. For $m = 20$ this is $\frac{m^2+2m}{n} = 4.4$, and the observed mean difference is 5.53.

While the averages of the GMM-AR statistic 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 nominal significance level. In the small sample setting ($n = 100$) with many instruments ($m = 20$) the actual rejection frequency is around 0.015 although the nominal significance level is set to 0.05. On the other hand, the centered GMM-AR test ($\widetilde{AR}$) overrejects more and more severely when $m$ increases (the actual size is 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, as predicted by Proposition 2.[3]

The GMM-AR statistic ($\widetilde{AR}_{df}$), which uses the 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 rejection frequency is around 0.061.

For the nonlinear specification we replace the linear index by $exp(\beta_0 x_i)$, set $\beta_0 = 1$, and let $u_i$ 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, corroborating the theoretical results. This pattern becomes more distinct when the number of moment conditions increases

[3] The rejection frequencies in the Monte Carlo simulations closely follow the patterns in Figures 1 and 2 indicating the accuracy of Proposition 2 in finite samples. These results are available from the authors upon request.

with respect to the sample size. The GMM-AR statistic ($\widetilde{AR}_{df}$), which uses the degrees-of-freedom correction, performs very well in this setting with actual size close to the nominal size.

# 5 Conclusion

In the construction of the GMM version of the Anderson and Rubin (1949) test statistic there is the choice to use either uncentered or centered moment conditions to form the weighting matrix. Our analytical and simulation results show that the centered GMM-AR test becomes oversized if the number of moments is moderately large, while the uncentered version is conservative. These size properties of the GMM-AR test statistic based on the uncentered or centered weighting matrix resemble the behavior of Lagrange multiplier or Wald test statistics without degrees-of-freedom correction.

Exploiting asymptotic expansion techniques, we derive a degrees-of-freedom correction in the centered version of the GMM-AR test, which implicitly incorporates an Edgeworth correction. Monte Carlo experiments corroborate our theoretical findings and illustrate the accuracy of the degrees-of-freedom corrected centered GMM-AR statistic in finite samples. Based on our findings, we recommend to use the centered GMM-AR test with the degrees-of-freedom correction in applied research. If controlling the size is the main aim, then using the conservative, uncentered GMM-AR test is a viable alternative.

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); Hayakawa and Pesaran (2015); Kleibergen (2019) |
| 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 (2009); Wright (2010); Hausman *et al.* (2011); Caner and Yildiz (2012); Caner (2014) |

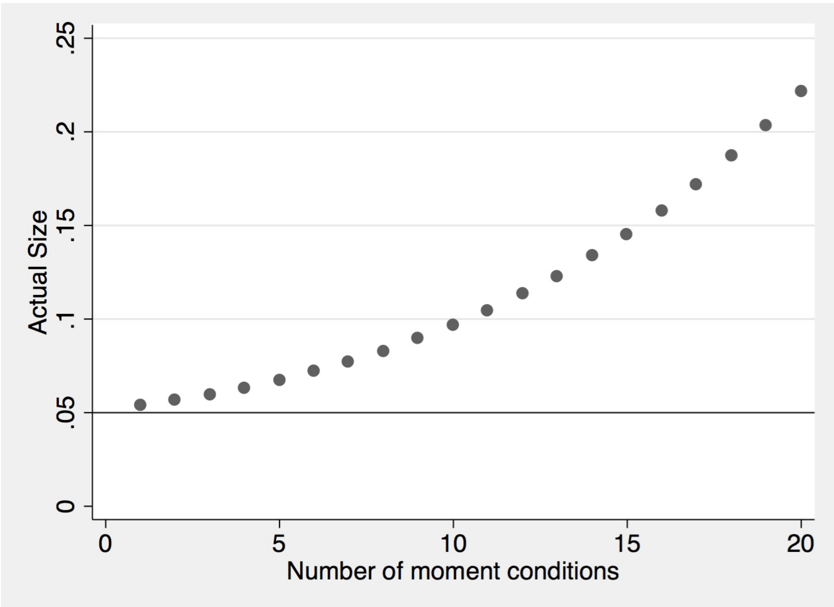
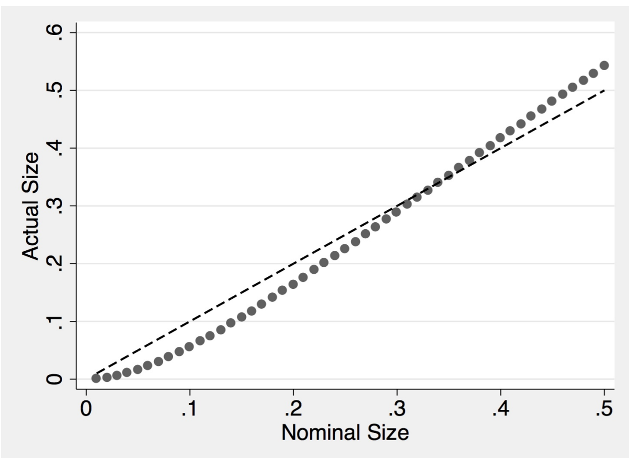
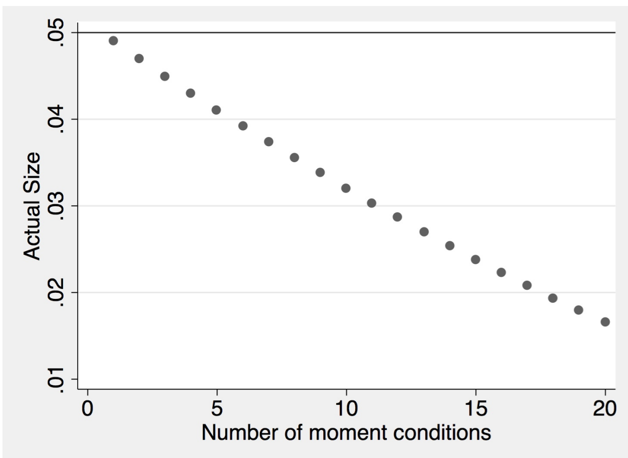



Figure 1: Illustration of Proposition 2 for centered GMM-AR test




(a) $p$-value plot for $m = 20$




(b) Actual size for $\alpha$=0.05

Figure 2: Illustration of Proposition 2 for uncentered GMM-AR test

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

| m | mean | $\widehat{AR}$ p95 | RF | mean | $\widetilde{AR}$ p95 | RF | mean | $\widetilde{AR}_{df}$ 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 |

$\rho = 0.5$; $F^{\infty} = 1$; 10,000 replications. Rejection frequencies (RF) for $H_0 : \beta = \beta_0$. 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).

| m | mean | $\widehat{AR}$ p95 | RF | mean | $\widetilde{AR}$ p95 | RF | mean | $\widetilde{AR}_{df}$ p95 | RF |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $n = 100$ | | | | |
| 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 |
| | | | | | $n = 1000$ | | | | |
| 3 | 2.99 | 7.90 | 0.052 | 3.00 | 7.97 | 0.053 | 2.99 | 7.94 | 0.052 |
| 5 | 5.00 | 11.04 | 0.049 | 5.03 | 11.17 | 0.052 | 5.01 | 11.11 | 0.051 |
| 10 | 10.12 | 18.39 | 0.052 | 10.24 | 18.73 | 0.056 | 10.14 | 18.55 | 0.054 |
| 20 | 20.06 | 30.65 | 0.040 | 20.51 | 31.62 | 0.053 | 20.10 | 30.98 | 0.044 |
| 30 | 29.94 | 42.78 | 0.041 | 30.93 | 44.70 | 0.062 | 30.00 | 43.36 | 0.047 |
| 40 | 40.07 | 55.00 | 0.041 | 41.82 | 58.21 | 0.075 | 40.15 | 55.88 | 0.051 |

$\rho = 0.5$; $F^{\infty} = 1$; 10,000 replications. Rejection frequencies (RF) for $H_0 : \beta = \beta_0$. Nominal significance level 5%. The asymptotic critical values are reported in Table 2.

# References

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

Anderson TW, 2003. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York

Anderson TW, 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

Andrews DWK, Stock JH, 2007. Testing with many weak instruments. *Journal of Econometrics* **138**(1): 24–46

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 ER, Savin NE, 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 TS, 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* **182**(2): 247–268

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

Donald SG, Imbens GW, Newey WK, 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* **117**(1): 55–93

Donald SG, Newey WK, 2000. A jackknife interpretation of the continuous updating estimator. *Economics Letters* **67**(3): 239–243

Evans GBA, Savin NE, 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 RJ, 2005. Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory* **21**(04): 667–709

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

Hansen LP, 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4): 1029–1054

Hansen LP, 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, Pesaran MH, 2015. Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics* **188**(8): 111–134

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

Kleibergen F, 2019. Improved accuracy of weak instrument robust GMM statistics through bootstrap and Edgeworth approximations. Discussion Paper February 2019, University of Amsterdam

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 WK, Smith RJ, 2004. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**(1): 219–255

Newey WK, West KD, 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review* **28**(3): 777–787

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

Phillips PCB, Park JY, 1988. On the formulation of Wald tests of nonlinear restrictions. *Econometrica* **56**(5): 1065–1083

Staiger D, Stock JH, 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**(3): 557–586

Stock JH, Wright JH, 2000. GMM with weak identification. *Econometrica* **68**(5): 1055–1096

Stock JH, Wright JH, 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*, 603–624. Springer

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

# A Appendix

## A.1 Proof of Proposition 1

Proof of Part A:
We suppress the dependence of $\widetilde{AR}$, $\widehat{AR}$, $\Omega$, $\hat{\Omega}$ and $\hat{g}$ on $\beta_0$ and $\sum_i$ is short for $\sum_{i=1}^n$. Kleibergen (2019) develops an asymptotic expansion of the centered AR statistic by replacing the covariance matrix estimator $\tilde{\Omega}$ in the construction of $\widetilde{AR}$ by the following Taylor expansion around the true value $\beta_0$

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

The order of the remainder term follows from the $\sqrt{n}$ convergence rate of the covariance matrix estimator, i.e. $\tilde{\Omega} - \Omega = O_p\left(n^{-1/2}\right)$. Substituting this in the centered AR statistic, we obtain under $H_0 : \beta = \beta_0$ the following higher-order expression

$$
\begin{aligned}
\widetilde{AR} &= n\hat{g}'\Omega^{-1}\hat{g} - n\hat{g}'\Omega^{-1}\left(\tilde{\Omega} - \Omega\right)\Omega^{-1}\hat{g} \\
&\quad + n\hat{g}'\Omega^{-1}\left(\tilde{\Omega} - \Omega\right)\Omega^{-1}\left(\tilde{\Omega} - \Omega\right)\Omega^{-1}\hat{g} + o_p\left(n^{-1}\right) \\
&= \widetilde{AR}_0 + \widetilde{AR}_1 + \widetilde{AR}_2 + o_p\left(n^{-1}\right).
\end{aligned}
$$

Theorem 1 of Kleibergen (2019) furthermore shows that

$$
\begin{aligned}
E\left[\widetilde{AR}_0\right] &= m, \\
E\left[\widetilde{AR}_1\right] &= -\frac{1}{n}E\left[\left(g_i'\Omega^{-1}g_i\right)^2\right] + \frac{m}{n} + \frac{1}{n}\left(m^2 + 2m\right) + O\left(n^{-2}\right), \\
E\left[\widetilde{AR}_2\right] &= \frac{1}{n}E\left[\left(g_i'\Omega^{-1}g_i\right)^2\right] - \frac{m}{n} + \frac{1}{n}E\left[g_i'\Omega^{-1}g_ig_i'\right]\Omega^{-1}E\left[g_ig_i'\Omega^{-1}g_i\right] + O\left(n^{-2}\right).
\end{aligned}
$$

Combining terms this results in

$$
E\left[\widetilde{AR}\right] = m + \frac{1}{n}\left(m^2 + 2m + E\left[g_i'\Omega^{-1}g_ig_i'\right]\Omega^{-1}E\left[g_ig_i'\Omega^{-1}g_i\right]\right) + O\left(n^{-2}\right).
$$

We develop an analogous result for the uncentered AR statistic by replacing the covariance matrix estimator $\hat{\Omega}$ in the construction of $\widehat{AR}$ by the following Taylor expansion around the true value $\beta_0$

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

Substituting this in the uncentered AR statistic, we obtain under $H_0 : \beta = \beta_0$ the following higher-order expression

$$
\begin{aligned}
\widehat{AR} &= n\hat{g}'\Omega^{-1}\hat{g} - n\hat{g}'\Omega^{-1}\left(\hat{\Omega} - \Omega\right)\Omega^{-1}\hat{g} \\
&\quad + n\hat{g}'\Omega^{-1}\left(\hat{\Omega} - \Omega\right)\Omega^{-1}\left(\hat{\Omega} - \Omega\right)\Omega^{-1}\hat{g} + o_p\left(n^{-1}\right) \\
&= \widehat{AR}_0 + \widehat{AR}_1 + \widehat{AR}_2 + o_p\left(n^{-1}\right).
\end{aligned}
$$

The leading term in the expansion $\widehat{AR}_0$ is equal to the leading term $\widetilde{AR}_0$ in the expansion of the centered AR statistic, and we have $E\left[\widehat{AR}_0\right] = m$. Regarding the higher-order terms $\widehat{AR}_1$ and $\widehat{AR}_2$ we find under Assumption 1

$$
\begin{aligned}
E\left[\widehat{AR}_1\right] &= -E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2}\left(g_{i_2}g'_{i_2}-\Omega\right)\Omega^{-1}\sum_{i_3} g_{i_3}\right] \\
&= -E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2} g_{i_2}g'_{i_2}\Omega^{-1}\sum_{i_3} g_{i_3}\right] + E\left[\frac{1}{n}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_3} g_{i_3}\right] \\
&= -E\left[\frac{1}{n^2}\sum_{i_1}\sum_{i_2\neq i_1} g'_{i_1}\Omega^{-1}g_{i_2}g'_{i_2}\Omega^{-1}g_{i1}\right] - E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}g_{i_1}g'_{i_1}\Omega^{-1}g_{i1}\right] + E\left[\frac{1}{n}\sum_{i_1} g'_{i_1}\Omega^{-1}g_{i_1}\right] \\
&= -\frac{n(n-1)}{n^2}m - \frac{1}{n}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right] + m \\
&= \frac{m}{n} - \frac{1}{n}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right],
\end{aligned}
$$

$$
\begin{aligned}
E\left[\widehat{AR}_2\right] &= E\left[\frac{1}{n^3}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2}\left(g_{i_2}g'_{i_2}-\Omega\right)\Omega^{-1}\sum_{i_3}\left(g_{i_3}g'_{i_3}-\Omega\right)\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&= E\left[\frac{1}{n^3}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2} g_{i_2}g'_{i_2}\Omega^{-1}\sum_{i_3}\left(g_{i_3}g'_{i_3}-\Omega\right)\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&- E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_3}\left(g_{i_3}g'_{i_3}-\Omega\right)\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&= E\left[\frac{1}{n^3}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2} g_{i_2}g'_{i_2}\Omega^{-1}\sum_{i_3} g_{i_3}g'_{i_3}\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&- E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2} g_{i_2}g'_{i_2}\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&- E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_3} g_{i_3}g'_{i_3}\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&+ E\left[\frac{1}{n}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_4} g_{i_4}\right].
\end{aligned}
$$

We use intermediate results from Kleibergen (2019) to derive the expectation of the four components in the above expression

$$E\left[\frac{1}{n^3}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2} g_{i_2}g'_{i_2}\Omega^{-1}\sum_{i_3} g_{i_3}g'_{i_3}\Omega^{-1}\sum_{i_4} g_{i_4}\right]$$
$$= \frac{(n-1)(n-2)}{n^2}m + 3\frac{n-1}{n^2}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right]$$
$$+ \frac{n-1}{n^2}E\left[g'_i\Omega^{-1}g_i g'_i\right]\Omega^{-1}E\left[g_i g'_i\Omega^{-1}g_i\right] + O\left(n^{-2}\right)$$
$$= m + \frac{3}{n}\left(E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right] - m\right) + \frac{1}{n}E\left[g'_i\Omega^{-1}g_i g'_i\right]\Omega^{-1}E\left[g_i g'_i\Omega^{-1}g_i\right] + O\left(n^{-2}\right),$$

$$E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2} g_{i_2}g'_{i_2}\Omega^{-1}\sum_{i_4} g_{i_4}\right]$$
$$= E\left[\frac{1}{n^2}\sum_{i_1}\sum_{i_2\neq i_1} g'_{i_1}\Omega^{-1}g_{i_2}g'_{i_2}\Omega^{-1}g_{i1}\right] + E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}g_{i_1}g'_{i_1}\Omega^{-1}g_{i1}\right]$$
$$= \frac{n-1}{n}m + \frac{1}{n}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right],$$

$$E\left[\frac{1}{n^2}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_3} g_{i_3}g'_{i_3}\Omega^{-1}\sum_{i_4} g_{i_4}\right]$$
$$= \frac{n-1}{n}m + \frac{1}{n}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right],$$
$$E\left[\frac{1}{n}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_4} g_{i_4}\right] = m.$$

Combining terms we have

$$E\left[\widehat{AR}_2\right] = m + \frac{3}{n}\left(E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right] - m\right) + \frac{1}{n}E\left[g'_i\Omega^{-1}g_i g'_i\right]\Omega^{-1}E\left[g_i g'_i\Omega^{-1}g_i\right]$$
$$- 2\left(\frac{n-1}{n}m + \frac{1}{n}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right]\right) + m + O\left(n^{-2}\right)$$
$$= -\frac{m}{n} + \frac{1}{n}E\left[\left(g'_i\Omega^{-1}g_i\right)^2\right] + \frac{1}{n}E\left[g'_i\Omega^{-1}g_i g'_i\right]\Omega^{-1}E\left[g_i g'_i\Omega^{-1}g_i\right] + O\left(n^{-2}\right).$$

To determine the order of the expectation of the remainder term, we analyze the next higher-order term in the expansion

$$\widehat{AR}_3 = \frac{1}{n^4}\sum_{i_1} g'_{i_1}\Omega^{-1}\sum_{i_2}\left(g_{i_2}g'_{i_2} - \Omega\right)\Omega^{-1}\sum_{i_3}\left(g_{i_3}g'_{i_3} - \Omega\right)\Omega^{-1}\sum_{i_4}\left(g_{i_4}g'_{i_4} - \Omega\right)\Omega^{-1}\sum_{i_5} g_{i_5}.$$

Taking expectations and assuming finite eight order moments (Assumption 1), the highest order will be $O\left(n^{-2}\right)$. This results from combining two fourth order products with identical indices

$$E\left[\frac{1}{n^4}\sum_{i_1}\sum_{i_2\neq i_1} g'_{i_1}\Omega^{-1}g_{i_1}g'_{i_1}\Omega^{-1}g_{i_2}g'_{i_2}\Omega^{-1}g_{i_2}g'_{i_2}\Omega^{-1}g_{i_1}\right] = O\left(n^{-2}\right).$$

Combining terms we find for the uncentered AR statistic

$$E\left[\widehat{AR}\right] = m + \frac{1}{n}E\left(g_i'\Omega^{-1}g_ig_i'\right)\Omega^{-1}E\left(g_ig_i'\Omega^{-1}g_i\right) + O\left(n^{-2}\right).$$

These results imply that

$$E\left[\widetilde{AR} - \widehat{AR}\right] = \frac{m^2+2m}{n} + O\left(n^{-2}\right),$$

which completes the proof.

Proof of Part B:
We suppress the dependence of $\widetilde{AR}$, $\widehat{AR}$, $\Omega$, $\hat{\Omega}$ and $\hat{g}$ on $\beta_0$ and $\sum_i$ is short for $\sum_{i=1}^n$. We adopt the many weak moments asymptotic framework of Newey and Windmeijer (2009). They assume that $mE\left[\|g_i\|^4\right]/n \longrightarrow 0$ (Assumption 2.b), which under finite moments (Assumption 1) implies $m^3/n \longrightarrow 0$. Newey and Windmeijer (2009, Theorem 4) derive that

$$\frac{\widehat{AR}-m}{\sqrt{2m}} \overset{d}{\to} N(0,1),$$

which implies that $\widehat{AR} = O_p(\sqrt{m})$, hence $\frac{\widehat{AR}}{n}$ is $O_p\left(\frac{\sqrt{m}}{n}\right)$. Note that this includes the case of standard (finite $m$) asymptotics for which $\frac{\widehat{AR}}{n}$ is $O_p(n^{-1})$. We can therefore expand (10) as follows

$$\begin{aligned}\widetilde{AR} &= \widehat{AR}\left(1+\frac{\widehat{AR}}{n}+\left(\frac{\widehat{AR}}{n}\right)^2+...\right)\\ &= \widehat{AR}+\frac{1}{n}\widehat{AR}^2+O_p\left(\frac{m\sqrt{m}}{n^2}\right).\end{aligned}$$

Newey and Windmeijer (2009) derive (see their Theorem 4)

$$\begin{aligned}\frac{\widehat{AR}}{n} &= \hat{g}'\Omega^{-1}\hat{g}+\hat{g}'\hat{\Omega}^{-1}\left(\Omega-\hat{\Omega}\right)\Omega^{-1}\hat{g}\\ &= \hat{g}'\Omega^{-1}\hat{g}+r,\end{aligned}$$

with the leading term $\hat{g}'\Omega^{-1}\hat{g} = O_p\left(\frac{\sqrt{m}}{n}\right)$ and the remainder term $r = o_p\left(\frac{\sqrt{m}}{n}\right)$. We therefore have

$$\begin{aligned}\frac{1}{n}\widehat{AR}^2 &= n\left(\hat{g}'\Omega^{-1}\hat{g}+r\right)^2\\ &= n\left(\hat{g}'\Omega^{-1}\hat{g}\right)^2+nr^2+2rn\hat{g}'\Omega^{-1}\hat{g}\\ &= \frac{1}{n}\left(n\hat{g}'\Omega^{-1}\hat{g}\right)^2+o_p\left(\frac{m}{n}\right),\end{aligned}$$

because both $nr^2$ and $2rn\hat{g}'\Omega^{-1}\hat{g}$ are $o_p\left(\frac{m}{n}\right)$. Furthermore, the quantity $\frac{1}{n}\left(n\hat{g}'\Omega^{-1}\hat{g}\right)^2 = O_p\left(\frac{m}{n}\right)$

with mean

$$
\begin{aligned}
E\left[\frac{1}{n}\left(n\hat{g}'\Omega^{-1}\hat{g}\right)^2\right] &= \frac{1}{n^3}E\left[\sum_{i_1} g_{i_1}'\Omega^{-1}\sum_{i_2} g_{i_2}\sum_{i_3} g_{i_3}'\Omega^{-1}\sum_{i_4} g_{i_4}\right] \\
&= \frac{1}{n^3}E\left[\sum_{i_1}\sum_{i_2\neq i_1} g_{i_1}'\Omega^{-1}g_{i_1}g_{i_2}'\Omega^{-1}g_{i_2}\right] + \frac{1}{n^3}E\left[\sum_{i_1}\sum_{i_2\neq i_1} g_{i_1}'\Omega^{-1}g_{i_2}g_{i_1}'\Omega^{-1}g_{i_2}\right] \\
&+ \frac{1}{n^3}E\left[\sum_{i_1}\sum_{i_2\neq i_1} g_{i_1}'\Omega^{-1}g_{i_2}g_{i_2}'\Omega^{-1}g_{i_1}\right] + \frac{1}{n^3}E\left[\sum_{i_1} g_{i_1}'\Omega^{-1}g_{i_1}g_{i_1}'\Omega^{-1}g_{i_1}\right] \\
&= \frac{n(n-1)}{n^3}\left(m^2+m+m\right) + \frac{1}{n^2}E\left[\left(g_i'\Omega^{-1}g_i\right)^2\right] \\
&= \frac{m^2+2m}{n} + \frac{1}{n^2}\left(E\left[\left(g_i'\Omega^{-1}g_i\right)^2\right] - \left(m^2+2m\right)\right) \\
&= \frac{m^2+2m}{n} + O\left(\frac{m^2}{n^2}\right),
\end{aligned}
$$

because $E\left[\left(g_i'\Omega^{-1}g_i\right)^2\right] = O(m^2)$ under Assumption 1.

Proof of Corollary 1:

We exploit the derived approximations for $E\left[\widetilde{AR}\right]$ and $E\left[\widehat{AR}\right]$ from Part A of Proposition 1. Upon collecting $O(n^{-2})$ and lower order terms in the remainder term, we have

$$
\begin{aligned}
E\left[\widetilde{AR}_{df} - \widehat{AR}\right] &= \frac{n-m-2}{n}E\left[\widetilde{AR}\right] - E\left[\widehat{AR}\right] \\
&= \left(1-\frac{m+2}{n}\right)\left[m + \frac{1}{n}\left(m^2+2m+E\left[g_i'\Omega^{-1}g_ig_i'\right]\Omega^{-1}E\left[g_ig_i'\Omega^{-1}g_i\right]\right)\right] \\
&- m - \frac{1}{n}E\left[g_i'\Omega^{-1}g_ig_i'\right]\Omega^{-1}E\left[g_ig_i'\Omega^{-1}g_i\right] + O\left(n^{-2}\right) \\
&= O\left(n^{-2}\right).
\end{aligned}
$$

## A.2 Proof of Proposition 2

We use Theorem 5 of Kleibergen (2019), which is an adaptation of Theorem 2.4 in Phillips and Park (1988), to provide the following Edgeworth approximation for the finite sample distribution of the centered AR statistic. Under Assumptions 1, 2.a and 3, Kleibergen (2019) derives that under $H_0 : \beta = \beta_0$

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

where $\Pr_{\chi^2_m}[x]$ is the distribution of a $\chi^2_m$ distributed random variable evaluated at $x$ and $S = \widetilde{AR}_1 + \widetilde{AR}_2$ with $\widetilde{AR}_1$ and $\widetilde{AR}_2$ defined in Proposition 1. In general the expectation of $S$, which is of order $O(N^{-1})$, depends on third moments. However, under Assumption 4 it simplifies to

$$E(S) = \frac{m^2 + 2m}{n},$$

which results 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

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

Finally, combining Proposition 2 with (10) we have

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

which completes the proof.