The Robust F-Statistic as a Test for Weak Instruments

Frank Windmeijer

Dept of Statistics and Nuffield College, University of Oxford, UK

August 6, 2024

Abstract

For the linear model with a single endogenous variable, Montiel Olea and Pflueger (2013) proposed the effective F-statistic as a test for weak instruments in terms of the Nagar bias of the two-stage least squares (2SLS) or limited information maximum likelihood (LIML) estimator relative to a benchmark worst-case bias. We show that their methodology for the 2SLS estimator applies to a class of linear generalized method of moments (GMM) estimators with an associated class of generalized effective F-statistics. The standard robust F-statistic is a member of this class. The associated GMMf estimator, with the extension "f" for first-stage, has the weight matrix based on the first-stage residuals. In the grouped-data IV designs of Andrews (2018) with moderate and high levels of endogeneity and where the robust F-statistic is large but the effective F-statistic is small, the GMMf estimator is shown to behave much better in terms of bias than the 2SLS estimator.

Keywords: Instrumental variables, weak instruments, heteroskedasticity, robust F-statistic, GMM

JEL Codes: C12, C26

1 Introduction

It is commonplace to report the first-stage F-statistic to test for weak instruments in linear models with a single endogenous variable, estimated by two-stage least squares (2SLS). 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 distortions. These weak-instruments critical values for the non-robust F-statistic are valid only under homoskedasticity (i.e. conditional homoskedasticity, no serial correlation and no clustering) of both the first-stage and structural errors, and do not apply to the robust F-statistic in general designs, see Bun and de Haan (2010), Montiel Olea and Pflueger (2013) and Andrews (2018). In particular, Andrews (2018) found for some cross-sectional heteroskedastic designs that the standard 2SLS confidence intervals had large coverage distortions even for very large values of the robust F-statistic. For example, he found for a high endogeneity design that "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, Supplementary Appendix, p 11).

For general heteroskedasticity, which we mean to cover the cases of conditional heteroskedasticity, serial correlation and clustering, Montiel Olea and Pflueger (2013) proposed the first-stage effective F-statistic and derived critical values for the null of weak instruments in terms of the Nagar bias of the 2SLS or LIML estimator, relative to an estimator-specific benchmark worst-case bias. We focus here on their results for the 2SLS estimator. As shown in Section 6.1, the effective F-statistics in the designs of Andrews (2018) do not reject the null of weak instruments. In their review paper Andrews, Stock, and Sun (2019, p 729) recommend "that researchers judge instrument strength based on the effective F-statistic of Montiel Olea and Pflueger (2013)".

The effective F-statistic is specific to the Nagar bias of the 2SLS or LIML estimator and the main contribution of this paper is that the 2SLS related methods of Montiel Olea and Pflueger (2013) apply to a wider class of linear generalized method of moments (GMM) estimators resulting in a class of associated generalized effective F-statistics. The robust F-statistic is a member of this class, and we call its associated GMM estimator the GMMf estimator, with the extension "f" for first-stage. This is because the weight matrix of the GMMf estimator is based on the first-stage residuals, with k_z times the robust F-statistic being the denominator of the GMMf estimator, where k_z is the number of excluded instruments. This is similar to the relationship of the non-robust F-statistic and the 2SLS estimator.

In practice, reported 2SLS estimation results with robust standard errors are often accompanied by the robust first-stage F-statistic, as most statistical packages automatically provide these together. Whilst the robust F-statistic can be used as a test for underidentification, it can not be used as a test for weak instruments related to the performance of the 2SLS estimator. Hence, for the single-endogenous variable case, robust 2SLS estimation results should be accompanied by the effective F-statistic and its critical value, as per the advice of Andrews et al. (2019). The critical value of the robust F-statistic in relation to the test for weak instruments for the GMMf estimator could also be computed. If the situation is such, like in the Andrews (2018) example, that the effective F-statistic is small and indicates weak instruments problems for the 2SLS estimator, but the robust F-statistic is large, rejecting the null of weak instruments for the GMMf estimator, then the latter could be preferred and reported.[1] This is illustrated in Section 6.1, where we replicate the Monte Carlo analysis of Andrews (2018). The 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. In the two designs considered, there is in each only one informative group, but the first-stage heteroskedasticity is such that the 2SLS estimator does not utilize this information well, whereas the GMMf estimator gives almost all the weight to the informative groups.

Section 2 introduces the single-endogenous variable linear model specification, main assumptions, effective and robust F-statistics and the GMMf estimator. Section 3 then formulates the class of generalized F-statistics for the class of linear GMM estimators and shows that the weak-instruments testing methods developed by Montiel Olea and Pflueger (2013) apply straightforwardly to this class. The section then provides a summary of the Montiel Olea and Pflueger (2013) results. Section 4 shows how the general results and specifications simplify for the robust-F statistic in relation to the Nagar bias of the GMMf estimator. As the Nagar bias is relative to a benchmark worse-case bias, which is estimator specific, we harmonize in Section 5 the benchmark bias and propose the use of the worst-case OLS bias as the benchmark, which applies to the class of GMM estimators considered.

The weak-instruments test procedures considered here only apply to linear models

[1] An extension of the "weakivtest" command of Pflueger and Wang (2015) in Stata, StataCorp. (2023), called "gfweakivtest" for calculating these is available from the author upon request.

with a single endogenous regressor. Lewis and Mertens (2022) develop an extension of the Montiel Olea and Pflueger (2013) method to the multiple endogenous variable case for the 2SLS estimator, but they do not consider such an extension for the wider class of GMM estimators.

2 Model, Assumptions and F-Statistics

We have a sample $\{y_i, x_i, z_i\}_{i=1}^{n}$, where z_i is a k_z-vector of instrumental variables. We are interested in the effect of x on y in a linear model specification, where x is endogenously determined. We consider the linear structural and first-stage specifications

$$y = x\beta + u \tag{1}$$

$$x = Z\pi + v_2, \tag{2}$$

where y, x, u and v_2 are n-vectors and Z an $n \times k_z$ matrix. Other exogenous explanatory variables, including the constant have been partialled out. The reduced-form specification for y is then given by

$$y = Z\pi\beta + v_1 = Z\pi_y + v_1, \tag{3}$$

where $v_1 = u + \beta v_2$, and $\pi_y = \pi\beta$.

Following Montiel Olea and Pflueger (2013) (henceforth MOP), we make the following assumptions.

Assumption 1.

1. *Weak-instruments asymptotics. The vector π is local to zero,*

$$\pi = \pi_n = c/\sqrt{n},$$

where c is a fixed vector $c \in \mathbb{R}^{k_z}$.

2. As $n \to \infty$,

$$\frac{1}{n}Z'Z \xrightarrow{p} Q_{zz};$$

$$\frac{1}{n}\left[v_1\, v_2\right]'\left[v_1\, v_2\right] \xrightarrow{p} \Sigma_v;$$

$$\frac{1}{\sqrt{n}}\begin{pmatrix} Z'v_1 \\ Z'v_2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \sim N\left(0, W\right),$$

with Q_{zz}, Σ_v *and* W *finite, positive definite matrices, and*

$$\Sigma_v = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix};$$

$$W = \begin{bmatrix} W_1 & W_{12} \\ W_{12}' & W_2 \end{bmatrix}.$$

3. There exists a sequence of positive definite estimates $\left\{\widehat{W}_n\right\}$*, such that* $\widehat{W}_n \xrightarrow{p} W$ *as* $n \to \infty$.

In the remainder, we drop the subscript n from $\widehat{W}_n$ for ease of exposition, and, commensurate with the partitioning of W,

$$\widehat{W} = \begin{bmatrix} \widehat{W}_1 & \widehat{W}_{12} \\ \widehat{W}_{12}' & \widehat{W}_2 \end{bmatrix}.$$

The two-stage least squares (2SLS) estimator is given by

$$\widehat{\beta}_{2sls} = \frac{x'P_Z y}{x'P_Z x},$$

where $P_Z = Z\left(Z'Z\right)^{-1} Z'$. The standard non-robust first-stage F-statistic is

$$\widehat{F} = \frac{x'P_Z x}{k_z \widehat{\sigma}_{v_2}^2},$$

where $\widehat{\sigma}_{v_2}^2 = \widehat{v}_2'\widehat{v}_2/n$, $\widehat{v}_2 = \left(I_n - P_Z\right)x$, where I_n is the identity matrix of order n. Note that we refrain throughout from finite sample degrees-of-freedom corrections in the exposition.

It follows that we can alternatively express the 2SLS estimator as

$$\widehat{\beta}_{2sls} = \frac{\widehat{\pi}' \left(V\widehat{a}r \left(\widehat{\pi} \right) \right)^{-1} \widehat{\pi}_y}{\widehat{\pi}' \left(V\widehat{a}r \left(\widehat{\pi} \right) \right)^{-1} \widehat{\pi}} = \frac{\widehat{\pi}' \left(V\widehat{a}r \left(\widehat{\pi} \right) \right)^{-1} \widehat{\pi}_y}{k_z \widehat{F}}, \tag{4}$$

where $\widehat{\pi}$ and $\widehat{\pi}_y$ are the OLS estimators of π and π_y in the first-stage and reduced-form models (2) and (3), and $V\widehat{a}r\left(\widehat{\pi}\right) = \widehat{\sigma}^2_{v_2}\left(Z'Z\right)^{-1}$, the non-robust estimator of the variance of $\widehat{\pi}$.

$\widehat{F}$ can be used as a test for underidentification and as a test for weak instruments in terms of the bias of the 2SLS estimator relative to that of the OLS estimator of β, or the size distortion of the Wald test for hypotheses on β, Stock and Yogo (2005). The Stock and Yogo critical values are valid only under conditional homoskedasticity of both u and v_2, i.e. $\mathbb{E}\left[u_i|z_i\right] = \sigma^2_u$ and $\mathbb{E}\left[v_{2i}|z_i\right] = \sigma^2_{v_2}$, or $W = \Sigma_v \otimes Q_{zz}$, with $\Sigma_v = \mathbb{E}\left[v_i v_i'\right]$ and $v_i = \left(v_{1i}, v_{2i}\right)'$.

The robust F-statistic is given by

$$\widehat{F}_r = \frac{x' Z \widehat{W}_2^{-1} Z' x}{n k_z} \tag{5}$$

and is a standard test statistic for testing $H_0 : \pi = 0$ under general forms of heteroskedasticity. But $\widehat{F}_r$ cannot be used as a test for weak instruments in relation to the behaviour of the 2SLS estimator, see the discussion in Andrews et al. (2019, pp 738-739), summarized at the end of this section. Andrews (2018) showed in a grouped-data IV design that $\widehat{F}_r$ could take very large values, of the order of 100,000, whereas the 2SLS estimator was still poorly behaved in terms of bias and Wald test size.

MOP proposed the effective F-statistic

$$\begin{aligned}\widehat{F}_{\text{eff}} &= \frac{x' P_Z x}{\operatorname{tr}\left(\left(\frac{1}{n} Z'Z\right)^{-1/2} \widehat{W}_2 \left(\frac{1}{n} Z'Z\right)^{-1/2} \right)} \\ &= \frac{x' P_Z x}{\operatorname{tr}\left(\widehat{W}_2 \left(\frac{1}{n} Z'Z\right)^{-1} \right)},\end{aligned} \tag{6}$$

and showed that this F-statistic can be used as a test for weak instruments in relation to the Nagar (1959) bias of the 2SLS and LIML estimators, relative to a worst-case benchmark. Andrews et al. (2019) advocate the use of $\widehat{F}_{\text{eff}}$ to gauge instrument strength for the 2SLS estimator. Although this weak-instrument test is related to the bias, the results presented in Andrews et al. (2019, Section 3) for a sample of 106 specifications from papers published in the *American Economic Review* suggest that the effective F-statistic "may convey useful information about the instrument strength more broadly, since we see that conventional asymptotic approximations appear reasonable in specifications where the effective F-statistic exceeds 10.", Andrews et al. (2019, p 739).

In the next section, we introduce a class of generalized effective F-statistics, denoted $\widehat{F}_{\text{geff}}$, associated with a class of linear Generalized Method of Moments (GMM) estimators. We show that the weak-instrument Nagar bias results of MOP, derived for the effective F-statistic in relation to the 2SLS estimator, applies to this general class. The robust F-statistic is a member of this class and the associated GMM estimator, denoted GMMf, is defined as

$$\widehat{\beta}_{gmmf} = \frac{x'Z\widehat{W}_2^{-1}Z'y}{x'Z\widehat{W}_2^{-1}Z'x}, \tag{7}$$

where the extension "f" is for first stage, as the weight matrix is based on the first-stage residuals. As we will show and explain below, for the Andrews (2018) design with large values for the robust F-statistic, but small values for the effective F-statistic, this estimator is much better behaved in terms of bias and also inference than the 2SLS estimator.

Like the expression of the 2SLS estimator in (4), we can write the GMMf estimator as

$$\widehat{\beta}_{gmmf} = \frac{\widehat{\pi}'\left(V\widehat{a}r_r\left(\widehat{\pi}\right)\right)^{-1}\widehat{\pi}_y}{\widehat{\pi}'\left(V\widehat{a}r_r\left(\widehat{\pi}\right)\right)^{-1}\widehat{\pi}} = \frac{\widehat{\pi}'\left(V\widehat{a}r_r\left(\widehat{\pi}\right)\right)^{-1}\widehat{\pi}_y}{k_z\widehat{F}_r}, \tag{8}$$

where $V\widehat{a}r_r\left(\widehat{\pi}\right) = n\left(Z'Z\right)^{-1}\widehat{W}_2\left(Z'Z\right)^{-1}$ is the robust estimator of the variance of $\widehat{\pi}$.

The discussion in Andrews et al. (2019, pp 738-739) provides the intuition of why $\widehat{F}_{\text{eff}}$ is an appropriate statistic for testing instrument strength when using 2SLS. As they argue, 2SLS behaves badly when its denominator $x'P_Zx$ is close to zero. $\widehat{F}_{\text{eff}}$ measures this object, as $x'P_Zx$ is its numerator, and gets the standard errors right on average in the case of general heteroskedasticity, unlike the non-robust $\widehat{F}$. The robust F-statistic $\widehat{F}_r$ measures a different object, $x'Z\widehat{W}_2^{-1}Z'x$, and, asymptotically, "while it has a noncentral chi-square distribution, its noncentrality parameter does not correspond to the distribution of $\widehat{\beta}_{2sls}$", Andrews et al. (2019, p 739). But $x'Z\widehat{W}_2^{-1}Z'x$ is the denominator of the GMMf estimator, and so $\widehat{F}_r$ is the appropriate statistic for testing instrument strength when using $\widehat{\beta}_{gmmf}$.

3 The Generalized Effective F-Statistic as a Test for Weak Instruments

Consider the class of linear Generalized Methods of Moments (GMM) estimators of β, given by

$$\widehat{\beta}_{\Omega_n} = \frac{x' Z \Omega_n Z' y}{x' Z \Omega_n Z' x}, \tag{9}$$

where Ω_n is a $k_z \times k_z$ possibly data dependent weight matrix. It satisfies the following assumption,

Assumption 2. *Under the conditions stated in Assumption 1, as $n \to \infty$, $\Omega_n \xrightarrow{p} \Omega$, with Ω a finite, full rank matrix.*

Assumption 2 precludes the standard two-step GMM estimator with $\Omega_n = \Omega_n\left(\widehat{\beta}_1\right)$, e.g. in the cross-sectional setting, $\Omega_n\left(\widehat{\beta}_1\right) = \left(\frac{1}{n}\sum_{i=1}^n \widehat{u}_{1i}^2 z_i z_i'\right)^{-1}$, where $\widehat{u}_{1i} = y_i - x_i \widehat{\beta}_1$ and $\widehat{\beta}_1$ is an initial estimator, for example the 2SLS estimator. The initial estimator is consistent under standard strong-instruments asymptotics, but it converges to a non-degenerate random variable under weak-instruments asymptotics, see (12) below, and hence Assumption 2 does not hold.

Let

$$W_\Omega = \begin{bmatrix} W_{\Omega,1} & W_{\Omega,12} \\ W_{\Omega,12}' & W_{\Omega,2} \end{bmatrix} = \left(I_2 \otimes \Omega^{1/2}\right) W \left(I_2 \otimes \Omega^{1/2}\right),$$

and

$$\widehat{W}_{\Omega_n} = \left(I_2 \otimes \Omega_n^{1/2}\right) \widehat{W} \left(I_2 \otimes \Omega_n^{1/2}\right). \tag{10}$$

Then consider the class of generalized effective F-statistics, given by

$$\widehat{F}_{\text{geff}}\left(\Omega_n\right) = \frac{x' Z \Omega_n Z' x}{n \operatorname{tr}\left(\widehat{W}_{\Omega_n,2}\right)} = \frac{x' Z \Omega_n Z' x}{n \operatorname{tr}\left(\Omega_n^{1/2} \widehat{W}_2 \Omega_n^{1/2}\right)} = \frac{x' Z \Omega_n Z' x}{n \operatorname{tr}\left(\widehat{W}_2 \Omega_n\right)}. \tag{11}$$

For the 2SLS estimator we have $\Omega_n = \left(\frac{1}{n} Z' Z\right)^{-1}$ and

$$\widehat{F}_{\text{geff}}\left(\left(Z'Z/n\right)^{-1}\right) = \frac{x' P_Z x}{\operatorname{tr}\left(\widehat{W}_2 \left(Z'Z/n\right)^{-1}\right)} = \widehat{F}_{\text{eff}}.$$

For the GMMf estimator as defined in (7), we have $\Omega_n = \widehat{W}_2^{-1}$. Therefore

$$\widehat{F}_{\text{geff}}\left(\widehat{W}_2^{-1}\right) = \frac{x' Z \widehat{W}_2^{-1} Z' x}{n k_z} = \widehat{F}_r.$$

This remainder of this section together with the proofs in the Appendix draw heavily on Montiel Olea and Pflueger (2013). We show that the MOP weak-instruments testing methodology for 2SLS applies to the class of generalized F-statistics in relation to the Nagar bias of the linear GMM estimators. We do this by restating parts of their Lemma 1 and Theorem 1, Montiel Olea and Pflueger (2013, p 362), that directly apply to our GMM setting.

Lemma 1. *Under Assumptions 1 and 2, the following limits hold jointly as $n \to \infty$.*

$$\widehat{\beta}_{\Omega_n} - \beta \xrightarrow{d} \beta^*_\Omega = \left(\gamma'_{\Omega,2}\gamma_{\Omega,2}\right)^{-1}\gamma'_{\Omega,2}\left(\gamma_{\Omega,1} - \beta\gamma_{\Omega,2}\right) \tag{12}$$

$$\widehat{F}_{geff}\left(\Omega_n\right) \xrightarrow{d} F^*_{geff}\left(\Omega\right) = \gamma'_{\Omega,2}\gamma_{\Omega,2}/tr\left(W_{\Omega,2}\right) = \gamma'_{\Omega,2}\gamma_{\Omega,2}/tr\left(W_2\Omega\right), \tag{13}$$

where

$$\begin{pmatrix} \gamma_{\Omega,1} \\ \gamma_{\Omega,2} \end{pmatrix} \sim N\left(\begin{pmatrix} c_\Omega\beta \\ c_\Omega \end{pmatrix}, W_\Omega\right),$$

with $c_\Omega = \Omega^{1/2}Q_{zz}c$.

Proof. See Appendix A.3. □

3.1 Nagar Bias Approximation

As $v_1 = u + \beta v_2$, it follows from Assumption 1 that, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \Omega_n^{1/2}Z'u \\ \Omega_n^{1/2}Z'v_2 \end{pmatrix} \xrightarrow{d} N\left(0, S\left(\beta, W_\Omega\right)\right),$$

with

$$S\left(\beta, W_\Omega\right) = \begin{bmatrix} S_1\left(\beta, W_\Omega\right) & S_{12}\left(\beta, W_\Omega\right) \\ S_{12}\left(\beta, W_\Omega\right)' & W_{\Omega,2} \end{bmatrix}, \tag{14}$$

where

$$S_1\left(\beta, W_\Omega\right) = W_{\Omega,1} - \beta\left(W_{\Omega,12} + W'_{\Omega,12}\right) + \beta^2 W_{\Omega,2}$$
$$S_{12}\left(\beta, W_\Omega\right) = W_{\Omega,12} - \beta W_{\Omega,2}.$$

We can now state Theorem 1 of Montiel Olea and Pflueger (2013, p 262) on the Nagar bias approximation for our general class of GMM estimators.

Theorem 1. *Nagar Bias Approximation. Let* $c_{\Omega,0} := c_\Omega / \|c_\Omega\|$*, where* $\|c_\Omega\| = \sqrt{c'_\Omega c_\Omega}$*, and let the concentration parameter* $\mu^2_\Omega := \|c_\Omega\|^2 / tr(W_{\Omega,2})$*. Define the benchmark bias as*

$$BM(\beta, W_\Omega) := \sqrt{\frac{tr(S_1(\beta, W_\Omega))}{tr(W_{\Omega,2})}}. \tag{15}$$

The Taylor series expansion of β^*_Ω *around* $\mu^{-1}_\Omega = 0$ *results in the Nagar (1959) bias approximation*

$$\mathbb{E}[\beta^*_\Omega] \approx N(\beta, c_\Omega, W_\Omega) = \frac{n(\beta, c_{\Omega,0}, W_\Omega)}{\mu^2_\Omega}, \tag{16}$$

with

$$n(\beta, c_{\Omega,0}, W_\Omega) = \frac{tr(S_{12}(\beta, W_\Omega)) - 2c'_{\Omega,0} S_{12}(\beta, W_\Omega)\, c_{\Omega,0}}{tr(W_{\Omega,2})}. \tag{17}$$

Further,

$$B(W_\Omega) := \sup_{\beta \in \mathbb{R}, c_{\Omega,0} \in \mathcal{S}^{k_z - 1}} \left(\frac{|n(\beta, c_{\Omega,0}, W_\Omega)|}{BM(\beta, W_\Omega)} \right) \leq 1 \tag{18}$$

where $\mathcal{S}^{k_z-1}$ *is the* k_z*-dimensional unit sphere.*

Proof. For the Nagar bias approximation, see Appendix A.4. For $B(W_\Omega) \leq 1$, see Montiel Olea and Pflueger (2013, Apendix A.3). □

MOP interpret the benchmark $\mathrm{BM}(\beta, W_\Omega)$ as a "worst-case" bias. It is derived by approximating the expectation of the ratio by the ratio of expectations,

$$\begin{aligned} \mathbb{E}[\beta^*_\Omega] &\approx \frac{\mathbb{E}\left[\gamma'_{\Omega,2}(\gamma_{\Omega,1} - \beta\gamma_{\Omega,2})\right]}{\mathbb{E}\left[\gamma'_{\Omega,2}\gamma_{\Omega,2}\right]} = \frac{\mathrm{tr}(S_{12}(\beta, W_\Omega))}{\mathrm{tr}(W_{\Omega,2})(1 + \mu^2_\Omega)} \\ &= \frac{1}{(1 + \mu^2_\Omega)} \frac{\mathrm{tr}(S_{12}(\beta, W_\Omega))}{\sqrt{\mathrm{tr}(W_{\Omega,2})}\sqrt{\mathrm{tr}(S_1(\beta, W_\Omega))}} \sqrt{\frac{\mathrm{tr}(S_1(\beta, W_\Omega))}{\mathrm{tr}(W_{\Omega,2})}}. \end{aligned}$$

This expected (absolute) bias expression is maximized when the concentration parameter $\mu^2_\Omega = 0$ and when the the first-stage and structural errors are perfectly correlated, and it follows that then $\mathbb{E}[\beta^*_\Omega] \leq \sqrt{\mathrm{tr}(S_1(\beta, W_\Omega)) / \mathrm{tr}(W_{\Omega,2})}$, see Montiel Olea and Pflueger (2013, pp 362-363).

The results of Lemma 1 and Theorem 1 are those in MOP for the 2SLS estimator, with $\Omega_n = \left(\frac{1}{n} Z'Z\right)^{-1}$ and $\Omega = Q^{-1}_{zz}$. The Lemma and Theorem are replicated here to show that their methodology applies directly to the larger class of GMM estimators (9), under Assumption 2. We can therefore also apply the MOP procedure for testing for weak instruments directly, as we describe next.

3.2 Null Hypothesis of Weak Instruments and Testing Procedure

The null hypothesis of weak instruments is specified as in MOP as

$$H_0 : \mu_\Omega^2 \in \mathcal{H}\left(W_\Omega, \tau\right) \text{ against } H_1 : \mu_\Omega^2 \notin \mathcal{H}\left(W_\Omega, \tau\right),$$

where

$$\mathcal{H}\left(W_\Omega, \tau\right) = \left\{ \mu_\Omega^2 \in \mathbb{R}_+ : \sup_{\beta \in \mathbb{R}, c_{\Omega,0} \in \mathcal{S}^{k_z - 1}} \left(\frac{\left| N\left(\beta, \mu_\Omega \sqrt{\operatorname{tr}\left(W_{\Omega,2}\right)} c_{\Omega,0}, W_\Omega\right) \right|}{\operatorname{BM}\left(\beta, W_\Omega\right)} \right) > \tau \right\},$$

or equivalently

$$\mathcal{H}\left(W_\Omega, \tau\right) = \left\{ \mu_\Omega^2 \in \mathbb{R}_+ : \mu_\Omega^2 < \frac{B\left(W_\Omega\right)}{\tau} \right\}.$$

Under the null hypothesis of weak instruments, the Nagar bias exceeds a fraction τ of the benchmark for at least some value of the structural parameter β and some direction of the first-stage parameters, $c_{\Omega,0}$. The parameter τ is a user specified threshold, commonly set to $\tau = 0.10$.

The generalization of the MOP test for weak instruments is then based on $\widehat{F}_{\text{geff}}\left(\Omega_n\right)$ which is asymptotically distributed as $\gamma_{\Omega,2}' \gamma_{\Omega,2} / \operatorname{tr}\left(W_{\Omega,2}\right)$, with $\gamma_{\Omega,2} \sim N\left(c_\Omega, W_{\Omega,2}\right)$, which has mean $1 + \mu_\Omega^2$. It follows that we reject H_0 when $\widehat{F}_{\text{geff}}\left(\Omega_n\right)$ is large. Denote by $F_{c_\Omega, W_{\Omega,2}}^{-1}\left(\alpha\right)$ the upper α quantile of the distribution of $\gamma_{\Omega,2}' \gamma_{\Omega,2} / \operatorname{tr}\left(W_{\Omega,2}\right)$ and let

$$cv\left(\alpha, W_{\Omega,2}, d_\Omega\right) := \sup_{c_\Omega \in \mathbb{R}^{k_z}} \left\{ F_{c_\Omega, W_{\Omega,2}}^{-1}\left(\alpha\right) 1_{\left(\frac{c_\Omega' c_\Omega}{\operatorname{tr}\left(W_{\Omega,2}\right)} < d_\Omega \right)} \right\},$$

where $1_{(A)}$ denotes the indicator function over a set A. The null of weak instruments is then rejected if

$$\widehat{F}_{\text{geff}}\left(\Omega_n\right) > cv\left(\alpha, \widehat{W}_{\Omega_n,2}, B\left(\widehat{W}_{\Omega_n}\right) / \tau\right),$$

which is shown in Lemma 2 of Montiel Olea and Pflueger (2013, p 363) to be pointwise asymptotically valid,

$$\sup_{\mathcal{H}\left(W_\Omega, \tau\right)} \lim_{n \to \infty} \mathbb{P}\left(\widehat{F}_{\text{geff}}\left(\Omega_n\right) > cv\left(\alpha, \widehat{W}_{\Omega_n,2}, B\left(\widehat{W}_{\Omega_n}\right) / \tau\right) \right) \leq \alpha,$$

and, provided that $B\left(\widehat{W}_{\Omega_n}\right)$ is bounded in probability,

$$\lim_{\mu_\Omega^2\to\infty}\lim_{n\to\infty}\mathbb{P}\left(\widehat{F}_{\text{geff}}\left(\Omega_n\right)>cv\left(\alpha,\widehat{W}_{\Omega_n,2},B\left(\widehat{W}_{\Omega_n}\right)/\tau\right)\right)=1.$$

After obtaining $B\left(\widehat{W}_{\Omega_n}\right)$ by a numerical routine, MOP show that the critical values can be obtained by Monte Carlo methods or by the Patnaik (1949) curve-fitting methodology. The Patnaik critical value is obtained as the the upper α quantile of $\chi^2_{\widehat{k}_{\text{geff}}(\Omega_n)}\left(d_{\Omega_n,\tau}\widehat{k}_{\text{geff}}\left(\Omega_n\right)\right)/\widehat{k}_{\text{geff}}\left(\Omega_n\right)$ where $\chi^2_{\widehat{k}_{\text{geff}}(\Omega_n)}\left(d_{\Omega_n,\tau}\widehat{k}_{\text{geff}}\left(\Omega_n\right)\right)$ denotes the noncentral χ^2 distribution with $\widehat{k}_{\text{geff}}(\Omega_n)$ degrees of freedom and noncentrality parameter $d_{\Omega_n,\tau}\widehat{k}_{\text{geff}}\left(\Omega_n\right)$, with

$$d_{\Omega_n\tau}=B\left(\widehat{W}_{\Omega_n}\right)/\tau; \tag{19}$$

$$\widehat{k}_{\text{geff}}\left(\Omega_n\right)=\frac{\left[\text{tr}\left(\widehat{W}_{\Omega_n,2}\right)\right]^2\left(1+2d_{\Omega_n,\tau}\right)}{\text{tr}\left(\widehat{W}'_{\Omega_n,2}\widehat{W}_{\Omega_n,2}\right)+2d_{\Omega_n,\tau}\text{tr}\left(\widehat{W}_{\Omega_n,2}\right)\lambda_{\max}\left(\widehat{W}_{\Omega_n,2}\right)}, \tag{20}$$

and where $\lambda_{\max}\left(\widehat{W}_{\Omega_n,2}\right)$ denotes the maximum eigenvalue of $\widehat{W}_{\Omega_n,2}$.

To summarize, and following MOP, the weak-instruments test procedure related to the Nagar approximation of the bias of the GMM estimator $\widehat{\beta}_{\Omega_n}$as defined in (9), under Assumptions 1 and 2 is as follows.

1. Compute the generalized effective F-statistic,

$$\widehat{F}_{\text{geff}}\left(\Omega_n\right)=\frac{x'Z\Omega_nZ'x}{n\text{tr}\left(\widehat{W}_2\Omega_n\right)}.$$

2. Obtain

$$B\left(\widehat{W}_{\Omega_n}\right)=\sup_{\beta\in\mathbb{R},c_{\Omega,0}\in\mathcal{S}^{k_z-1}}\left(\frac{\left|n\left(\beta,c_{\Omega,0},\widehat{W}_{\Omega_n}\right)\right|}{\text{BM}\left(\beta,\widehat{W}_{\Omega_n}\right)}\right)$$

by a numerical maximization routine, where $\widehat{W}_{\Omega_n}$ is as defined in (10), $n\left(\beta,c_{\Omega,0},\widehat{W}_{\Omega_n}\right)$ and $\text{BM}\left(\beta,\widehat{W}_{\Omega_n}\right)$ in (17) and (18) respectively, with the estimated $\widehat{W}_{\Omega_n}$ replacing W_Ω.

3. Applying the Patnaik (1949) curve-fitting methodology, estimate the effective degrees of freedom $\widehat{k}_{\text{geff}}\left(\Omega_n\right)$ as given in (20) and compute the critical value $cv\left(\alpha,\widehat{W}_{\Omega_n,2},d_{\Omega_n,\tau}\right)$ for a user specified threshold value τ as the upper α quantile of

$\chi^2_{\widehat{k}_{\text{geff}}(\Omega_n)}\left(d_{\Omega_n,\tau}\widehat{k}_{\text{geff}}(\Omega_n)\right)/\widehat{k}_{\text{geff}}(\Omega_n)$, where $d_{\Omega_n,\tau}$ is defined in (19).

4. Reject the null of weak instruments, that the proportion of the Nagar approximation of the bias of $\widehat{\beta}_{\Omega_n}$ relative to the benchmark bias is larger than τ, if $\widehat{F}_{\text{geff}}(\Omega_n) > cv\left(\alpha, \widehat{W}_{\Omega_n,2}, d_{\Omega_n,\tau}\right)$.

As an alternative to step 3. one can use Monte Carlo methods, see Montiel Olea and Pflueger (2013, Section 5). As $B(W_\Omega) \leq 1$, MOP propose a simplified asymptotically valid but conservative test. For this simplified test procedure, step 2. is not needed, instead replacing $B\left(\widehat{W}_{\Omega_n}\right)$ by 1, and so $d_{\Omega_n,\tau}$ by $1/\tau$, in steps 3. and 4. Hence the simplified test rejects the null hypothesis of weak instruments if

$$\widehat{F}_{\text{geff}} > cv\left(\alpha, \widehat{W}_{\Omega_n,2}, 1/\tau\right),$$

with $cv\left(\alpha, \widehat{W}_{\Omega_n,2}, 1/\tau\right) \geq cv\left(\alpha, \widehat{W}_{\Omega,2}, d_{\Omega_n,\tau}\right)$.

4 The Robust F-Statistic as a Test for Weak Instruments

For the robust F-statistic in relation to the Nagar bias of the GMMf estimator as defined in (7),

$$\widehat{\beta}_{gmmf} = \frac{x'Z\widehat{W}_2^{-1}Z'y}{x'Z\widehat{W}_2^{-1}Z'x},$$

the above expressions apply, but simplify significantly. With $\Omega_n = \widehat{W}_2^{-1} \xrightarrow{p} W_2^{-1} = \Omega$, it follows that

$$W_\Omega = \begin{bmatrix} W_2^{-1/2}W_1W_2^{-1/2} & W_2^{-1/2}W_{12}W_2^{-1/2} \\ W_2^{-1/2}W_{12}'W_2^{-1/2} & I_{k_z} \end{bmatrix}.$$

For the Nagar bias expression, we get

$$N_{gmmf}(\beta, c_\Omega, W_\Omega) = \frac{1}{\mu_\Omega^2} n_{gmmf}(\beta, c_\Omega, W_\Omega),$$

with $\mu_\Omega^2 = c'Q_{zz}W_2^{-1}Q_{zz}c/k_z$ and

$$n_{gmmf}(\beta, c_{\Omega,0}, W_\Omega) = \frac{\text{tr}(S_{12}(\beta, W_\Omega)) - 2c_{\Omega,0}'S_{12}(\beta, W_\Omega)c_{\Omega,0}}{k_z},$$

where

$$S_{12}(\beta, W_\Omega) = W_{\Omega,12} - \beta I_{k_z}.$$

The expression therefore simplifies to

$$n_{gmmf}(\beta, c_{\Omega,0}, W_\Omega) = \frac{1}{k_z}\left(\text{tr}(W_{\Omega,12}) - 2c'_{\Omega,0}W_{\Omega,12}c_{\Omega,0} - (k_z - 2)\beta\right).$$

The benchmark worst-case bias for the GMMf estimator is then given by

$$\text{BM}(\beta, W_\Omega) = \sqrt{\frac{\text{tr}(S_1(\beta, W_\Omega))}{k_z}}$$

where

$$\text{tr}(S_1(\beta, W_\Omega)) = \text{tr}(W_{\Omega,1}) - 2\beta\text{tr}(W_{\Omega,12}) + k_z\beta^2.$$

Then

$$\begin{aligned} B_{gmmf}(W_\Omega) &= \sup_{\beta\in\mathbb{R}, c_{\Omega,0}\in\mathcal{S}^{k_z-1}}\left(\frac{|n_{gmmf}(\beta, c_{\Omega,0}, W_\Omega)|}{\text{BM}(\beta, W_\Omega)}\right) \\ &= \sup_{\beta\in\mathbb{R}, c_{\Omega,0}\in\mathcal{S}^{k_z-1}}\left(\frac{\left|\text{tr}(W_{\Omega,12}) - 2c'_{\Omega,0}W_{\Omega,12}c_{\Omega,0} - (k_z-2)\beta\right|}{\sqrt{k_z\left(\text{tr}(W_{\Omega,1}) - 2\beta\text{tr}(W_{\Omega,12}) + k_z\beta^2\right)}}\right) \end{aligned}$$

As $\Omega = W_2^{-1}$ we have that $\gamma_{\Omega,2} \sim N(c_\Omega, I_{k_z})$ and so it follows that $\gamma'_{\Omega,2}\gamma_{\Omega,2} \sim \chi^2_{k_z}(c'_\Omega c_\Omega) = \chi^2_{k_z}(k_z\mu^2_\Omega)$. Therefore, the null of weak instruments for the GMMf estimator, specified as

$$H_0 : \mu^2_\Omega \in \mathcal{H}_{gmmf}(W_\Omega, \tau),$$

$$\mathcal{H}_{gmmf}(W_\Omega, \tau) = \left\{\mu^2_\Omega \in \mathbb{R}_+ : \mu^2_\Omega < \frac{B_{gmmf}(W_\Omega)}{\tau}\right\},$$

is rejected if

$$\widehat{F}_r > cv(\alpha, k_z, d_{\Omega_n,\tau}),$$

with $cv(\alpha, k_z, d_{\Omega_n,\tau})$ the upper α quantile of $\chi^2_{k_z}(k_z d_{\Omega_n})/k_z$, and where $d_{\Omega_n,\tau} = B_{gmmf}\left(\widehat{W}_{\Omega_n}\right)/\tau$ and $\widehat{W}_{\Omega_n} = \left(I_2 \otimes \widehat{W}_2^{-1/2}\right)\widehat{W}\left(I_2 \otimes \widehat{W}_2^{-1/2}\right)$. Relative to the general results for the generalized effective F-statistic, we see that for the GMMf estimator, $B_{gmmf}\left(\widehat{W}_{\Omega_n}\right)$ is a simpler function to maximize with respect to β and $c_{\Omega,0}$ than $B\left(\widehat{W}_{\Omega_n}\right)$ from (18) when $\Omega_n \neq \widehat{W}_2^{-1}$. There is further no need for Monte Carlo simulations or Patnaik's curve-fitting methodology to compute the critical values, as $\widehat{F}_r$ follows an asymptotic scaled noncentral chi-square distribution, with the "effective" degrees of freedom here equal to

k_z.

To summarize the weak-instruments testing procedure in relation to the approximate Nagar bias of $\widehat{\beta}_{gmmf}$,

1. Compute the robust F-statistic

$$\widehat{F}_r = \frac{x'Z\widehat{W}_2^{-1}Z'x}{nk_z}.$$

2. Obtain

$$B_{gmmf}\left(\widehat{W}_{\Omega_n}\right) = \sup_{\beta\in\mathbb{R},c_{\Omega,0}\in\mathcal{S}^{k_z-1}}\left(\frac{\left|\text{tr}\left(\widehat{W}_{\Omega_n,12}\right) - 2c_{\Omega,0}'\widehat{W}_{\Omega_n,12}c_{\Omega,0} - (k_z-2)\,\beta\right|}{\sqrt{k_z\left(\text{tr}\left(\widehat{W}_{\Omega_n,1}\right) - 2\beta\text{tr}\left(\widehat{W}_{\Omega_n,12}\right) + k_z\beta^2\right)}}\right)$$

by a numerical maximization routine, where $\widehat{W}_{\Omega_n} = \left(I_2 \otimes \widehat{W}_2^{-1/2}\right)\widehat{W}\left(I_2 \otimes \widehat{W}_2^{-1/2}\right)$.

3. Reject the null of weak instruments if $\widehat{F}_r > cv\left(\alpha, \widehat{W}_{\Omega_n,2}, d_{\Omega_n,\tau}\right)$, where $d_{\Omega_n,\tau} = B_{gmmf}\left(\widehat{W}_{\Omega_n}\right)/\tau$ and where $cv\left(\alpha, \widehat{W}_{\Omega_n,2}, d_{\Omega_n,\tau}\right)$ is the upper α quantile of $\chi^2_{k_z}\left(k_z d_{\Omega_n,\tau}\right)/k_z$.

The simplified conservative test based on the fact that $B_{gmmf}\left(W_\Omega\right) \leq 1$ is obtained using the critical value $cv\left(\alpha, k_z, 1/\tau\right)$, which is simply the upper α quantile of $\chi^2_{k_z}\left(k_z/\tau\right)/k_z$.

We can illustrate divergence of $\widehat{F}_{\text{eff}}$ and $\widehat{F}_r$, building on examples in Montiel Olea and Pflueger (2013) and Andrews et al. (2019). Let $k_z = 2$, $Q_{zz} = I_2$ and $W_2 = \begin{pmatrix} \xi_1^2 & 0 \\ 0 & \xi_2^2 \end{pmatrix}$. Then for 2SLS/$\widehat{F}_{\text{eff}}$ we have the concentration parameter $\mu^2_{Q_{zz}^{-1}} = \frac{c_1^2+c_2^2}{\xi_1^2+\xi_2^2}$, and for GMMf/$\widehat{F}_r$ we have $\mu^2_{W_2^{-1}} = \frac{1}{2}\left(\frac{c_1^2}{\xi_1^2} + \frac{c_2^2}{\xi_2^2}\right)$. Then, for $c_1^2 > 0$ and $c_2^2 > 0$, if for example $\xi_1^2 \to 0$ and $\xi_2^2 > c_1^2 + c_2^2$ then $\mu^2_{Q_{zz}^{-1}} \to \frac{c_1^2+c_2^2}{\xi_2^2} < 1$ and $\mu^2_{W_2^{-1}} \to \infty$. In this situation, there is no weak-instruments problem for the GMMf estimator, but there is for the 2SLS estimator. A design where such heteroskedasticity can be generated is the grouped-data IV one, and is the one used in Andrews (2018). We consider this design in Section 6, where we further show how the GMMf estimator utilizes the information in such designs better than the 2SLS estimator.[2]

[2] Note that for this example it is the case that if $\left(\xi_1^2 > \xi_2^2\right)$ *and* $\left(\frac{c_1^2}{\xi_1^2} > \frac{c_2^2}{\xi_2^2}\right)$ then $\mu^2_{Q_{zz}^{-1}} > \mu^2_{W_2^{-1}}$.

4.1 Efficiency

Under standard strong-instruments asymptotics, the 2SLS estimator is consistent and asymptotically efficient if $\mathbb{E}\left[u_i^2 z_i z_i'\right] = \sigma_u^2 Q_{zz}$, and for the GMMf estimator this is the case if $\mathbb{E}\left[v_{2,i}^2 z_i z_i'\right] = \delta \mathbb{E}\left[u_i^2 z_i z_i'\right]$, for some constant $\delta > 0$. However, under the weak-instruments asymptotics setting of Assumption 1, it follows from the expression of the random variable β_Ω^* in (12) that the GMM estimators are biased and inconsistent. Therefore, the weak-instruments test is based on the relative Nagar bias, and GMM estimators can then be ranked on their bias performance, not on their relative efficiency. As the MOP benchmark bias is estimator specific, we first propose to harmonize the bias by using the worst-case OLS bias as a benchmark.

5 Harmonizing the Benchmark Bias

The benchmark bias $\mathrm{BM}\left(\beta, W_\Omega\right)$ as defined in (15) is estimator/F-statistic specific and is derived using an ad hoc approximation of $\mathbb{E}\left(\beta_\Omega^*\right)$. This makes a comparison of weak-instrument test results between generalized F-statistics for different GMM estimators difficult, as the critical values of the test are based on the maximum of the Nagar bias approximation relative to the estimator-specific benchmark. A solution is to harmonize the benchmark bias, such that it is the same for each estimator/F-statistic. For the homoskedastic case, the bias related weak-instruments critical values of Stock and Yogo (2005) for the nonrobust F-statistic are in terms of the relative bias, that of the 2SLS estimator relative to that of the OLS estimator. We propose here to harmonize the benchmark worst-case bias and to replace it with the worst-case bias of the OLS estimator.

The probability limit of the worst-case weak-instrument OLS (absolute) bias, under Assumption 1 is given by

$$
\begin{aligned}
\widehat{\beta}_{LS} - \beta &= \frac{x'u}{x'x} \\
&= \frac{c'Z'u/\sqrt{n} + v_2'u}{c'Z'Zc/n + 2c'Z'v_2/\sqrt{n} + v_2'v_2} \\
&\xrightarrow{p} \frac{\sigma_{uv_2}\left(\beta, \Sigma_v\right)}{\sigma_2^2} = \rho_{uv_2} \frac{\sigma_u\left(\beta, \Sigma_v\right)}{\sigma_2} \\
&\leq \frac{\sigma_u\left(\beta, \Sigma_v\right)}{\sigma_2} = \sqrt{\frac{\sigma_1^2 - 2\beta\sigma_{12} + \beta^2\sigma_2^2}{\sigma_2^2}} = \mathrm{BM}_{LS}\left(\beta, \Sigma_v\right).
\end{aligned}
$$

Note that this benchmark bias is only a function of β and Σ_v, the marginal variance of $(v_1\, v_2)'$, applies to all forms of general heteroskedasticity and does not depend on homoskedasticity assumptions. However, it can be seen to be a valid benchmark for all GMM estimators defined in (9) by considering the worst-case benchmark bias $\mathrm{BM}\left(\beta, W_\Omega\right)$ under the maintained assumption of homoskedasticity. In that case we have $W = \Sigma_v \otimes Q_{zz}$ and thus $W_\Omega = \Sigma_v \otimes \Omega^{-1/2} Q_{zz} \Omega^{-1/2}$. Then it follows that

$$\begin{aligned}\mathrm{BM}\left(\beta, W_\Omega\right) &= \sqrt{\frac{\left(\sigma_1^2 - 2\beta\sigma_{12} + \beta^2\sigma_2^2\right)\mathrm{tr}\left(Q_{zz}\Omega^{-1}\right)}{\sigma_2^2 \mathrm{tr}\left(Q_{zz}\Omega^{-1}\right)}} \\ &= \sqrt{\frac{\sigma_1^2 - 2\beta\sigma_{12} + \beta^2\sigma_2^2}{\sigma_2^2}} = \mathrm{BM}_{LS}\left(\beta, \Sigma_v\right).\end{aligned}$$

We have that

$$\lim_{\beta \to \pm\infty} \frac{\left|n\left(\beta, c_{\Omega,0}, W_\Omega\right)\right|}{\mathrm{BM}\left(\beta, W_\Omega\right)} = \lim_{\beta \to \pm\infty} \frac{\left|n\left(\beta, c_{\Omega,0}, W_\Omega\right)\right|}{\mathrm{BM}_{LS}\left(\beta, \Sigma_v\right)} = 1 - \frac{2\lambda_{\min}\left(W_{\Omega,2}\right)}{\mathrm{tr}\left(W_{\Omega,2}\right)},$$

where $\lambda_{\min}\left(W_{\Omega,2}\right)$ is the minimum eigenvalue of $W_{\Omega,2}$. It follows further from Assumptions 1 and 2 that

$$0 < \frac{\mathrm{BM}\left(\beta, W_\Omega\right)}{\mathrm{BM}_{LS}\left(\beta, \Sigma_v\right)} < C$$

$\forall \beta \in \mathbb{R}$, for some finite $C > 1$. Note that the positive definiteness, or full rank assumption of Σ_v is important, as otherwise $\mathrm{BM}_{LS}\left(\beta, \Sigma_v\right)$ would be zero for some value of β. Therefore the case of $\rho_{12}^2 = 1$ is excluded. It then follows from Theorem 1 that,

$$B_{LS}\left(W_\Omega, \Sigma_v\right) \coloneqq \sup_{\beta \in \mathbb{R}, c_{\Omega,0} \in \mathcal{S}^{k_z - 1}} \left(\frac{\left|n\left(\beta, c_{\Omega,0}, W_\Omega\right)\right|}{\mathrm{BM}_{LS}\left(\beta, \Sigma_v\right)}\right) < \infty$$

and the test procedures described in Section 3 apply, replacing $B\left(W_\Omega\right)$ by $B_{LS}\left(W_\Omega, \Sigma_v\right)$ from step 2 onwards in the summaries of the testing procedures. The null of weak instruments is then rejected if

$$\widehat{F}_{\text{geff}}\left(\Omega_n\right) > cv^{LS}\left(\alpha, \widehat{W}_{\Omega_n,2}, d_{\Omega_n,\tau}^{LS}\right),$$

where $d_{\Omega_n,\tau}^{LS} = B_{LS}\left(\widehat{W}_{\Omega_n}, \widehat{\Sigma}_v\right)/\tau$.

The interpretation of the weak-instruments null hypothesis $H_0 : \mu_\Omega^2 \in \mathcal{H}\left(W_\Omega, \Sigma_v, \tau\right)$, where

$$\mathcal{H}\left(W_\Omega, \Sigma_v, \tau\right) = \left\{\mu_\Omega^2 \in \mathbb{R}_+ : \sup_{\beta \in \mathbb{R}, c_{\Omega,0} \in \mathcal{S}^{k_z - 1}} \left(\frac{\left|N\left(\beta, \mu_\Omega \sqrt{\mathrm{tr}\left(W_{\Omega,2}\right)} c_{\Omega,0}, W_\Omega\right)\right|}{\mathrm{BM}_{LS}\left(\beta, \Sigma_v\right)}\right) > \tau\right\},$$

is then that the Nagar bias exceeds a fraction τ of the benchmark maximum OLS bias for at least some value of the structural parameter β and some direction of the first-stage coefficients. Whilst this is not the same as the asymptotic relative bias results of Stock and Yogo (2005), it is more aligned with it. It makes the interpretation of the null hypothesis and hence that of its rejection the same for different GMM estimators and their associated generalized effective F-statistics and is not based on an ad hoc approximation. In the remainder, this benchmark is used. Note that the simplified conservative test procedure does not apply here, as it is not the case that $B_{LS}\left(W_{\Omega},\Sigma_v\right)\leq 1$ in general.

6 Grouped-Data IV Model

We now consider the heteroskedastic model designs from Andrews (2018). In these designs, very large values of the robust F-statistic are accompanied by a poor performance of the 2SLS estimator, where Andrews (2018) focused on coverage distortions of confidence sets. We find that in these designs the effective F-statistic is indeed small, indicating a weak-instrument problem for 2SLS, but large values of the robust F-statistic indicate there is not a weak-instrument problem for the GMMf estimator.

Following the example of divergence between $\widehat{F}_{\text{eff}}$ and $\widehat{F}_r$ given at at the end of Section 4, the design in Andrews (2018, Supplementary Appendix C.3) is that of a grouped-data IV setup,

$$\begin{aligned} y_i &= x_i\beta + u_i \\ x_i &= z_i'\pi + v_{2,i}, \end{aligned}$$

for $i=1,...,n$, where the G-vector $z_i\in\{e_1,...,e_G\}$, with e_g a G-vector with gth entry equal to 1 and zeros everywhere else, for $g=1,\ldots,G$.

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

$$\left(\left(\begin{array}{c} u_i \\ v_{2,i} \end{array}\right)|z_i=e_g\right)\sim\left(0,\Sigma_g\right),$$

$$\Sigma_g=\left[\begin{array}{cc} \sigma^2_{u,g} & \sigma_{uv_2,g} \\ \sigma_{uv_2,g} & \sigma^2_{v_2,g} \end{array}\right]. \tag{21}$$

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

$$y_{jg} = x_{jg}\beta + u_{jg} \tag{22}$$

$$x_{jg} = \pi_g + v_{2,jg} \tag{23}$$

$$\begin{pmatrix} u_{jg} \\ v_{2,jg} \end{pmatrix} \sim (0, \Sigma_g),$$

for $j = 1, ..., n_g$ and $g = 1, ..., G$, with n_g the number of observations in group g, $\sum_{g=1}^{G} n_g = n$, see also Bekker and Ploeg (2005). We assume that $\lim_{n\to\infty} \frac{n_g}{n} = f_g$, with $0 < f_g < 1$.

The OLS estimator of π_g is given by $\widehat{\pi}_g = \overline{x}_g = \frac{1}{n_g}\sum_{j=1}^{n_s} x_{jg}$ and $Var(\widehat{\pi}_g) = \sigma^2_{v_2,g}/n_g$. The OLS residual is $\widehat{v}_{2,jg} = x_{jg} - \overline{x}_g$ and the estimator for the variance is given by $V\widehat{a}r(\widehat{\pi}_g) = \widehat{\sigma}^2_{v_2,g}/n_g$, where $\widehat{\sigma}^2_{v,g} = \frac{1}{n_g}\sum_{j=1}^{n_g} \widehat{v}^2_{2,jg}$. Let Z be the $n \times G$ matrix of instruments. For the vector π the OLS estimator is given by

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

Let

$$\begin{aligned} \widehat{W}_2 &= \frac{1}{n}\sum_{i=1}^{n} \widehat{v}^2_{2,i} z_i z_i' \\ &= \text{Diag}\left(n_g \widehat{\sigma}^2_{v_2,g}\right)/n, \end{aligned} \tag{24}$$

where $\text{Diag}(q_g)$ is a diagonal matrix with gth diagonal element q_g. Then the robust estimator of $Var(\widehat{\pi})$ is given by

$$\begin{aligned} V\widehat{a}r_r(\widehat{\pi}) &= (Z'Z)^{-1} n\widehat{W}_2 (Z'Z)^{-1} \\ &= \text{Diag}\left(\widehat{\sigma}^2_{v_2,g}/n_g\right). \end{aligned}$$

The non-robust variance estimator is

$$\begin{aligned} V\widehat{a}r(\widehat{\pi}) &= \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{v}^2_{2,i}\right)(Z'Z)^{-1} \\ &= \left(\sum_{g=1}^{G} \frac{n_g}{n}\widehat{\sigma}^2_{v_2,g}\right) \text{Diag}\left(\frac{1}{n_g}\right). \end{aligned}$$

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

$$\widehat{\beta}_g = \frac{z_g' y}{z_g' x} = \frac{\overline{y}_g}{\overline{x}_g}, \tag{25}$$

with $\overline{y}_g = \frac{1}{n_g} \sum_{j=1}^{n_g} y_{jg}$, and the 2SLS estimator for β is

$$\begin{aligned} \widehat{\beta}_{2sls} &= \left(x' P_Z x\right)^{-1} x' P_Z y \\ &= \frac{\sum_{g=1}^{G} n_g \overline{x}_g \overline{y}_g}{\sum_{g=1}^{G} n_g \overline{x}_g^2} \\ &= \frac{\sum_{g=1}^{G} n_g \overline{x}_g^2 \left(\overline{y}_g / \overline{x}_g\right)}{\sum_{g=1}^{G} n_g \overline{x}_g^2} = \sum_{g=1}^{G} \widehat{w}_{2sls,g} \widehat{\beta}_g, \end{aligned}$$

the standard result that $\widehat{\beta}_{2sls}$ is a linear combination of the instrument specific IV estimators, (see e.g. Windmeijer, 2019). The weights are given by

$$\widehat{w}_{2sls,g} = \frac{n_g \overline{x}_g^2}{\sum_{s=1}^{G} n_s \overline{x}_s^2} \geq 0 \tag{26}$$

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 given by

$$\widehat{F}_g = \frac{\widehat{\pi}_g^2}{V\widehat{a}r\left(\widehat{\pi}_g\right)} = \frac{n_g \overline{x}_g^2}{\widehat{\sigma}_{v_2,g}^2} \tag{27}$$

for $g = 1, ..., G$. As the errors $(u_{jg}, v_{2,jg})$ are iid within groups, the standard weak-instruments results of Staiger and Stock (1997) and Stock and Yogo (2005) apply to each group-specific IV estimator $\widehat{\beta}_g$. As these are just-identified models, we can relate the values of the F-statistics to Wald-test size distortions.

From (26) and (27) it follows that the weights for the 2SLS estimator are related to the individual F-statistics as follows

$$\widehat{w}_{2sls,g} = \frac{n_g \overline{x}_g^2}{\sum_{s=1}^{G} n_s \overline{x}_s^2} = \frac{\widehat{\sigma}_{v_2,g}^2 \widehat{F}_g}{\sum_{s=1}^{G} \widehat{\sigma}_{v_2,s}^2 \widehat{F}_s}. \tag{28}$$

Under first-stage homoskedasticity, $\sigma_{v_2,g}^2 = \sigma_{v_2,s}^2$, for $g, s = 1, \ldots, G$, then $\widehat{\sigma}_{v_2,s}^2 \approx \widehat{\sigma}_{v_2,s}^2$ for all g, s, and hence $\widehat{F} \approx \frac{1}{G} \sum_{g=1}^{G} \widehat{F}_g$. Then the weights are given by $\widehat{w}_{2sls,g} \approx \frac{\widehat{F}_g}{\sum_{s=1}^{G} \widehat{F}_{\pi_g}} \approx \frac{\widehat{F}_g}{G\widehat{F}}$, so we see that 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 equal sized groups with approximately the same value of the signal $\widehat{\pi}_g^2$, $\widehat{F}_g$ is larger for groups with, and because of, the smaller values of $\widehat{\sigma}_{v_2,g}^2$. The 2SLS weights ignore this information and give in this example approximately equal weights to groups with similar values of $\widehat{\pi}_g^2$. In practice it could then be the case that a small variance, large $\widehat{F}_g$ group could receive a small weight in the 2SLS estimator. As shown in the Monte Carlo exercises below, this is exactly what happens in the design of Andrews (2018). There is one group with a large individual F-statistic. However, this group has a very small population variance $\sigma_{v_2,g}^2$ resulting in a relatively small weight in the 2SLS estimator and a poor performance of the estimator in terms of bias and size of the Wald-test.

The non-robust F-statistic for π is given by

$$\begin{aligned}
\widehat{F} &= \frac{1}{G}\widehat{\pi}'\left(V\widehat{a}r\left(\widehat{\pi}\right)\right)^{-1}\widehat{\pi} \\
&= \frac{1}{G}\frac{\sum_{g=1}^{G} n_g \overline{x}_g^2}{\left(\sum_{g=1}^{G}\frac{n_g}{n}\widehat{\sigma}_{v_2,g}^2\right)} = \frac{1}{G}\sum_{g=1}^{G}\frac{\widehat{\sigma}_{v_2,g}^2}{\left(\sum_{s=1}^{G}\frac{n_s}{n}\widehat{\sigma}_{v_2,s}^2\right)}\widehat{F}_g.
\end{aligned}$$

The effective F-statistic is given by

$$\widehat{F}_{\text{eff}} = \sum_{g=1}^{G}\frac{\widehat{\sigma}_{v_2,g}^2}{\left(\sum_{s=1}^{G}\widehat{\sigma}_{v_2,s}^2\right)}\widehat{F}_g,$$

and so $\widehat{F}_{\text{eff}} = \widehat{F}$ if groups sizes are equal, $n_g = n/G$ for $g = 1, \ldots, G$. In the designs of Andrews (2018) group sizes are equal in expectation. Both these F-statistics will therefore correctly reflect weak-instruments problems for the 2SLS estimator in these designs.

The robust first-stage F-statistic is given by

$$\begin{aligned}
\widehat{F}_r &= \frac{1}{G}\widehat{\pi}'\left(V\widehat{a}r_r\left(\widehat{\pi}\right)\right)^{-1}\widehat{\pi} \\
&= \frac{1}{G}\sum_{g=1}^{G}\frac{n_g\overline{x}_g^2}{\widehat{\sigma}_{v_2,g}^2} = \frac{1}{G}\sum_{g=1}^{G}\widehat{F}_g.
\end{aligned}$$

It is therefore clear, that if $\widehat{F}_r$ is large, then at least one of the $\widehat{F}_g$ is large. For the GMMf estimator we have that

$$\widehat{\beta}_{gmmf} = \frac{\sum_{g=1}^{G} n_g \overline{x}_g \overline{y}_g / \widehat{\sigma}^2_{v2,g}}{\sum_{g=1}^{G} n_g \overline{x}_g^2 / \widehat{\sigma}^2_{v2,g}} = \sum_{g=1}^{G} \frac{n_g \overline{x}_g^2 / \widehat{\sigma}^2_{v2,g}}{\sum_{s=1}^{G} n_s \overline{x}_s^2 / \widehat{\sigma}^2_{v2,s}} \widehat{\beta}_g = \sum_{g=1}^{G} \widehat{w}_{gmmf,g} \widehat{\beta}_g, \tag{29}$$

with

$$\widehat{w}_{gmmf,g} = \frac{\widehat{F}_g}{\sum_{s=1}^{G} \widehat{F}_s} = \frac{\widehat{F}_g}{G \widehat{F}_r},$$

hence the groups with the larger F-statistics get the larger weights, independent of the values of $\widehat{\sigma}^2_{v2,s}$, mimicking the 2SLS weights under homoskedasticity of the first-stage errors.

6.1 Some Monte Carlo Results

We consider here the two heteroskedastic designs of Andrews (2018) with $G = 10$ groups, $\beta = 0$ and one with moderate and one with high endogeneity. Tables 9 and 12 in the Supplementary Appendix C.3 of Andrews (2018) present the values of the conditional group-specific variance matrices Σ_g as defined in (21) and the first-stage parameters, denoted π_{0g}, for $g = 1, \ldots, 10$. The correlation between u_i and v_{2i} is -0.59 in the moderate and we set it equal to 0.99 in the high endogeneity case. We multiply the first-stage parameters π_0 by 0.040 and 0.026, such that the value of the robust $\widehat{F}_r$ is just over 80 on average for $10,000$ replications and sample size $n = 10,000$ in both designs. The group sizes are equal in expectation with $P(z_i = e_g) = 0.1$ for all g. The first two rows in each panel of Table 3 present the values of π_g and $\sigma^2_{v_2,g}$ for $g = 1, \ldots, 10$.

Tables 1 and 2 presents the weak-instruments test results and estimation results. We use and present the critical values based on the OLS bias benchmark, $\text{BM}_{LS}\left(\widehat{W}_{\Omega_n}, \widehat{\Sigma}_v\right)$, and set $\tau = 0.10$. The means of the effective F-statistics are small in both designs, and the $\widehat{F}_{\text{eff}}$-based test does not reject the null of weak instruments for the 2SLS estimator in any of the replications. This is reflected in the bias of the 2SLS estimator and the 2SLS-based Wald test for $H_0 : \beta = 0$ overrejects. The means and standard deviations of the non-robust F-statistics are virtually the same as those of the effective F-statistics, confirming the results derived in Section 6. The means of the robust F-statistics are large, and the $\widehat{F}_r$-based test for weak instruments rejects the null of weak instruments in all replications. The GMMf estimator is virtually unbiased and the GMMf-based Wald

Table 1: Weak-instruments test results

	$\widehat{F}$	$\widehat{F}_{\text{eff}}$	cv_{eff}^{LS}	RF	$\widehat{F}_r$	cv_r^{LS}	RF
Mod Endog	1.411	1.411	17.09	0	80.23	13.45	1
	(0.738)	(0.739)	(0.110)		(7.025)	(0.290)	
High Endog	0.993	0.993	17.12	0	80.12	12.26	1
	(0.599)	(0.599)	(0.113)		(7.081)	(0.010)	

Notes: Means and (st.dev.), of $10,000$ replications. $n = 10,000$, $\tau = 0.1$, Rej.freq. (RF) at 5% level.

Table 2: Estimation results

	$\widehat{\beta}_{ols}$	$\widehat{\beta}_{2sls}$	$\widehat{\beta}_{gmmf}$	$Wald_{2sls}$	$Wald_{gmmf}$
Mod Endog	-0.608	-0.424	-0.001	0.534	0.049
	(0.011)	(0.257)	(0.563)		
High Endog	0.747	0.742	0.007	0.999	0.065
	(0.001)	(0.057)	(0.029)		

Notes: Bias and (st.dev.). Rej.freq. of robust Wald tests at 5% level.

test has good size properties. The means of the critical values for $\widehat{F}_{\text{eff}}$ are 17.09 and 17.12, whereas those for $\widehat{F}_r$ are 13.45 in the moderate and 12.26 in the high endogeneity design. In comparison, the 10% relative bias Stock and Yogo critical value for the homoskedastic case is here given by 11.46, see Skeels and Windmeijer (2018).

The details as given in Table 3 below make clear what is happening. It reports the population values of π_g, $\sigma^2_{v_2,g}$, $\mu^2_{n,g} = 1000\pi^2_g/\sigma^2_{v_2,g}$ and the mean values of $\widehat{F}_g$, $\widehat{w}_{2sls,g}$ and $\widehat{w}_{gmmf,g} = \widehat{F}_g / \sum_{s=1}^{G} \widehat{F}_g$. For the moderate endogeneity design identification in the first group is strong, with an average value of $\widehat{F}_1 = 789.5$. Identification in all other 9 groups is very weak, with the largest average value for $\widehat{F}_5 = 2.23$. The signal for group 1, π_1^2, is somewhat larger than those for the other groups, but the population value $\mu^2_{n,1}$ is large mainly due to the relatively very small value of $\sigma^2_{v_2,1}$. As detailed in (28), the 2SLS weights ignore the $\sigma^2_{v_2,1}$ part of the information in group 1 which leads to the low average value of $\widehat{w}_{2sls,1} = 0.127$. This shows that the 2SLS estimator does not utilize the identification strength of the first group well, with some larger weights given to higher variance, but lower concentration-parameter groups.

Table 3 further shows that for the GMMf estimator almost all weight is given to the first group, with the average of $\widehat{w}_{gmmf,1}$ 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 $\sigma^2_{v_2,1}$. Reducing the value of $\sigma^2_{u,1}$ (and the value for $\sigma_{uv_2,1}$ accordingly to keep the same

Table 3: Group information and estimator weights

	g	1	2	3	4	5	6	7	8	9	10
ME	π_g	0.058	-0.023	0.049	0.015	0.022	0.008	-0.017	0.011	-0.036	-0.040
	$\sigma^2_{v_2,g}$	0.004	2.789	4.264	0.779	0.395	7.026	1.226	0.308	1.709	6.099
	$\mu^2_{n,g}$	785.7	0.184	0.556	0.284	1.190	0.009	0.236	0.387	0.770	0.266
	$\widehat{F}_g$	789.5	1.170	1.564	1.279	2.225	0.997	1.203	1.372	1.798	1.246
	$\widehat{w}_{2sls,g}$	0.126	0.098	0.178	0.035	0.031	0.180	0.049	0.015	0.096	0.192
	$\widehat{w}_{gmmf,g}$	0.984	0.002	0.002	0.002	0.003	0.001	0.002	0.002	0.002	0.002
HE	$100 \cdot \pi_g$	-0.021	0.095	-0.484	-0.069	0.159	-0.028	0.101	-0.418	0.450	-0.546
	$\sigma^2_{v_2,g}$	1.600	0.478	2.975	1.142	0.174	0.145	4.658	1.963	2.990	$0.38{\cdot}a$
	$\mu^2_{n,g}$	$0.28{\cdot}a$	0.002	0.008	$4.2{\cdot}a$	0.015	$5.6{\cdot}a$	$2.2{\cdot}a$	0.009	0.007	789.9
	$\widehat{F}_g$	0.998	1.017	0.979	1.010	1.034	0.984	0.977	1.031	0.997	792.2
	$\widehat{w}_{2sls,g}$	0.111	0.040	0.177	0.085	0.016	0.013	0.242	0.134	0.181	0.003
	$\widehat{w}_{gmmf,g}$	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.989

Notes: $\mu^2_{n,g} = 1000\pi^2_g/\sigma^2_{v_2,g}$; $a = 10^{-4}$

correlation structure within group 1), will reduce the standard deviation of the GMMf estimator.

The pattern of group information for the high endogeneity case is similar to that of the moderate endogeneity case, with one informative group, $g = 10$, with an average value of $\widehat{F}_{10} = 792.2$. However, the variance $\sigma^2_{v_2,10}$ is now so small in relative terms, that the 2SLS weight for group 10 has an average value of only $\widehat{w}_{2sls,10} = 0.003$. The GMMf estimator corrects this, with the average value of $\widehat{w}_{gmmf,10} = 0.989$. The standard deviation of the GMMf estimates, 0.029, is in this case smaller than that of the 2SLS estimates, 0.057.

The left panels of Figure 1 displays the relative bias of the 2SLS and GMMf estimators, relative to that of the OLS estimator, as a function of the mean values of the robust F-statistic $\widehat{F}_r$, together with the rejection frequency of the $\widehat{F}_r$-based test for weak instruments, using the critical values from the least-squares benchmark bias. We present the relative bias here to be in line with the homoskedastic case as presented below. Different values of $\widehat{F}_r$ are obtained by different values of the scalar e when setting the first-stage parameters $\pi = e\pi_0$. The relative bias of the GMMf estimator decreases quite rapidly with increasing values of $\widehat{F}_r$. For the moderate endogeneity case, the test has a rejection frequency of 5% at a mean $\widehat{F}_r$ of 10.03, with the relative bias of the GMMf estimator at that point equal to 0.092. As shown in the top right-hand panel of Figure 1, the GMMf estimator based Wald test is well behaved in terms of size, with hardly any size distortion for mean values of $\widehat{F}_r$ larger than 5. The GMMf relative bias picture for the

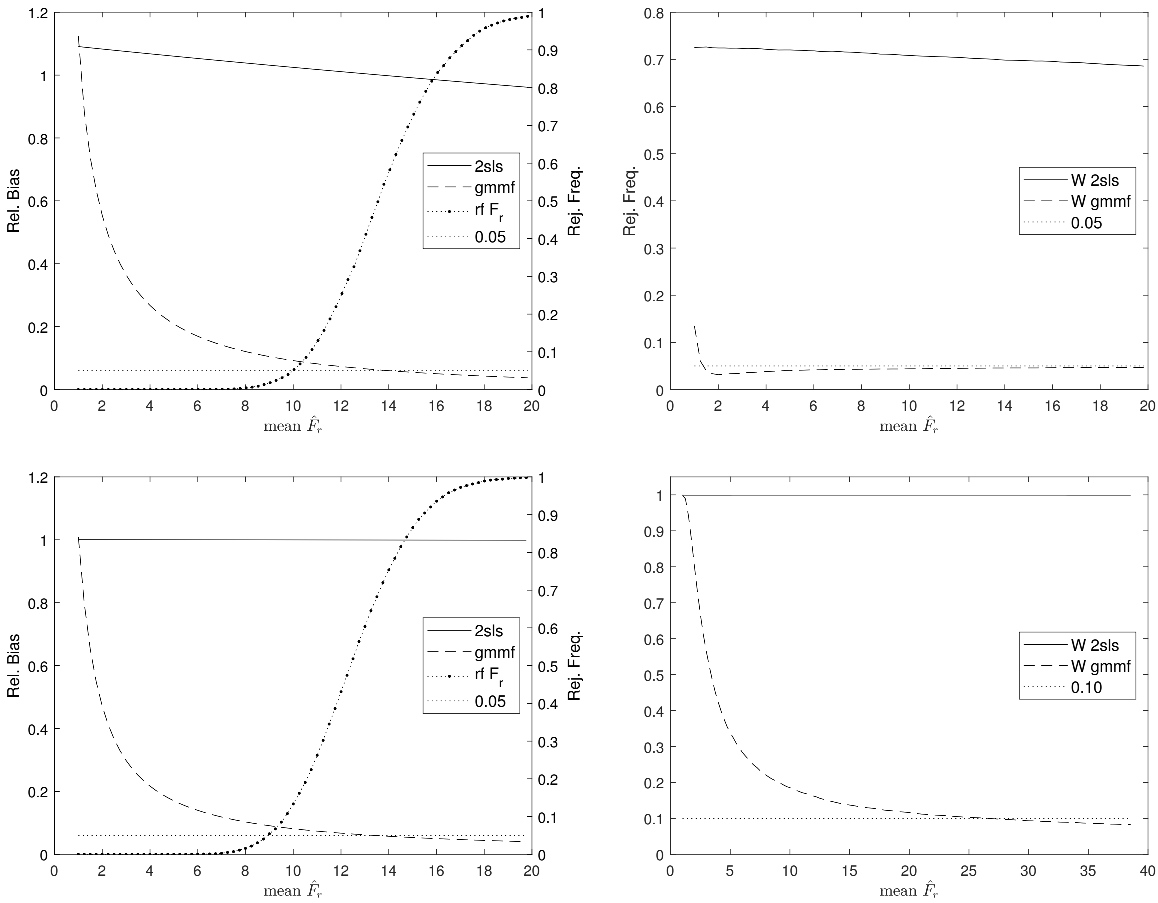

Figure 1: Heteroskedastic design. Top: Moderate Endogeneity. Bottom: High Endogeneity. Left: Bias of 2SLS and GMMf estimators relative to OLS bias, and $\widehat{F}_r$-based weak-instrument test rejection frequencies, $\alpha = 0.05$, $\tau = 0.10$, least-squares benchmark bias. The "0.05" line refers to the rej. freq. Right: Rejection frequencies of robust Wald tests, $\alpha = 0.05$.

high-endogeneity case is very similar to that of the moderate-endogeneity case. Here the $\widehat{F}_r$ based test for weak instruments has a rejection frequency of 5% at a mean $\widehat{F}_r$ of 8.98, with the relative bias there being 0.091. As for the homoskedastic case, where the Wald test size deviation from nominal size is larger for larger values of $\rho^2_{uv_2}$, the GMMf Wald test has a worse size performance in the high-endogeneity design, and has a 10% rejection frequency at a mean $\widehat{F}_r$ of 26.64. This would imply a critical value at the 5% level of around 32, which compares to the Stock and Yogo weak-instruments critical value of 38.54 for a Wald test size of 10% at the 5% nominal level.

6.1.1 Homoskedastic Design

We next consider the homoskedastic design for the moderate endogeneity case with $\Sigma_{uv_2} = \frac{1}{G}\sum_{g=1}^{G}\Sigma_{uv_2,g}$, resulting in

$$\Sigma_{uv_2} = \begin{bmatrix} 2.57 & -1.50 \\ -1.50 & 2.46 \end{bmatrix},$$

with $\rho_{uv_2} = -0.59$, as above. We consider smaller sample sizes of $n = 250$ and $n = 500$, or group sizes of 25 or 50 on average, to compare the weak-instrument finite sample behaviour of the GMMf estimator to that of the 2SLS estimator. In particular, the noise induced by estimation of W_2^{-1} may adversely affect the GMMf estimator.

The results in Figure 2 shows that for this design and sample sizes the relative biases and Wald rejection frequencies are virtually identical for the two estimators, with the standard deviations of the GMMf estimates slightly larger than those of the 2SLS estimator, as expected. The rejection frequencies of the $\widehat{F}_r$-based test are here closer to those of the standard Stock and Yogo $\widehat{F}$-based test compared to the rejection frequencies of the $\widehat{F}_{\text{eff}}$-based test, with the latter test more conservative.

6.2 Relative Values of $\widehat{F}_{\text{eff}}$ and $\widehat{F}_r$ and Biases of 2SLS and GMMf Estimators

From the results on the relative magnitude of the concentration parameters as discussed in Section 4 and footnote 2, we can change the parameter values in the grouped-data design of Section 6.1 such that $\widehat{F}_{\text{eff}} > \widehat{F}_r$ in expectation. The design is described in Appendix A.1 where Table A1 presents Monte Carlo results for the weak-instruments tests and estimation results. The null of weak instruments is rejected in all replications for the 2SLS estimator, whereas it is not rejected in virtually all replications for the GMMf estimator and the bias of the 2SLS estimator is smaller than that of the GMMf estimator.

As per the results of Tables 1 and 2, a large value of $\widehat{F}_r \gg cv_r^{LS}$ indicates that the GMMf estimator is quite well behaved in terms of bias. But when that is the case, a value of $\widehat{F}_{\text{eff}} < cv_{\text{eff}}^{LS}$, and so not rejecting the null of weak instruments for the 2SLS estimator, does not necessarily imply that the 2SLS bias is larger than the GMMf bias (and vice versa). Consider an example with fixed group sizes $\frac{n_g}{n} = f_g$, $0 < f_g < 1$, for $g = 1, \ldots, G$.

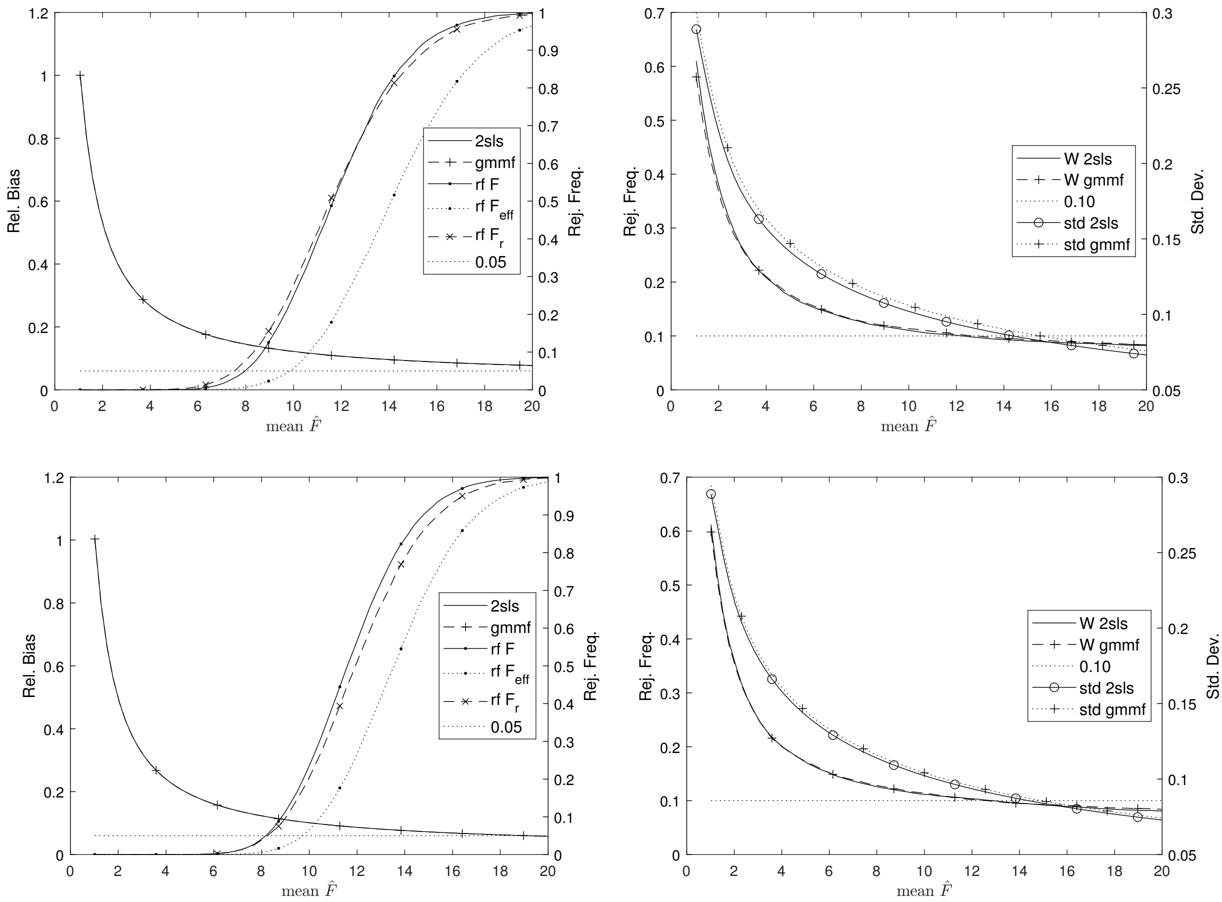

Figure 2: Homoskedastic design, moderate endogeneity. Top: $n = 250$. Bottom: $n = 500$. Left: Relative bias and weak-instrument tests rejection frequencies. Right: Wald test rejection frequencies and standard deviations of 2SLS and GMMf estimates.

Then for 2SLS,

$$\mu_{2sls}^2 := \mu_{Q_{zz}^{-1}}^2 = \frac{\sum_{g=1}^G c_g^2 f_g}{\sum_{g=1}^G \sigma_{v_2,g}^2},$$

and for GMMf,

$$\mu_{gmmf}^2 := \mu_{W_2^{-1}}^2 = \frac{1}{G}\sum_{g=1}^G c_g^2 f_g / \sigma_{v_2,g}^2.$$

Then the Nagar bias (16) for the 2SLS and GMMf estimators are respectively given by

$$N_{2sls} = \frac{\sum_{g=1}^G \left(1 - 2\frac{c_g^2 f_g}{\sum_{s=1}^G c_s^2 f_s}\right)\sigma_{uv_2,g}}{\sum_{g=1}^G c_g^2 f_g}$$

$$N_{gmmf} = \frac{\sum_{g=1}^G \left(1 - 2\frac{c_g^2 f_g / \sigma_{v_2,g}^2}{\sum_{s=1}^G c_s^2 f_s / \sigma_{v_2,s}^2}\right)\frac{\sigma_{uv_2,g}}{\sigma_{v_2,g}^2}}{\sum_{g=1}^G c_g^2 f_g / \sigma_{v_2,g}^2}.$$

One can find therefore parameter values in this design, such that $N_{2sls} < N_{gmmf}$ whereas $\mu_{2sls} < \mu_{gmmf}$. It follows for example, that if $c_g^2 f_g = c_f$, $c_f > 0$ for $g = 1, \ldots, G$, then $N_{2sls} = 0$ if $\sum_{g=1}^{G} \sigma_{uv_2,g} = 0$, irrespective of the value of μ_{2sls}. In that case, there is no overall endogeneity, as $\rho_{uv_2} = 0$. (Conversely, in the constant $c_g^2 f_g$ case, $N_{gmmf} = 0$ if $\sum_{g=1}^{G} \frac{\sigma_{uv_2,g}}{\sigma_{v_2,g}^4} = 0$ irrespective of the value of μ_{gmmf}).

As another example, consider again $G = 10$, with $f_g = 0.1$ for $g = 1, \ldots, 10$, with parameter values c_g and group variance matrices $\Sigma_{uv_2,g}$ given in Appendix A.2. The resulting overall endogeneity is a moderate $\rho_{uv_2} = 0.244$. We get a value of $\mu_{gmmf}^2 = 43.09$ and a smaller value of $\mu_{2sls}^2 = 8.45$, whereas the Nagar bias approximations are virtually identical at $N_{2sls} = 0.022 < N_{gmmf} = 0.023$. Table A3 in Appendix A.2 presents some Monte Carlo results for this design, for $n = 10,000$, confirming the theoretical results. We find there that the null of weak instruments is not rejected in virtually all replications for the 2SLS estimator, whereas it is rejected in all replications for the GMMf estimator. But the Monte Carlo bias of the 2SLS estimator, 0.022, is slightly smaller than that of the of the GMMf estimator, 0.024, as indicated by the Nagar bias approximations.

From these latter examples it is clear that a situation with $\widehat{F}_r > cv_r^{LS}$ and $\widehat{F}_{\text{eff}} < cv_{\text{eff}}^{LS}$ does not necessarily imply that the 2SLS estimator is more biased than the GMMf estimator (and vice versa). However, randomly sampling values c_g and group variance matrices $\Sigma_{uv_2,g}$, for $g = 1, \ldots, 10$, and collecting the first 1000 of those sets where $|\rho_{uv_2}| > 0.2$, $5 < \mu_{2sls} < 10$ and $40 < \mu_{gmmf} < 45$, we find for 989, or 99% of those, the Nagar bias of the 2SLS estimator to be larger than that of the GMMf estimator.

7 Considerations for Practice

The Andrews (2018) grouped-data IV designs are quite extreme in the variation of $\sigma_{v_2,g}^2$, leading to the large differences between the values of $\widehat{F}_{\text{eff}}$ and $\widehat{F}_r$ and between the performances of the 2SLS and GMMf estimators. Note that these results carry over to a model with a constant and a full set of mutually exclusive binary indicators as instruments, when the variances $\sigma_{v_2,g}^2$ for at least two groups are relatively small and their π_g coefficients are different. This is the case if we for example change $\sigma_{v_2,2}^2$ in the moderate endogeneity design above to be equal to the small $\sigma_{v_2,1}^2$. An example where this could be relevant is the judge fixed effects design as pioneered by Kling (2006). There are many papers using judge effects as instrumental variables, Table 1 in Frandsen et al. (2023) presents a subset of these. Stevenson (2018) studied the effect of pretrial detention on conviction, using judge indicators as instruments, as cases are randomly assigned to judges. As the treatment is here binary, with variance $\pi(1-\pi)$, a very lenient (small π) and a very strict judge (large π) in terms of sending defendants to pretrial detention have small values of $\sigma^2_{v_2}$, but clearly different values of π. Unlike the 2SLS estimator, the GMMf estimator takes the differential strengths of the instruments due to the different values of $\sigma^2_{v_2}$ into account, giving more weight to very lenient and very strict judges.[3]

For any single-endogenous variable application, most packages will compute the robust F-statistic for estimation results with robust standard errors. For example, in Stata, StataCorp. (2023), the robust first-stage F-statistic is provided with the output of "ivregress" or "ivreg2", Baum, Schaffer, and Stillman (2010), whereas "weakivtest", Pflueger and Wang (2015), calculates the effective F-statistic and critical values for the weak-instruments test. An extended version of the latter, called "gfweakivtest"[4] also calculates the robust F-statistic and its weak-instruments critical values. It further includes the critical values based on the least-squares benchmark bias for both $\widehat{F}_r$ and $\widehat{F}_{\text{eff}}$, and presents the estimation results for the GMMf estimator. As per the advice of Andrews et al. (2019), 2SLS estimation results with robust standard errors should be accompanied by the effective F-statistic and its critical value. If the situation is as in the Andrews (2018) examples above, then the GMMf estimator could be the preferred estimator.

Windmeijer (2023) presents such comparisons for a study with a set of mutually exclusive binary indicators as instruments, and one of the *American Economic Review* studies as considered in the review paper by Andrews et al. (2019). This is from Stephens and Yang (2014) who study the effect of schooling on wages, using data from the 1960-1980 US Censuses of Population. The endogenous variable is years of schooling for individual i, born in state s in year t, and the instruments are three indicator variables $RS7$, $RS8$ and $RS9$, corresponding to being required to attend seven, eight or nine or more years of schooling, respectively. All specifications include state-of-birth and year-of-birth fixed effects, and the computed standard errors are robust to heteroskedasticity and clustering at the state-of-birth/year-of-birth cell. Stephens and Yang (2014) report the robust first-stage F-statistics $\widehat{F}_r$ in their Table 1, which presents eight sets of estimates of the returns of schooling on log weekly wages for four different samples and two different model specifications for each sample. None of the $\widehat{F}_r$ statistics indicate an underidentification problem. But no effective F-statistics were reported. As the estimator used is the 2SLS estimator, it

[3]I would like to thank an anonymous referee for this example.
[4]Available from the author upon request.

is therefore important to consider whether the $\widehat{F}_r$ statistic misrepresents weak-instruments bias of the 2SLS estimator, in the sense that a large value of $\widehat{F}_r$ may not be an indicator of a good performance of the 2SLS estimator. Table 3 in Windmeijer (2023) shows that this is not the case here. The $\widehat{F}_{\text{eff}}$ and $\widehat{F}_r$ based tests for weak instruments both reject and don't reject the null of weak instruments for the same specifications, their values are similar in magnitude and the 2SLS and GMMf estimation results are virtually identical when the null of weak instruments is rejected and the Hansen J-test does not indicate misspecification. This is a reassuring result for the 2SLS estimates that were accompanied by larger values of $\widehat{F}_r$, with the only two cases where the null of weak instruments was not rejected had values of $\widehat{F}_r$ equal to 8.22 and 6.34 with those of $\widehat{F}_{\text{eff}}$ equal to 8.11 and 6.13 respectively.

8 Concluding Remarks

For models with a single endogenous explanatory variable, we have introduced a class of generalized effective F-statistics as defined in (11) in relation to a class of linear GMM estimators given in (9) and have shown that the Montiel Olea and Pflueger (2013) weak-instruments testing procedure that they established for the effective F-statistic in relation to the Nagar bias of the 2SLS estimator applies to this extended class. In particular, the standard robust F-statistic is a member of this class and is associated with the behaviour in terms of Nagar bias of the GMMf estimator, which has its weight matrix based on the first-stage residuals. We then focused on a comparison of the effective F-statistic and the robust F-statistic and the associated weak-instrument behaviours of the 2SLS and GMMf estimators. In particular, we have shown that and explained why the GMMf estimator's performance is much better in terms of bias than that of the 2SLS estimator in the grouped-data designs of Andrews (2018), where the robust F-statistic can take very large values, but the effective F-statistic is very small. One should therefore in general not use the robust F-statistic to gauge instrument strength in relation to the performance of the 2SLS estimator, Andrews et al. (2019, pp 738-739), but as shown here, it can be used as a weak-instruments test in relation to the Nagar bias of the GMMf estimator. In practice, therefore, both the effective F-statistic and robust F-statistic should be reported, together with their critical values, and the GMMf estimator could be considered in cases where there is a clear discrepancy with a large value for the robust F-statistic rejecting the null of weak instruments, and when the effective F-statistic is small and does not reject its null of weak instruments.

We have not focused here on the wider applicability of the class of generalized effective F-statistics and their associated GMM estimators, but an example is the one-step Arellano and Bond (1991) GMM estimator for panel data models with a single endogenous variable. Two-step estimators do not fall in the class because of the presence of estimated structural parameters in the weight matrix, but one could test for weak instruments in this setting, fixing the parameter of the endogenous variable in the weight matrix, for example under a specific null value of interest.

A topic for future research for the general heteroskedasticity setting is an extension to the linear model with more than one endogenous variable. Lewis and Mertens (2022) is an extension of the Montiel Olea and Pflueger (2013) method to the multiple endogenous variable case for the 2SLS estimator, but they do not consider such an extension for the wider class of GMM estimators. Future research should also address the weak-instruments Wald size properties for both the single and multiple endogenous variables settings.

Acknowledgments

I would like to thank the editor Xiaohong Chen, an associate editor and two anonymous referees for their detailed comments and suggestions that helped to improve the paper substantially. Further thanks to Benedikt Koch and seminar participants for helpful comments.

Appendix

A.1 A Grouped-Data Design with $\widehat{F}_{\text{eff}} > \widehat{F}_r$

From the results on the relative magnitude of the concentration parameters as discussed in Section 4 and footnote 2, we can change the parameter values such that $\widehat{F}_{\text{eff}} > \widehat{F}_r$ in expectation in the grouped-data design. Here we take the moderate-endogeneity design of Andrews (2018), but change the value of $\sigma^2_{v_2,1}$ from 0.004 to 20, adjusting $\sigma^2_{u,1}$ and $\sigma_{uv_2,1}$ accordingly, and change the value of π_1 from 0.058 to 1.414. This results in the expected group specific concentration parameter $\mu^2_{n,1} = 100$, and so the first group is again the informative group, but has now a relatively large variance $\sigma^2_{v_2}$ and a relatively large signal π^2_1. The estimation results are given in Table A1. $\widehat{F}_{\text{eff}}$ indicates that there is no weak-instruments problem for 2SLS, whereas $\widehat{F}_r$ shows that there is for the GMMf estimator. The 2SLS estimator is virtually unbiased and the 2SLS Wald test is well behaved, whereas the GMMf estimator displays bias with the its Wald test overrejecting. In this case, the 2SLS estimator gives a large weight of 0.98 to group 1, whereas this weight is 0.89 for GMMf.

Table A1: Weak-instruments tests and estimation results, $\sigma^2_{v_2,1} = 20$, $\pi_1 = 1.414$

$\widehat{F}$	$\widehat{F}_{\text{eff}}$	cv^{LS}_{eff}	RF	$\widehat{F}_r$	cv^{LS}_r	RF
46.10	46.12	18.55	1	11.42	16.03	0.025
(9.11)	(9.17)	(0.116)		(2.16)	(0.191)	
$\widehat{\beta}_{ols}$	$\widehat{\beta}_{2sls}$	$\widehat{\beta}_{gmmf}$	$Wald_{2sls}$	$Wald_{gmmf}$		
-0.316	-0.007	-0.065	0.056	0.217		
(0.011)	(0.024)	(0.061)				

Notes: See notes to Tables 1 and 2. Moderate Endogeneity.

A.2 Parameter Values for Design of Section 6.2

The parameter values c_g and $\Sigma_{uv_2} = \begin{bmatrix} \sigma^2_{u,g} & \sigma_{uv_2,g} \\ \sigma_{uv_2,g} & \sigma^2_{v_2,g} \end{bmatrix}$ are given by

Table A2: Parameter values.

g	c_g	$\sigma^2_{u,g}$	$\sigma_{uv_2,g}$	$\sigma^2_{v_2,g}$
1	20.6393	9.0052	1.7135	4.2487
2	27.6284	3.4060	1.7847	9.9668
3	-3.3019	2.3741	2.8222	6.0015
4	-38.7569	1.7522	-0.7409	0.4370
5	-11.1463	3.5420	-2.4995	8.6788
6	18.2092	3.2771	3.0059	4.0456
7	-0.4646	0.0538	0.3084	6.9979
8	25.0219	6.2319	4.8593	8.2675
9	-25.6606	5.8019	-0.4336	4.2698
10	5.9592	7.3973	0.8086	0.0968

Table A3 presents some Monte Carlo results for this design, for $n = 10,000$, confirming the theoretical results. We find here that the null of weak instruments is not rejected in virtually all replications for the 2SLS estimator, whereas it is rejected in all replications for the GMMf estimator. But the Monte Carlo bias of the 2SLS estimator, 0.022, is slightly smaller than that of the of the GMMf estimator, 0.024, as indicated by the Nagar bias approximations as detailed in Section 6.2.

Table A3: Weak-instruments tests and estimation results

$\widehat{F}_{\text{eff}}$	cv_{eff}^{LS}	RF	$\widehat{F}_r$	cv_r^{LS}	RF
9.49	15.85	0.00	44.24	19.47	1
(1.83)	(0.10)		(4.45)	(0.15)	
$\widehat{\beta}_{ols}$	$\widehat{\beta}_{2sls}$	$\widehat{\beta}_{gmmf}$	$Wald_{2sls}$	$Wald_{gmmf}$	
0.218	0.022	0.024	0.062	0.049	
(0.009)	(0.092)	(0.149)			

Notes: See notes to Tables 1 and 2. Design as in text and Appendix.

A.3 Proof of Lemma 1

It follows from the first-stage and reduced-form model specifications (2) and (3) and Assumptions 1 and 2 that, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\begin{pmatrix} \Omega_n^{1/2} Z'y \\ \Omega_n^{1/2} Z'x \end{pmatrix} = \begin{pmatrix} \Omega_n^{1/2}\left(\frac{1}{n}Z'Z\right)c\beta + \frac{1}{\sqrt{n}}\Omega_n^{1/2}Z'v_1 \\ \Omega_n^{1/2}\left(\frac{1}{n}Z'Z\right)c + \frac{1}{\sqrt{n}}\Omega_n^{1/2}Z'v_2 \end{pmatrix}$$
$$\xrightarrow{d} \begin{pmatrix} \gamma_{\Omega,1} \\ \gamma_{\Omega,2} \end{pmatrix} := \begin{pmatrix} c_\Omega\beta + \Omega^{1/2}\psi_1 \\ c_\Omega + \Omega^{1/2}\psi_2 \end{pmatrix} \sim N\left(\begin{pmatrix} c_\Omega\beta \\ c_\Omega \end{pmatrix}, W_\Omega\right), \quad \text{(A.1)}$$

where $c_\Omega = \Omega^{1/2}Q_{zz}c$.

As

$$\widehat{\beta}_{\Omega_n} = \frac{x'Z\Omega_n Z'y}{x'Z\Omega_n Z'x},$$

it follows that

$$\widehat{\beta}_{\Omega_n} - \beta = \frac{x'Z\Omega_n Z'u}{x'Z\Omega_n Z'x} = \frac{x'Z\Omega_n Z'\left(v_1 - \beta v_2\right)}{x'Z\Omega_n Z'x}.$$

Under Assumptions 1 and 2 it follows from (A.1) and the continuous mapping theorem that

$$\widehat{\beta}_{\Omega_n} - \beta \xrightarrow{d} \beta_\Omega^* = \left(\gamma_{\Omega,2}'\gamma_{\Omega,2}\right)^{-1}\gamma_{\Omega,2}'\left(\gamma_{\Omega,1} - \beta\gamma_{\Omega,2}\right).$$

As

$$\widehat{F}_{\text{geff}}\left(\Omega_n\right) = \frac{x'Z\Omega_n Z'x}{n\text{tr}\left(\widehat{W}_2\Omega_n\right)}$$

it follows Assumptions 1 and 2 that

$$\widehat{F}_{\text{geff}}\left(\Omega_n\right) \xrightarrow{d} \frac{\gamma_{\Omega,2}'\gamma_{\Omega,2}}{\text{tr}\left(W_2\Omega\right)}$$

again from (A.1) and the continuous mapping theorem.

A.4 Proof of Nagar Bias Result of Theorem 1

The Nagar bias result is obtained as follows. Let $S(\beta, W_\Omega)$ be as defined in (14), and

$$\xi = S_1^{-1/2}(\beta, W_\Omega)(\gamma_{\Omega,1} - \beta\gamma_{\Omega,2}) \sim \mathcal{N}(0,1),$$
$$\nu = W_{\Omega,2}^{-1/2}(\gamma_{\Omega,2} - c_\Omega) \sim \mathcal{N}(0,1).$$

We then have that

$$\beta_\Omega^* = \frac{\gamma_{\Omega,2}'(\gamma_{\Omega,1} - \beta\gamma_{\Omega,2})}{\gamma_{\Omega,2}'\gamma_{\Omega,2}}$$
$$= \frac{c_\Omega' S_1^{1/2}(\beta, W_\Omega)\xi + \nu' W_{\Omega,2}^{1/2} S_1^{1/2}(\beta, W_\Omega)\xi}{c_\Omega' c_\Omega + 2c_\Omega' W_{\Omega,2}^{1/2} S_1^{1/2}\nu + \nu' W_{\Omega,2}\nu}.$$

It follows that

$$\|c_\Omega\|\beta_\Omega^* = \frac{c_{\Omega,0}' S_1^{1/2}(\beta, W_\Omega)\xi + \frac{\nu' W_{\Omega,2}^{1/2} S_1^{1/2}(\beta, W_\Omega)\xi}{\sqrt{\operatorname{tr}\left(W_{\Omega,2}\right)}}\mu_\Omega^{-1}}{1 + \frac{2c_{\Omega,0}' W_{\Omega,2}^{1/2} S_1^{1/2}\nu}{\sqrt{\operatorname{tr}\left(W_{\Omega,2}\right)}}\mu_\Omega^{-1} + \frac{\nu' W_{\Omega,2}\nu}{\operatorname{tr}\left(W_{\Omega,2}\right)}\mu_\Omega^{-2}},$$

where $\mu_\Omega^2 = \|c_\Omega\|^2/\operatorname{tr}(W_{\Omega,2})$. Then from Rothenberg (1984, (6.2)), we get the second-order Edgeworth, Nagar (1959) approximation

$$\mathbb{E}(\beta_\Omega^*) \approx \frac{1}{\mu_\Omega^2}\frac{1}{\operatorname{tr}(W_{\Omega,2})}\mathbb{E}\left[\nu' W_{\Omega,2}^{1/2} S_1^{1/2}(\beta, W_\Omega)\xi - 2c_{\Omega,0}' S_1^{1/2}(\beta, W_\Omega)\xi c_{\Omega,0}' W_{\Omega,2}^{1/2}\nu\right]$$
$$= \frac{1}{\mu_\Omega^2}\left(\frac{\operatorname{tr}\left(S_{12}(\beta, W_\Omega) - 2c_{\Omega,0}' S_{12}(\beta, W_\Omega)c_{\Omega,0}\right)}{\operatorname{tr}(W_{\Omega,2})}\right).$$

References

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *The Review of Economics and Statistics*, 100, 337–348.

Andrews, I., J. H. Stock, and L. Sun (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753.

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.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277.

Baum, C. F., M. E. Schaffer, and S. Stillman (2010): "ivreg2: Stata Module for Extended Instrumental Variables/2SLS, GMM and AC/HAC, LIML and k-Class Regression," *http://ideas.repec.org/c/boc/bocode/s425401.html*.

Bekker, P. A. and J. Ploeg (2005): "Instrumental Variable Estimation Based on Grouped Data," *Statistica Neerlandica*, 59, 239–267.

Bun, M. and M. de Haan (2010): "Weak Instruments and the First-Stage F-Statistic in IV Models with a Nonscalar Error Covariance Structure," Tech. Rep. Discussion Paper: 2010/02, University of Amsterdam.

Frandsen, B., L. Lefgren, and E. Leslie (2023): "Judging Judge Fixed Effects," *American Economic Review*, 113, 253–277.

Kling, J. R. (2006): "Incarceration Length, Employment, and Earnings," *American Economic Review*, 96, 863–876.

Lewis, D. J. and K. Mertens (2022): "A Robust Test for Weak Instruments with Multiple Endogenous Regressors," Working Papers 2208, Federal Reserve Bank of Dallas.

Montiel Olea, J. L. and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

Nagar, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575.

Patnaik, P. B. (1949): "The Non-Central Chi-Square and F-Distributions and Their Applications," *Biometrika*, 36, 202–232.

Pflueger, C. E. and S. Wang (2015): "A Robust Test for Weak Instruments in Stata," *The Stata Journal*, 216–225.

Rothenberg, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics, Chapter 15*, ed. by Z. Griliches and M. D. Intriligator, Elsevier, vol. 2, 881–935.

Skeels, C. and F. Windmeijer (2018): "On the Stock–Yogo Tables," *Econometrics*, 6, 44.

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557.

StataCorp. (2023): "2023. Stata Statistical Software: Release 18," *College Station, TX: StataCorp LLC.*

Stephens, M. and D.-Y. Yang (2014): "Compulsory Education and the Benefits of Schooling," *American Economic Review*, 104, 1777–1792.

Stevenson, M. T. (2018): "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes," *The Journal of Law, Economics, and Organization.*

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models*, Cambridge University Press, 80–108.

Windmeijer, F. (2019): "Two-Stage Least Squares as Minimum Distance," *The Econometrics Journal*, 22, 1–9.

——— (2023): "The Robust F-Statistic as a Test for Weak Instruments," *arXiv*, 2309.01637.