Weak Instruments, First-Stage Heteroskedasticity and the Robust F-test

Frank Windmeijer

Discussion Paper 19 / 708

February 2019

Department of Economics
University of Bristol
Priory Road Complex
Bristol BS8 1TU
United Kingdom

Weak Instruments, First-Stage Heteroskedasticity and the Robust F-test

Frank Windmeijer
Department of Economics
University of Bristol, UK
and Cemmap, UCL/IFS, London, UK

February 2019

Abstract

This paper is concerned with the findings related to the robust first-stage F-statistic in the Monte Carlo analysis of Andrews (2018), who found in a heteroskedastic design that even for very large values of the robust F-statistic, the standard 2SLS confidence intervals had large coverage distortions. This finding appears to discredit the robust F-statistic as a test for underidentification. However, it is shown here that large values of the robust F-statistic do imply that there is first-stage information, but this may not be utilised well by the 2SLS estimator, or the standard GMM estimator. An estimator that corrects for this is a robust GMM estimator, with the robust weight matrix not based on the structural residuals, but on the first-stage residuals. For the grouped data setting of Andrews (2018), this estimator gives the weights to the group specific estimators according to the group specific concentration parameters in the same way as 2SLS does under homoskedasticity, which is formally shown using weak instrument asymptotics. This estimator is much better behaved than the 2SLS estimator in this design, behaving well in terms of relative bias and Wald test size distortion at more 'standard' values of the robust F-statistic. We further derive the conditions under which the Stock and Yogo (2005) weak instruments critical values apply to the robust F-statistic in relation to the behaviour of this GMM estimator.

JEL Classification: C12, C36
Keywords: weak instruments, heteroskedasticity, F-test, Stock-Yogo critical values

1 Introduction

It is commonplace to report the first-stage F-statistic as a test for underidentification in linear single endogenous variable models estimated by two-stage least squares (2SLS). This could either be a non-robust or robust version of the test, with robustness to for example heteroskedasticity, serial correlation and/or clustering. Under maintained assumptions, these are valid tests for the null $H_0 : \pi = 0$ in the first-stage linear specification $x = Z\pi + v$, where x is the endogenous explanatory variable in the model of interest $y = x\beta + u$, and Z are the instruments. If the null is not rejected, then this is an indication that the relevance condition of the instruments does not hold and that the 2SLS estimator does not provide a meaningful estimate of the parameter of interest β. A rejection of the null does, however, not necessarily imply that the 2SLS estimator is well behaved. This follows the work of Staiger and Stock (1997) and Stock and Yogo (2005), with the latter providing critical values for the first-stage non-robust F-statistic for null hypotheses of weak instruments in terms of bias of the 2SLS estimator relative to that of the OLS estimator and Wald test size distortion. These non-robust weak instruments F-tests are valid only under conditional homoskedasticity, no serial correlation and no clustering of both the first-stage errors v and the structural errors u, and do not apply to the robust F-test in general designs, see Bun and de Haan (2010), Olea and Pflueger (2013) and Andrews (2018). For general designs Olea and Pflueger (2013) proposed the effective first-stage F-statistic and critical values linked to the Nagar bias of the 2SLS estimator, whereas Andrews (2018) obtained valid two-step identification robust confidence sets.

This paper is concerned with the findings related to the robust F-statistic in the Monte Carlo analysis of Andrews (2018, Supplementary Appendix (SA)). In a cross sectional heteroskedastic design he found that even for very large values of the robust F-statistic, the standard 2SLS confidence intervals had large coverage distortions. For example, for a high endogeneity design, "the 2SLS confidence set has a 15% coverage distortion even when the mean of the first-stage robust F-statistic is 100,000", Andrews (2018, SA, p 11). This is a striking finding and appears to discredit the robust F-statistic as a test for underidentification. However, I show here that large values of the robust F-statistic do imply that there is first-stage information, but this may not be utilised well by the 2SLS estimator, or GMM estimators that incorporate heteroskedasticity in the structural error u only.

Andrews (2018) design is the same as a grouped data one, see Angrist (1991) and the discussion in Angrist and Pischke (2009), where the instruments are mutually exclusive group membership indicators. Denoting the groups by $s = 1, ..., S$, the group specific concentration parameter values are determined by the ratios $\pi_s^2/\sigma_{v,s}^2$, where $\sigma_{v,s}^2$ is the group specific variance of the first-stage error v. The 2SLS estimator is a weighted average of the group specific estimators of β, giving more weight to large concentration parameter groups if v is homoskedastic. However, as shown in Section 3, this may not happen under heteroskedasticity, where 2SLS gives more weight to high variance $\sigma_{v,s}^2$ groups, everything else constant. In the design of Andrews (2018) we consider here, there is one informative group, leading to the large value of the robust F-statistic, but this group has a small variance $\sigma_{v,s}^2$, and therefore gets only a relatively small weight in the 2SLS estimator.

An estimator that correctly gives larger weights to more informative groups is a robust GMM estimator, not using the structural residuals $\widehat{u}$, but the first-stage residuals $\widehat{v}$ in the robust weight matrix. This estimator, called GMMf, is introduced in Section 4 and gives the weights to the group specific estimators according to the group specific concentration parameters in the same way as 2SLS does under homoskedasticity. This is further formally shown using weak instrument asymptotics in Section 5. Section 6 discusses the potential problems of the standard GMM estimator that uses a robust weight matrix based on the conditional variances of the structural errors u. Monte Carlo results in Section 7 show that the GMMf estimator exploits the available information well, with much better relative bias and Wald test size properties than the 2SLS estimator for values of the robust F-statistic in line with those of the non-robust F-statistic and behaviour of the 2SLS estimator in the homoskedastic case.

For a general setting, we report in Section 8 the conditions under which the Stock and Yogo (2005) critical values can be applied to the robust F-statistic in relation to the behaviour of the GMMf estimator. These conditions are derived in Appendix A.2. Whilst these have limited applicability, the fully homoskedastic design is a special case.

2 Model and Assumptions

We consider the model as in Andrews (2018, SA C.3), which is the same as a grouped data IV setup,

$$
\begin{aligned}
y_i &= x_i\beta + u_i \\
x_i &= z_i'\pi + v_i,
\end{aligned}
$$

for $i = 1, ..., n$, where z_i is a vector of mutually exclusive binary indicator variables, $z_i \in \{e_1, ..., e_S\}$, with e_s is a $S \times 1$ vector with 1 in the sth entry and zeros everywhere else. Assumptions for standard asymptotic normality results hold and the robust variance of the limiting distribution of the parameters can be estimated consistently.

The variance-covariance structure for the errors is modelled fully flexibly by group, and specified as

$$
\left(\left(\begin{array}{c} u_i \\ v_i \end{array}\right) | z_i = e_s\right) \sim (0, \Sigma_s).
$$

$$
\Sigma_s = \left[\begin{array}{cc} \sigma_{u,s}^2 & \sigma_{uv,s} \\ \sigma_{uv,s} & \sigma_{v,s}^2 \end{array}\right].
$$

At the group level, we therefore have for group member j

$$
\begin{aligned}
y_{js} &= x_{js}\beta + u_{js} && (1) \\
x_{js} &= \pi_s + v_{js} && (2)
\end{aligned}
$$

$$
\left(\begin{array}{c} u_{js} \\ v_{js} \end{array}\right) \sim (0, \Sigma_s)
$$

for $j = 1, ..., n_s$ and $s = 1, ..., S$, with n_s the number of observations in group s, $\sum_{s=1}^{S} n_s = n$, see also Bekker and Van der Ploeg (2005). We assume that $\lim_{n\to\infty} \frac{n_s}{n} = f_s$, with $0 < f_s < \infty$.

3 First-Stage F and 2SLS Weights

The OLS estimator of π_s is given by $\widehat{\pi}_s = \overline{x}_s = \frac{1}{n_s}\sum_{j=1}^{n_s} x_{js}$ and $Var(\widehat{\pi}_s) = \sigma_{v,s}^2/n_s$. The OLS residual is $\widehat{v}_{js} = x_{js} - \overline{x}_s$ and the estimator for the variance is given by $V\widehat{a}r(\widehat{\pi}_s) = \widehat{\sigma}_{v,s}^2/n_s$, where $\widehat{\sigma}_{v,s}^2 = \frac{1}{n_s}\sum_{j=1}^{n_s} \widehat{v}_{js}^2$. Let Z be the $n \times S$ matrix of instruments. For the vector π the OLS estimator is given by

$$
\widehat{\pi} = (Z'Z)^{-1} Z'x = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_S)'.
$$

Let

$$\begin{aligned}\widehat{\Omega}_v &= \sum_{i=1}^{n} \widehat{v}_i^2 z_i z_i' \\ &= \operatorname{diag}\left(n_s \widehat{\sigma}_{v,s}^2\right)\end{aligned} \tag{3}$$

where $\operatorname{diag}(a_s)$ is a diagonal matrix with sth diagonal element a_s. Then the robust estimator of $Var(\widehat{\pi})$ is given by

$$\begin{aligned}V\widehat{a}r_r(\widehat{\pi}) &= (Z'Z)^{-1} \widehat{\Omega}_v (Z'Z)^{-1} \\ &= \operatorname{diag}\left(\widehat{\sigma}_{v,s}^2 / n_s\right).\end{aligned}$$

The non-robust variance estimator is

$$\begin{aligned}V\widehat{a}r(\widehat{\pi}) &= \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{v}_i^2\right)(Z'Z)^{-1} \\ &= \left(\sum_{s=1}^{S} \frac{n_s}{n} \widehat{\sigma}_{v,s}^2\right) \operatorname{diag}\left(\frac{1}{n_s}\right).\end{aligned}$$

The group (or instrument) specific IV estimators for β are given by

$$\widehat{\beta}_s = \frac{z_s' y}{z_s' x} = \frac{\overline{y}_s}{\overline{x}_s}, \tag{4}$$

with $\overline{y}_s = \frac{1}{n_s}\sum_{j=1}^{n_s} y_{js}$, and the 2SLS estimator for β is, with $P_Z = Z(Z'Z)^{-1} Z'$,

$$\begin{aligned}\widehat{\beta}_{2sls} &= (x' P_Z x)^{-1} x' P_Z y \\ &= \frac{\sum_{s=1}^{S} n_s \overline{x}_s \overline{y}_s}{\sum_{s=1}^{S} n_s \overline{x}_s^2} \\ &= \frac{\sum_{s=1}^{S} n_s \overline{x}_s^2 (\overline{y}_s / \overline{x}_s)}{\sum_{s=1}^{S} n_s \overline{x}_s^2} = \sum_{s=1}^{S} w_{2sls,s} \widehat{\beta}_s,\end{aligned}$$

the standard result that $\widehat{\beta}_{2sls}$ is a linear combination of the instrument specific IV estimators, (see e.g. Windmeijer, 2018). The weights are given by

$$w_{2sls,s} = \frac{n_s \overline{x}_s^2}{\sum_{l=1}^{S} n_l \overline{x}_l^2} \geq 0 \tag{5}$$

and hence the 2SLS estimator is here a weighted average of the group specific estimators.

For the group specific estimates, the first-stage F-statistics are equal to the Wald statistics for testing the null hypotheses $H_0 : \pi_s = 0$, and are given by

$$F_{\pi_s} = \frac{\widehat{\pi}_s^2}{V\widehat{a}r\left(\widehat{\pi}_s\right)} = \frac{n_s \overline{x}_s^2}{\widehat{\sigma}_{v,s}^2} \tag{6}$$

for $s = 1, ..., S$. For each group specific IV estimator $\widehat{\beta}_s$ the standard weak instruments results of Staiger and Stock (1997) and Stock and Yogo (2005) apply. As these are just-identified models, we can relate the values of the F-statistics to Wald test size distortions.

The robust first-stage F-statistic for testing $H_0 : \pi = 0$ is given by

$$\begin{aligned} F_r &= \frac{1}{S}\widehat{\pi}' \left(V\widehat{a}r_r\left(\widehat{\pi}\right)\right)^{-1} \widehat{\pi} \\ &= \frac{1}{S}\sum_{s=1}^{S} \frac{n_s \overline{x}_s^2}{\widehat{\sigma}_{v,s}^2} = \frac{1}{S}\sum_{s=1}^{S} F_{\pi_s}. \end{aligned}$$

It is therefore clear, that if F_r is large, then at least one of the F_{π_s} is large.

The non-robust F-statistic is given by

$$\begin{aligned} F &= \frac{1}{S}\widehat{\pi}' \left(V\widehat{a}r\left(\widehat{\pi}\right)\right)^{-1} \widehat{\pi} \\ &= \frac{1}{S}\frac{\sum_{s=1}^{S} n_s \overline{x}_s^2}{\left(\sum_{s=1}^{S} \frac{n_s}{n}\widehat{\sigma}_{v,s}^2\right)} = \frac{1}{S}\sum_{s=1}^{S} \frac{\widehat{\sigma}_{v,s}^2}{\left(\sum_{l=1}^{S} \frac{n_l}{n}\widehat{\sigma}_{v,l}^2\right)} F_{\pi_s}. \end{aligned}$$

From (5) and (6) it follows that the weights for the 2SLS estimator are related to the individual F-statistics as follows

$$w_{2sls,s} = \frac{n_s \overline{x}_s^2}{\sum_{l=1}^{S} n_l \overline{x}_l^2} = \frac{\widehat{\sigma}_{v,s}^2 F_{\pi_s}}{\sum_{l=1}^{S} \widehat{\sigma}_{v,l}^2 F_{\pi_l}}.$$

Under homoskedasticity, $\widehat{\sigma}_{v,s}^2 \approx \widehat{\sigma}_{v,l}^2$ for all s, l, and hence $F \approx \frac{1}{S}\sum_{s=1}^{S} F_{\pi_s}$. Then the weights are given by $w_{2sls,s} \approx \frac{F_{\pi_s}}{\sum_{l=1}^{S} F_{\pi_l}} \approx \frac{F_{\pi_s}}{SF}$, so we see that then the groups with the larger individual F-statistics get the larger weights in the 2SLS estimator under homoskedasticity.

This is not necessarily the case under heteroskedasticity. For two groups with equal value of the F-statistic, the group with the larger variance gets the larger weight, and indeed, a large variance weakly identified group could dominate the 2SLS estimator. As shown in the Monte Carlo exercises below, this is exactly what happens in the design of Andrews (2018). The robust F-statistic is large because one of the groups has a large value of the individual F-statistic. However, this group has a very small variance $\sigma^2_{v,s}$ and hence gets a small weight in the 2SLS estimator, resulting in a poor performance of the estimator in terms of (relative) bias and Wald test size.

4 Alternative GMM Estimator

Clearly, one would like to use an estimator that gives larger weights to more strongly identified groups, independent of the value of $\sigma^2_{v,s}$, mimicking the weights of the 2SLS estimator under homoskedasticity of the first-stage errors. This is achieved by the following GMM estimator, denoted GMMf, with the extension f for first-stage,

$$\begin{aligned}
\widehat{\beta}_{gmmf} &= \left(x'Z\widehat{\Omega}_v^{-1}Z'x\right)^{-1} x'Z\widehat{\Omega}_v^{-1}Z'y \\
&= \left(\widehat{\pi}'Z'Z\widehat{\Omega}_v^{-1}Z'Z\widehat{\pi}\right)^{-1} \widehat{\pi}'Z'Z\widehat{\Omega}_v^{-1}Z'y,
\end{aligned}$$

with $\widehat{\Omega}_v = \sum_{i=1}^n \widehat{v}_i^2 z_i z_i'$ as defined in (3). This looks like the usual GMM estimator, but instead of the structural residuals $\widehat{u}$, the first-stage residuals $\widehat{v}$ are used in the weight matrix. It clearly links directly to the robust F-statistic, as the denominator is equal to SF_r.

It follows that

$$\begin{aligned}
\widehat{\beta}_{gmmf} &= \frac{\sum_{s=1}^S n_s \overline{x}_s \overline{y}_s / \widehat{\sigma}^2_{v,s}}{\sum_{s=1}^S n_s \overline{x}_s^2 / \widehat{\sigma}^2_{v,s}} = \frac{\sum_{s=1}^S \left(n_s \overline{x}_s^2 / \widehat{\sigma}^2_{v,s}\right) \widehat{\beta}_s}{\sum_{s=1}^S n_s \overline{x}_s^2 / \widehat{\sigma}^2_{v,s}} \\
&= \sum_{s=1}^S w_{gmmf,s} \widehat{\beta}_s,
\end{aligned} \tag{7}$$

with

$$w_{gmmf,s} = \frac{F_{\pi_s}}{\sum_{l=1}^S F_{\pi_l}} = \frac{F_{\pi_s}}{SF_r},$$

and hence the groups with the larger F-statistics get the larger weights, independent of the values of $\sigma^2_{v,s}$, mimicking the 2SLS weights under homoskedasticity of the first-stage errors.

5 Weak Instrument Asymptotics

We can formalise the results obtained above further using weak instruments asymptotics (WIA). For each group $s = 1, ..., S$ define

$$\pi_s = \frac{c_s}{\sqrt{n_s}}.$$

The limit of the group specific concentration parameters are then given by

$$\mu_s^2 = \frac{c_s^2}{\sigma_{v,s}^2}. \tag{8}$$

Then

$$\widehat{\pi}_s = \overline{x}_s = \frac{1}{n_s}\sum_{j=1}^{n_s}\left(\frac{c_s}{\sqrt{n_s}} + v_{js}\right) = \frac{c_s}{\sqrt{n_s}} + \overline{v}_s,$$

and

$$\begin{aligned} n_s\overline{x}_s^2 &= n_s\left(\frac{c_s}{\sqrt{n_s}} + \overline{v}_s\right)^2 \\ &= \left(c_s^2 + 2c_s\sqrt{n_s}\overline{v}_s + \left(\sqrt{n_s}\overline{v}_s\right)^2\right) \\ &\xrightarrow{d} \left(c_s + \sigma_{v,s}a_s\right)^2 = \sigma_{v,s}^2\left(\mu_s + a_s\right)^2 \end{aligned}$$

where $\mu_s = c_s/\sigma_{v,s}$ and $\alpha_s \sim N(0,1)$. We get the standard WIA result that

$$F_{\pi_s} = \frac{n_s\overline{x}_s^2}{\widehat{\sigma}_{v,s}^2} \xrightarrow{d} \left(\mu_s + a_s\right)^2 \sim \chi^2_{1,\mu_s^2},$$

where χ^2_{1,μ_s^2} is the non-central chi-squared distribution with 1 degree of freedom and non-centrality parameter μ_s^2.

From (5) it then follows that

$$w_{2sls,s} \xrightarrow{d} \frac{\sigma_{v,s}^2\left(\mu_s + a_s\right)^2}{\sum_{l=1}^S \sigma_{v,l}^2\left(\mu_l + a_l\right)^2},$$

with the a_l independent $N(0,1)$ variables, for $l = 1, ..., S$.

For the weights $w_{gmmf,s}$,

$$w_{gmmf,s} \xrightarrow{d} \frac{\left(\mu_s + a_s\right)^2}{\sum_{l=1}^S\left(\mu_l + a_l\right)^2}.$$

Consider for illustration the case where there are two groups. Table 1 presents some results for the average values of $w_{2sls,1}$ and $w_{gmmf,1}$ after randomly drawing 100,000 values of a_1 and a_2.

Table 1. WIA weights for 2SLS and GMMf

$\sigma^2_{v,1}$	$\sigma^2_{v,2}$	μ^2_1	μ^2_2	$w_{2sls,1}$	$w_{gmmf,1}$
5	5	5.76	5.76	0.50	0.50
5	0.1	5.76	5.76	0.95	0.50
5	0.1	1.96	5.76	0.84	0.32

Notes: Average weights from $100,000$ draws of a_1 and a_2.

If there is homoskedasticity, $\sigma^2_{v,1} = \sigma^2_{v,2}$, and both groups have equal concentration parameters, $\mu^2_1 = \mu^2_2$, then $E(w_{2sls,1}) = E(w_{gmmf,1}) = 0.5$, both estimators will give on average equal weight to the group specific estimators. Next consider the case where there is a large difference in the variances, $\sigma^2_{v,1} = 5$, and $\sigma^2_{v,2} = 0.1$, and $\mu^2_1 = \mu^2_2 = 5.76$, which is the value of the concentration parameter for the group specific Wald tests to have a maximal rejection frequency of 10% at the 5% level. We find for this case that $E(w_{2sls,1}) = 0.95$, i.e. almost all weight will on average be given to the high variance group 1. The expected weight for the GMMf estimator is in this case not affected by the relative values of the $\sigma^2_{v,s}$ and remains at $E(w_{gmmf,1}) = 0.5$. If we subsequently reduce the value of c_1 such that $\mu^2_1 = 1.96$, then $E(w_{2sls,1}) = 0.84$, i.e. the 2SLS estimator will give more weight to $\widehat{\beta}_1$, the estimator in the group with the smaller concentration parameter, but larger variance. In contrast, $E(w_{gmmf,1}) = 0.32$ for this case, giving less weight to the less informative group.

6 Variance of u

So far, we have focused on the first-stage heteroskedasticity, with the robust GMMf estimator exploiting the first-stage information by assigning larger weights to the groups with larger group specific concentration parameters independent of the values of $\sigma^2_{v,s}$. Next, consider the infeasible robust GMM group IV estimator, given by

$$\begin{aligned}\widehat{\beta}_{gmm} &= \frac{\sum_{s=1}^{S} n_s \overline{x}_s \overline{y}_s / \sigma^2_{u,s}}{\sum_{s=1}^{S} n_s \overline{x}^2_s / \sigma^2_{u,s}} = \frac{\sum_{s=1}^{S} \left(n_s \overline{x}^2_s / \sigma^2_{u,s}\right) \widehat{\beta}_s}{\sum_{s=1}^{S} n_s \overline{x}^2_s / \sigma^2_{u,s}} \\ &= \sum_{s=1}^{S} w_{gmm,s} \widehat{\beta}_s.\end{aligned}$$

Whereas $\widehat{\beta}_{gmm}$ is the best, normal, consistent and efficient estimator under standard asymptotics, from the analysis above, it is clear that the weights may not be optimal under WIA. We have under WIA that

$$\frac{n_s \overline{x}_s^2}{\sigma_{u,s}^2} \xrightarrow{d} \frac{\sigma_{v,s}^2}{\sigma_{u,s}^2} \left(\mu_s + a_s\right)^2$$

and so

$$w_{gmm,s} \xrightarrow{d} \frac{\frac{\sigma_{v,s}^2}{\sigma_{u,s}^2} \left(\mu_s + a_s\right)^2}{\sum_{l=1}^{S} \frac{\sigma_{v,l}^2}{\sigma_{u,l}^2} \left(\mu_s + a_s\right)^2}.$$

Clearly, if u is homoskedastic, $\sigma_{u,s}^2 = \sigma_{u,l}^2$ for all s, l, then the infeasible GMM estimator has the same WIA limiting distribution as the 2SLS estimator and suffers from the same problems as described above for 2SLS. If $\sigma_{u,s}^2 = \kappa \sigma_{v,s}^2$ for all s then $\widehat{\beta}_{gmm}$ behaves like the GMMf estimator, the latter in that case also the efficient estimator under standard asymptotics. For other cases the behaviour of $\widehat{\beta}_{gmm}$ clearly depends on whether $\sigma_{v,s}^2/\sigma_{u,s}^2$ assigns relatively larger or smaller weights to the more informative groups.

An alternative is to weight by $\sigma_{u,s}^2 \sigma_{v,s}^2$, such that

$$\begin{aligned} \widehat{\beta}_{gmmuf} &= \sum_{s=1}^{S} w_{gmmuf,s} \widehat{\beta}_s, \\ w_{gmmuf,s} &= \frac{n_s \overline{x}_s^2 / \left(\sigma_{u,s}^2 \sigma_{v,s}^2\right)}{\sum_{l=1}^{S} n_l \overline{x}_l^2 / \left(\sigma_{u,s}^2 \sigma_{v,s}^2\right)} \xrightarrow{d} \frac{\frac{1}{\sigma_{u,s}^2} \left(\mu_s + a_s\right)^2}{\sum_{l=1}^{S} \frac{1}{\sigma_{u,l}^2} \left(\mu_s + a_s\right)^2}. \end{aligned}$$

The resulting weights are then as for the standard GMM estimator under first-stage homoskedasticity. This would clearly improve efficiency if $\sigma_{u,s}^2$ is relatively small for the more informative groups, but can assign again less weight to more informative groups if their values of $\sigma_{u,s}^2$ are relatively large.

7 Some Monte Carlo Results

We consider here the heteroskedastic design of Andrews (2018) with 10 groups, $\beta = 0$ and moderate endogeneity. Results for the high endogeneity case are given in the Appendix. We multiply the first-stage parameters by 0.04, such that the value of the robust F_r is just over 80 on average for 10,000 replications and sample size $n = 10,000$. The group sizes are equal in expectation with $Pr\left\{z_i = e_s\right\} = \frac{1}{S}$ for all $s \in \{1, ..., S\}$.

Table 2 presents the estimation results. The non-robust F-statistic is small, $F = 1.41$ and the effective F-statistic of Olea and Pflueger (2013), denoted F_{eff}, is equal to the non-robust F in this group IV design. Although the robust F-statistic is large, $F_r = 80.2$, the 2SLS estimator $\widehat{\beta}_{2sls}$ is poorly behaved. Its relative bias equal to 0.699 and the Wald test rejection frequency for $H_0 : \beta = 0$ is equal to 0.534 at the 5% level. In contrast, the GMMf estimator is unbiased and its Wald test rejection frequency equal to 0.049 at the 5% level.

Table 2. Estimation results for $S = 10$, moderate endogeneity

F	F_{eff}	F_r	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{2sls}$	$\widehat{\beta}_{gmmf}$	W_{2sls}	W_{gmmf}
1.411	1.411	80.23	-0.608	-0.424	-0.001	0.534	0.049
			(0.011)	(0.257)	(0.563)		

Notes: means and (st. dev.) of $10,000$ replications. Rej. freq. of robust Wald tests W at 5% level.

The details as given in Table 3 make clear what is happening. It reports the fixed values of π_s, $\sigma^2_{v,s}$ and $\mu^2_{n,s} = 1000\pi^2_s/\sigma^2_{v,s}$ and the mean values of F_{π_s}, $w_{2sls,s}$ and $w_{gmmf,s} = F_{\pi_s}/\sum_{l=1}^{S} F_{\pi_l}$. Identification in the first group is strong, with an average value of $F_{\pi_1} = 789.5$, whereas identification in all other 9 groups is very weak, with the largest average value for $F_{\pi_5} = 2.23$. But the variance in group 1 is very small, and some of the variances in the other groups are quite large. This leads to the low average value of $w_{2sls,1} = 0.127$, showing that the 2SLS estimator doesn't utilise the identification strength of the first group, with larger weight given to higher variance, but lower concentration parameter groups.

Table 3. Group information and estimator weights

s	1	2	3	4	5	6	7	8	9	10
π_s	0.058	-0.023	0.049	0.015	0.022	0.008	-0.017	0.011	-0.036	-0.040
$\sigma^2_{v,s}$	0.004	2.789	4.264	0.779	0.395	7.026	1.226	0.308	1.709	6.099
$\mu^2_{n,s}$	785.7	0.184	0.556	0.284	1.190	0.009	0.236	0.387	0.770	0.266
F_{π_s}	789.5	1.170	1.564	1.279	2.225	0.997	1.203	1.372	1.798	1.246
$w_{2sls,s}$	0.126	0.098	0.178	0.035	0.031	0.180	0.049	0.015	0.096	0.192
$w_{gmmf,s}$	0.984	0.002	0.002	0.002	0.003	0.001	0.002	0.002	0.002	0.002

Notes: $\mu^2_{n,s} = 1000\pi^2_s/\sigma^2_{v,s}$.

Table 3 further shows that for the GMMf estimator almost all weight is given to the first group, with the average of $w_{gmmf,1}$ equal to 0.984, resulting in the good behaviour of the GMMf estimator in terms of bias and Wald test size. In this case the standard deviation of the GMMf estimator is quite large relative to that of the 2SLS estimator. This is driven by the value of $\sigma^2_{u,1}$, which in this design is equal to 1.10, much larger than $\sigma^2_{v,1}$. Reducing the value of $\sigma^2_{u,1}$ (and the value for $\sigma_{uv,1}$ accordingly to keep the same correlation structure within group 1), will reduce the standard deviation of the GMMf estimator.

Figure 1 displays the rejection frequencies of the robust Wald tests for testing $H_0 : \beta = 0$ for varying values of the robust F-statistic F_r for the 2SLS and GMMf estimators. Different values of F_r are obtained by different values of d when setting the first-stage parameters $\pi = d\pi_0$. It is clear that the Wald test based on the GMMf estimator is much better behaved in terms of size than that based on the 2SLS estimator, with hardly any size distortion for mean values of F_r larger than 5. Figure 2 shows that the bias of the GMMf estimator, relative to that of the OLS estimator, is also a lot smaller than that of the 2SLS estimator, with the relative bias smaller than 0.10 for mean values of F_r larger than 9.

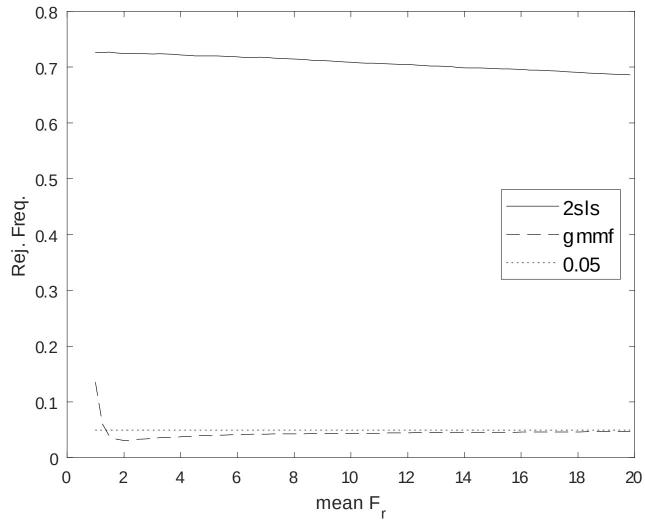

Figure 1. Rejection frequencies of robust Wald tests

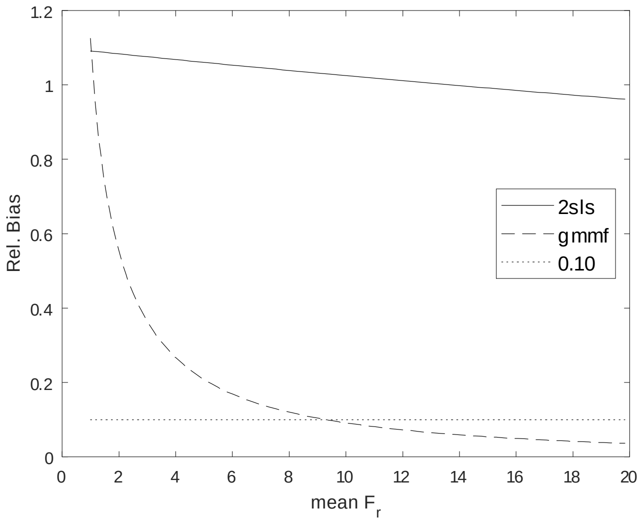

Figure 2. Relative bias

8 Testing for Weak Instruments

Using the GMMf estimator as a generalisation of the 2SLS estimator to deal with general forms of first-stage heteroskedasticity, we derive in the Appendix under what conditions the weak instruments Stock and Yogo (2005) critical values derived for the non-robust F-test and the properties of the 2SLS estimator under full homoskedasticity apply to the robust F-test and the properties of the GMMf estimator. We focus here on standard cross-sectional heteroskedasticity, but results apply to cluster and/or serially correlated designs.

Consider again the standard linear model

$$\begin{aligned} y_i &= x_i\beta + u_i; \\ x_i &= z_i'\pi + v_i, \end{aligned}$$

where z_i is a k_z-vector of instruments, and where other exogenous variables, including the constant have been partialled out. General conditional heteroskedasticity is specified as

$$\begin{aligned} E\left[u_i^2|z_i\right] &= \sigma_u^2\left(z_i\right); \\ E\left[v_i^2|z_i\right] &= \sigma_v^2\left(z_i\right); \\ E\left[u_iv_i|z_i\right] &= \sigma_{uv}\left(z_i\right). \end{aligned}$$

Further, let

$$\Omega_u = E\left[\sigma_u^2(z_i)\, z_i z_i'\right]; \quad \Omega_v = E\left[\sigma_v^2(z_i)\, z_i z_i'\right]; \quad \Omega_{uv} = E\left[\sigma_{uv}(z_i)\, z_i z_i'\right],$$

and the unconditional variances and covariance

$$\sigma_u^2 = E_z\left[\sigma_u^2(z_i)\right]; \quad \sigma_v^2 = E_z\left[\sigma_v^2(z_i)\right]; \quad \sigma_{uv} = E_z\left[\sigma_{uv}(z_i)\right].$$

The robust F-statistic and GMMf estimator are given by

$$\begin{aligned} F_r &= x'Z\widehat{\Omega}_v^{-1}Z'x/k_z \\ \widehat{\beta}_{gmmf} &= \left(x'Z\widehat{\Omega}_v^{-1}Z'x\right)^{-1} x'Z\widehat{\Omega}_v^{-1}Z'y \end{aligned}$$

Stock and Yogo (2005) derived critical values for the non-robust F-statistic under homoskedasticity for the weak instruments hypothesis on the relative bias of the 2SLS estimator, relative to that of the OLS estimator. In Appendix A.2 we show that these critical values apply to the robust F-statistic for relative bias of the GMMf estimator, relative to that of the OLS estimator if $\Omega_{uv} = \delta\Omega_v$ and $\sigma_{uv} = \delta\sigma_v^2$.

For the Wald test size distortion, we show in Appendix A.2 that the Stock and Yogo (2005) critical values apply to the GMMf based Wald test if $\Omega_{uv} = \delta\Omega_v$ and $\Omega_u = \kappa\Omega_v$, the latter implying that the GMMf estimator is also the efficient estimator under standard asymptotics.

Whilst these conditions imply a limited applicability of the Stock and Yogo (2005) critical values for the robust F-statistic in relation to the behaviour of the GMMf estimator, it is a generalisation of, and includes, the homoskedastic case. It also encompasses the illustrative example of Olea and Pflueger (2013, Section 3.1), where they considered a design with $E\left[(u_i\ v_i)'(u_i\ v_i)\right] = \Sigma$ and $E\left[\left((u_i\ v_i)'(u_i\ v_i)\right) \otimes z_i z_i'\right] = a^2\Sigma \otimes I_{k_z}$, and where the non-robust F-statistic gives an overestimate of the information content for the 2SLS estimator when $a > 1$.

9 Conclusions

This paper has shown why large values of the first-stage robust F-statistic may not translate in good behaviour of the 2SLS estimator. In the heteroskedastic grouped data design of Andrews (2018), this is the case because a highly informative group had a relatively small first-stage variance, and the 2SLS estimator gives more weight to groups with small concentration parameters but large first-stage variances. A robust GMM estimator, called GMMf, with the robust weight matrix estimated using the first-stage residuals, remedies this problem and gives larger weights to more informative groups. This is independent of the values of the first-stage variances and is a generalisation of the 2SLS estimator in that it mimics what the 2SLS estimator does under first-stage homoskedasticity. A large value of the robust F-statistic indicates that there is first-stage information resulting in a well behaved GMMf estimator. We have provided the conditions under which the Stock and Yogo (2005) weak instruments critical values developed for the non-robust F-statistic and relative bias and Wald test size distortion of the 2SLS estimator apply to the robust F-statistic and the behaviour of the GMMf estimator.

References

Andrews, I., 2018, Valid two-step identification-robust confidence sets for GMM. *Review of Economics and Statistics* 100, 337-348.

Angrist, J.D., 1991, Grouped-data estimation and testing in simple labor-supply models, *Journal of Econometrics* 47, 243-266.

Angrist, J.D. and J.-S. Pischke, 2009, *Mostly Harmless Econometrics. An Empiricist's Companion*, Princeton University Press, Princeton and Oxford.

Bekker, P.A. and J. van der Ploeg, 2005, Instrumental variable estimation based on grouped data, *Statistica Neerlandica* 59, 239-267.

Bun, M.J.G. and M. de Haan, 2010. Weak instruments and the first stage F-statistic in IV models with a nonscalar error covariance structure. Discussion Paper 2010/02, UvA Econometrics, University of Amsterdam.

Staiger, D. and J.H. Stock, 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

Stock, J.H. and M. Yogo, 2005. Testing for weak instruments in linear IV regression. In D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg*, 80-108. New York: Cambridge University Press.

Olea, J.L.M. and C. Pflueger, 2013, A robust test for weak instruments. *Journal of*

Business & Economic Statistics 31, 358-369.

Windmeijer, F., 2018, Two-stage least squares as minimum distance. *The Econometrics Journal*, in press, https://doi.org/10.1111/ectj.12115

Appendix

A.1 Results for high endogeneity design

Tables A1 and A2 present estimation results for the $s = 10$, high endogeneity design of Andrews (2018). As in Section 7, the first stage parameters have been multiplied by a factor such that the robust F-statistic has an average value of just over 80. As shown in Table A1, the pattern of group information is similar to that in that in the moderate endogeneity case, with one informative group, group $s = 10$, with an average value of $F_{\pi_{10}} = 792.2$. However, the variance $\sigma^2_{v,10}$ is now so small in relative terms, that the 2SLS weight for group 10 has an average value of only $w_{2sls,10} = 0.003$. The GMMf estimator corrects this, with the average value of $w_{gmmf,10} = 0.989$, and is again much better behaved than the 2SLS estimator both in terms of (relative) bias and Wald test size, as displayed in Table A2.

Table A1. Group information and estimator weights, high endogeneity

s	1	2	3	4	5	6	7	8	9	10
$100 \cdot \pi_s$	-0.021	0.095	-0.484	-0.069	0.159	-0.028	0.101	-0.418	0.450	-0.546
$\sigma^2_{v,s}$	1.600	0.478	2.975	1.142	0.174	0.145	4.658	1.963	2.990	$0.38 \cdot a$
$\mu^2_{n,s}$	$0.28 \cdot a$	0.002	0.008	$4.2 \cdot a$	0.015	$5.6 \cdot a$	$2.2 \cdot a$	0.009	0.007	789.9
F_{π_s}	0.998	1.017	0.979	1.010	1.034	0.984	0.977	1.031	0.997	792.2
$w_{2sls,s}$	0.111	0.040	0.177	0.085	0.016	0.013	0.242	0.134	0.181	0.003
$w_{gmmf,s}$	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.989

Notes: $a = 10^{-4}$; $\mu^2_{n,s} = 1000\pi^2_s/\sigma^2_{v,s}$.

Table A2. Estimation results for $S = 10$, high endogeneity

F	F_{eff}	F_r	$\widehat{\beta}_{OLS}$	$\widehat{\beta}_{2sls}$	$\widehat{\beta}_{gmmf}$	W_{2sls}	W_{gmmf}
0.994	0.994	80.12	0.754	0.749	0.007	1.000	0.067
			(0.000)	(0.023)	(0.029)		

Notes: means and (st. dev.) of 10,000 replications. Rej. freq. of robust Wald tests W at 5% level.

Figures A1 and A2 show respectively the rejection frequencies of the robust Wald tests and the relative bias of the 2SLS and GMMf estimators as a function of the value of the robust F-statistic, showing a much better performance of the GMMf estimator.

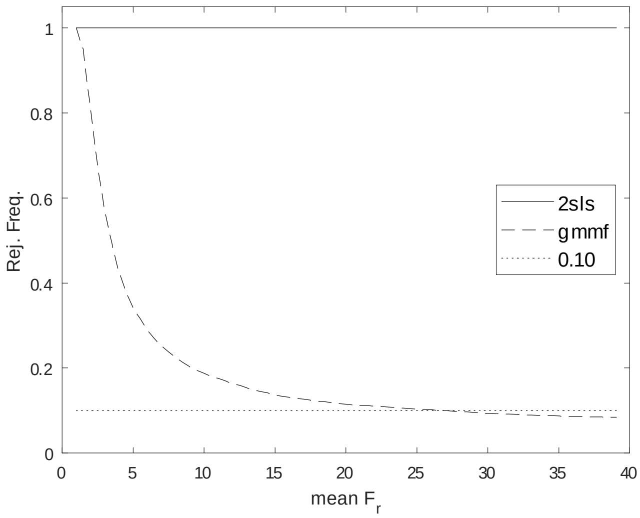

Figure A1. Rejection frequencies of robust Wald tests, high endogeneity

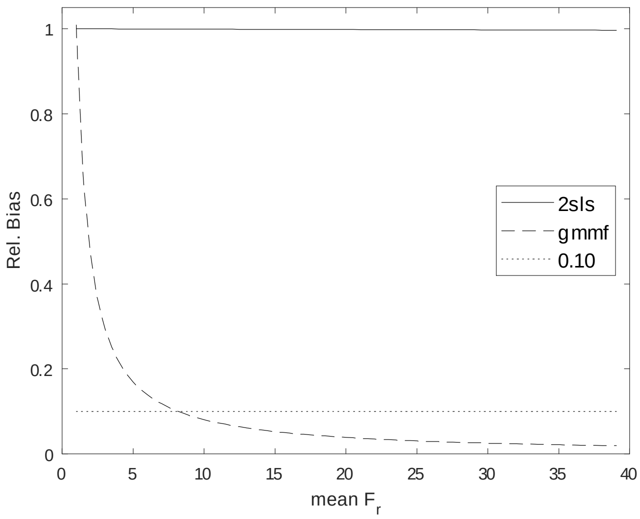

Figure A2. Relative bias, high endogeneity

A.2 Testing for Weak Instruments

Using the GMMf estimator as a generalisation of the 2SLS estimator to deal with general forms of first-stage heteroskedasticity, we investigate here under what conditions the

Stock and Yogo (2005) weak instruments critical values derived for the non-robust F-test and the properties of 2SLS estimator under full homoskedasticity apply to the robust F-test and the properties of the GMMf estimator.

Consider again the standard linear model

$$\begin{aligned} y_i &= x_i \beta + u_i ; \\ x_i &= z_i' \pi + v_i , \end{aligned}$$

with conditional heteroskedasticity specified as

$$\begin{aligned} E\left[u_i^2 | z_i\right] &= \sigma_u^2\left(z_i\right); \\ E\left[v_i^2 | z_i\right] &= \sigma_v^2\left(z_i\right); \\ E\left[u_i v_i | z_i\right] &= \sigma_{uv}\left(z_i\right), \end{aligned}$$

and, unconditionally,

$$\sigma_u^2 = E_z\left[\sigma_u^2\left(z_i\right)\right]; \quad \sigma_v^2 = E_z\left[\sigma_v^2\left(z_i\right)\right]; \quad \sigma_{uv} = E_z\left[\sigma_{uv}\left(z_i\right)\right].$$

Further, let

$$\Omega_u = E\left[\sigma_u^2\left(z_i\right) z_i z_i'\right]; \quad \Omega_v = E\left[\sigma_v^2\left(z_i\right) z_i z_i'\right]; \quad \Omega_{uv} = E\left[\sigma_{uv}\left(z_i\right) z_i z_i'\right],$$

and assume that

$$\begin{aligned} \begin{pmatrix} \frac{1}{\sqrt{n}} Z' u \\ \frac{1}{\sqrt{n}} Z' v \end{pmatrix} &\xrightarrow{d} \begin{pmatrix} \psi_{zu} \\ \psi_{zv} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_u & \Omega_{uv} \\ \Omega_{uv}' & \Omega_v \end{pmatrix}\right) \\ \begin{pmatrix} \Omega_u^{-1/2} \frac{1}{\sqrt{n}} Z' u \\ \Omega_v^{-1/2} \frac{1}{\sqrt{n}} Z' v \end{pmatrix} &\xrightarrow{d} \begin{pmatrix} z_u \\ z_v \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{k_z} & R \\ R' & I_{k_z} \end{pmatrix}\right), \end{aligned}$$

where

$$R = \Omega_u^{-1/2} \Omega_{uv} \Omega_v^{-1/2}.$$

Let $\widehat{\pi} = \left(Z'Z\right)^{-1} Z'x$ be the OLS estimator of π and $\widehat{v} = x - Z\widehat{\pi}$ the OLS residual. For the GMMf estimator, we have

$$\widehat{\beta}_{gmmf} = \beta + \left(x' Z \widehat{\Omega}_v^{-1} Z' x\right)^{-1} x' Z \widehat{\Omega}_v^{-1} Z' u$$

where, as before, $\widehat{\Omega}_v = \sum_{i=1}^n \widehat{v}_i^2 z_i z_i'$. Assume that conditions are such that

$$\frac{1}{n} Z'Z \xrightarrow{p} E\left[z_i z_i'\right] = Q_{zz}$$

$$\frac{1}{n} \widehat{\Omega}_v \xrightarrow{p} \Omega_v = E\left[\sigma_v^2\left(z_i\right) z_i z_i'\right].$$

For weak instrument asymptotics, let

$$\pi = \frac{c}{\sqrt{n}},$$

then

$$\begin{aligned}
x'Z\widehat{\Omega}_v^{-1}Z'x &= \left(Z\frac{c}{\sqrt{n}} + v\right)' Z\widehat{\Omega}_v^{-1}Z'\left(Z\frac{c}{\sqrt{n}} + v\right) \\
&= \frac{1}{n}c'Z'Z\widehat{\Omega}_v^{-1}Z'Zc + \frac{2}{\sqrt{n}}c'Z'Z\widehat{\Omega}_v^{-1}Z'v + v'Z\widehat{\Omega}_v^{-1}Z'v \\
&\xrightarrow{d} \left(\lambda + z_v\right)'\left(\lambda + z_v\right)
\end{aligned}$$

where

$$\lambda = \Omega_v^{-1/2} Q_{zz} c.$$

It follows that

$$F_r = \frac{1}{k_z} x'Z\widehat{\Omega}_v^{-1}Z'x \xrightarrow{d} \chi_{k_z}^2\left(\lambda'\lambda\right)/k_z. \tag{A.1}$$

For the numerator, we have

$$\begin{aligned}
x'Z\widehat{\Omega}_v^{-1}Z'u &= \left(Z\frac{c}{\sqrt{n}} + v\right)' Z\widehat{\Omega}_v^{-1}Z'u \\
&= \frac{2}{\sqrt{n}}c'Z'Z\widehat{\Omega}_v^{-1}Z'u + v'Z\widehat{\Omega}_v^{-1}Z'u \\
&\xrightarrow{d} \left(\lambda + z_v\right)'\Omega_v^{-1/2}\Omega_u^{1/2} z_u.
\end{aligned}$$

For the OLS estimator,

$$\widehat{\beta}_{ols} - \beta = \frac{x'u}{x'x} = \frac{\frac{c'}{\sqrt{n}}Z'u + v'u}{\frac{1}{n}c'Z'Zc + \frac{2}{\sqrt{n}}c'Z'v + v'v} \xrightarrow{p} \frac{\sigma_{uv}}{\sigma_v^2}.$$

As $E\left[z_u | z_v\right] = Rz_v$, it follows for the relative bias that

$$\frac{E\left[\widehat{\beta}_{gmmf} - \beta\right]}{E\left[\widehat{\beta}_{ols} - \beta\right]} \rightarrow \frac{\sigma_v^2}{\sigma_{vu}} E\left[\frac{\left(\lambda + z_v\right)'\Omega_v^{-1/2}\Omega_u^{1/2}Rz_v}{\left(\lambda + z_v\right)'\left(\lambda + z_v\right)}\right] \tag{A.2}$$

$$= \frac{\sigma_v^2}{\sigma_{uv}} E\left[\frac{(\lambda + z_v)' \,\Omega_v^{-1/2}\Omega_{uv}\Omega_v^{-1/2} z_v}{(\lambda + z_v)' \,(\lambda + z_v)}\right].$$

Therefore it follows that if $\Omega_{uv} = \delta\Omega_v$ and $\sigma_{uv} = \delta\sigma_v^2$, then

$$\frac{E\left[\widehat{\beta}_{gmmf} - \beta\right]}{E\left[\widehat{\beta}_{ols} - \beta\right]} \to E\left[\frac{(\lambda + z_v)' \, z_v}{(\lambda + z_v)' \,(\lambda + z_v)}\right]. \tag{A.3}$$

The conditions $\Omega_{uv} = \delta\Omega_v$ and $\sigma_{uv} = \delta\sigma_v^2$ are satisfied if $\sigma_{uv}(z_i) = \delta\sigma_v^2(z_i)$ for all $z_i \in \mathcal{Z}$.

The results (A.1) and (A.3) are the same as the Staiger and Stock (1997) and Stock and Yogo (2005) results for the 2SLS estimator and full conditional homoskedasticity,

$$\begin{pmatrix} \psi_{zu} \\ \psi_{zv} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \otimes Q_{zz}\right)$$

and with $\lambda = \sigma_v^{-1} Q_{zz}^{1/2} c$. Therefore the Stock and Yogo (2005) critical values apply to the robust F-statistic and relative bias (A.2) of the GMMf estimator if $\Omega_{uv} = \delta\Omega_v$ and $\sigma_{uv} = \delta\sigma_v^2$. For the grouped data IV example, this condition is fulfilled if $\sigma_{uv,s} = \delta\sigma_{v,s}^2$ for all $s = 1, ..., S$.

For the Wald test, we have

$$V\hat{a}r\left(\widehat{\beta}_{gmmf}\right) = \left(x'Z\widehat{\Omega}_v^{-1}Z'x\right)^{-1} x'Z\widehat{\Omega}_v^{-1}\widehat{\Omega}_u\widehat{\Omega}_v^{-1}Z'x\left(x'Z\widehat{\Omega}_v^{-1}Z'x\right)^{-1}$$

and so

$$W_{gmmf}(\beta) = \frac{\left(\widehat{\beta}_{gmmf} - \beta\right)^2 \left(x'Z\widehat{\Omega}_v^{-1}Z'x\right)^{-2}}{x'Z\widehat{\Omega}_v^{-1}\widehat{\Omega}_u\widehat{\Omega}_v^{-1}Z'x} = \frac{\left(x'Z\widehat{\Omega}_v^{-1}Z'u\right)^2}{x'Z\widehat{\Omega}_v^{-1}\widehat{\Omega}_u\widehat{\Omega}_v^{-1}Z'x}.$$

Then,

$$\begin{aligned}
\frac{1}{n}\widehat{\Omega}_u &= \frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 z_i z_i' \\
&= \frac{1}{n}\sum_{i=1}^n \left(u_i - x_i\left(\widehat{\beta}_{gmmf} - \beta\right)\right)^2 z_i z_i' \\
&= \frac{1}{n}\sum_{i=1}^n \left(u_i^2 - 2u_i x_i\left(\widehat{\beta}_{gmmf} - \beta\right) + x_i^2\left(\widehat{\beta}_{gmmf} - \beta\right)^2\right) z_i z_i' \\
&\xrightarrow{d} \Omega_u - 2\Omega_{uv}\left(\widehat{\beta}_{gmmf} - \beta\right) + \Omega_v\left(\widehat{\beta}_{gmmf} - \beta\right)^2,
\end{aligned}$$

and so

$$
\begin{aligned}
& x' Z \widehat{\Omega}_v^{-1} \widehat{\Omega}_u \widehat{\Omega}_v^{-1} Z' x \xrightarrow{d} (\lambda + z_v)' \, \Omega_v^{-1/2} \Omega_u \Omega_v^{-1/2} (\lambda + z_v) \\
& -2 (\lambda + z_v)' \, \Omega_v^{-1/2} \Omega_{uv} \Omega_v^{-1/2} (\lambda + z_v) \left(\widehat{\beta}_{gmmf} - \beta \right) \\
& + (\lambda + z_v)' (\lambda + z_v) \left(\widehat{\beta}_{gmmf} - \beta \right)^2 .
\end{aligned}
$$

This results in

$$
W_{gmmf} \xrightarrow{d} \frac{q_2^2}{a - 2bq_2/\eta_1 + q_2^2/\eta_1},
$$

where

$$
\begin{aligned}
q_2 &= (\lambda + z_v)' \, \Omega_v^{-1/2} \Omega_u^{1/2} z_u \\
a &= (\lambda + z_v)' \, \Omega_v^{-1/2} \Omega_u \Omega_v^{-1/2} (\lambda + z_v) \\
b &= (\lambda + z_v)' \, \Omega_v^{-1/2} \Omega_{uv} \Omega_v^{-1/2} (\lambda + z_v) \\
\eta_1 &= (\lambda + z_v)' (\lambda + z_v) .
\end{aligned}
$$

If $\Omega_u = \kappa \Omega_v$ and $\Omega_{uv} = \delta \Omega_v$, so $R = \frac{\delta}{\sqrt{\kappa}} I_{k_z}$, then

$$
W_{gmmf}(\beta) \xrightarrow{d} \frac{\kappa \eta_2^2}{\kappa \eta_1 - 2\delta \sqrt{\kappa} \eta_2 + \kappa \eta_2^2/\eta_1} = \frac{\eta_2^2/\eta_1}{1 - 2\rho \eta_2/\eta_1 + (\eta_2/\eta_1)^2} \tag{A.4}
$$

where

$$
\begin{aligned}
\eta_2 &= (\lambda + z_v)' z_u \\
\rho &= \frac{\delta}{\sqrt{\kappa}}.
\end{aligned}
$$

Conditions $\Omega_u = \kappa \Omega_v$ and $\Omega_{uv} = \delta \Omega_v$ are satisfied if $\sigma_{uv}(z_i) = \delta \sigma_v^2(z_i)$ and $\sigma_u^2(z_i) = \kappa \sigma_v^2(z_i)$ for all $z_i \in \mathcal{Z}$, and then $\rho = \frac{\sigma_{uv}}{\sigma_u \sigma_v}$.

Result (A.4) is the same as that of Staiger and Stock (1997) and Stock and Yogo (2005) result for the 2SLS based Wald test under conditional homoskedasticity with the maximum size distortion at $\rho^2 = 1$. Hence the Stock and Yogo (2005) Wald size based critical values apply in the heteroskedastic case to the GMMf based Wald test if $\Omega_u = \kappa \Omega_v$ and $\Omega_{uv} = \delta \Omega_v$, with again the maximum size distortion at $\delta^2/\kappa = 1$.