The Robust F-Statistic as a Test for Weak Instruments

Frank Windmeijer

Dept of Statistics and Nuffield College
University of Oxford

Rotterdam ESEM, August 2024

Introduction

- Classical linear model with a single endogenous variable x and instruments Z,

$$
\begin{aligned}
y &= x\beta + u \\
x &= Z\pi + v_2
\end{aligned}
$$

- Model is overidentified, $k_z > 1$.
- $\mathbb{E}(z_i u_i) = 0$.
- First-stage F-statistic, test for $H_0 : \pi = 0$, also valid test for weak instruments, in terms of bias and/or Wald test size distortion of 2sls, Liml and Fuller, when $(u, v_2|Z)$ homoskedastic (Stock and Yogo, 2005). Rule of thumb, $F > 10$.
- Most (all?) packages produce robust F-statistics when estimating by 2sls with robust standard errors (heteroskedasticity, serial correlation, clustering).
- While still test for $H_0 : \pi = 0$, robust F does not convey weak instrument information for 2sls.
- Isaiah Andrews (2018) famously finds a design where mean $F_r = 100,000$ and the 2sls confidence set has a 15% size distortion.

Introduction

```
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                4.342
                         (Kleibergen-Paap rk Wald F statistic):          5.021
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias    13.91
                                         10% maximal IV relative bias     9.08
                                         20% maximal IV relative bias     6.46
                                         30% maximal IV relative bias     5.39
                                         10% maximal IV size             22.30
                                         15% maximal IV size             12.83
                                         20% maximal IV size              9.54
                                         25% maximal IV size              7.80
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
```

Introduction

- Montiel-Olea and Pflueger (2013) Effective F-statistic is test for weak instruments related to Nagar bias of 2sls, valid under general forms of nonhomoskedasticity.
- Andrews, Stock and Sun (2019) recommend "that researchers judge instrument strength based on the effective F-statistic of Montiel Olea and Pflueger ".
- This paper: shows that MOP method applies to class of linear GMM estimators, with associated class of generalized effective F-statistics.
- It follows that the robust F-statistic is a test for weak instruments related to the Nagar bias of a GMM estimator.
- I call this the GMMf estimator, as the weight matrix is based on the first-stage residuals.
- So when F_r is large, this estimator performs well, as in the Andrews (2018) example.

Model and Assumptions

The general model specification is given by

$$y = x\beta + u = Z\pi_y + v_1$$
$$x = Z\pi + v_2$$

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

Weak instrument asymptotics: $\pi = \pi_n = \frac{c}{\sqrt{n}}$.

Further, as $n \to \infty$,

$$\frac{1}{n}Z'Z \xrightarrow{p} Q_{zz}$$
$$\frac{1}{n}\left[v_1 \;\; v_2\right]'\left[v_1 \;\; v_2\right] \xrightarrow{p} \Sigma_v = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$
$$\frac{1}{\sqrt{n}}\begin{pmatrix} Z'v_1 \\ Z'v_2 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, W = \begin{bmatrix} W_1 & W_{12} \\ W_{12}' & W_2 \end{bmatrix}\right)$$
$$\widehat{W} \xrightarrow{p} W.$$

For example,

$$\widehat{W}_2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{v}_{2i}^2 z_i z_i'.$$

2sls and F-Statistics

The 2sls estimator is

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

$P_z = Z\left(Z'Z\right)^{-1} Z'$.

Standard F-statistic is

$$\widehat{F} = \frac{x' P_z x}{k_z \widehat{\sigma}_2^2}.$$

The Effective F-statistic is given by,

$$\widehat{F}_{\text{eff}} = \frac{x' P_z x}{\text{tr}\left(\left(\frac{1}{n}Z'Z\right)^{-1/2} \widehat{W}_2 \left(\frac{1}{n}Z'Z\right)^{-1/2}\right)} = \frac{x' P_z x}{\text{tr}\left(\widehat{W}_2 \left(\frac{1}{n}Z'Z\right)^{-1}\right)},$$

and the robust F is

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

Generalized Effective F-statistic

- Montiel Olea and Pflueger (2013) show that the effective F-statistic is a test for weak instruments in terms of Nagar bias of 2sls estimator as a proportion of a benchmark, worst-case bias.
- I show that the their approach applies to the class of one-step linear GMM estimators

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

with $\Omega_n \xrightarrow{p} \Omega$, with the generalized effective F-statistic given by

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

For 2sls, $\Omega_n = \left(\frac{1}{n}Z'Z\right)^{-1}$ and so $\widehat{F}_{\text{geff}}\left(\left(\frac{1}{n}Z'Z\right)^{-1}\right) = \widehat{F}_{\text{eff}}$.

Generalized Effective F-statistic

- Let $\Omega_n = \widehat{W}_2^{-1}$, then

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

  and then

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

- So the robust F-statistic can be used as a test for weak instruments in relation to the Nagar bias of the GMMf estimator.
- "Canonical" relationship, expressions and limiting distribution simplify.

Steps, General

1. Let $\widehat{W}_{\Omega_n} = \left(I_2 \otimes \Omega_n^{1/2}\right) \widehat{W} \left(I_2 \otimes \Omega_n^{1/2}\right)$. Obtain
$$B\left(\widehat{W}_{\Omega_n}\right) = \sup_{\beta \in \mathbb{R}, c_{\Omega,0} \in \mathcal{S}^{k_z - 1}} \left(\frac{\left| n\left(\beta, c_{\Omega,0}, \widehat{W}_{\Omega_n}\right)\right|}{\mathrm{BM}\left(\beta, \widehat{W}_{\Omega_n}\right)} \right)$$
by a numerical maximization routine.

2. Compute $d_{\Omega_n \tau} = B\left(\widehat{W}_{\Omega_n}\right)/\tau$ and the effective degrees of freedom
$$\widehat{k}_{\mathrm{geff}}\left(\Omega_n\right) = \frac{\left[\mathrm{tr}\left(\widehat{W}_{\Omega_n,2}\right)\right]^2 \left(1 + 2d_{\Omega_n,\tau}\right)}{\mathrm{tr}\left(\widehat{W}'_{\Omega_n,2}\widehat{W}_{\Omega_n,2}\right) + 2d_{\Omega_n,\tau}\mathrm{tr}\left(\widehat{W}_{\Omega_n,2}\right)\lambda_{\max}\left(\widehat{W}_{\Omega_n,2}\right)},$$
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
$\chi^2_{\widehat{k}_{\mathrm{geff}}(\Omega_n)}\left(d_{\Omega_n,\tau}\widehat{k}_{\mathrm{geff}}\left(\Omega_n\right)\right)/\widehat{k}_{\mathrm{geff}}\left(\Omega_n\right)$

3. Reject the null of weak instruments, that the proportion of the Nagar bias of $\widehat{\beta}_{\Omega_n}$ relative to the benchmark bias is larger than τ, for at least some value of β and some direction $c_{\Omega,0}$, if $\widehat{F}_{\mathrm{geff}}\left(\Omega_n\right) > cv\left(\alpha, \widehat{W}_{\Omega_n,2}, d_{\Omega_n,\tau}\right)$.

Steps, Robust F and GMMf

1. Obtain

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

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

2. 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 $\chi^2_{k_z}\left(k_z d_{\Omega_n,\tau}\right)/k_z$.

Harmonizing the Benchmark Bias

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

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

This is an appropriate benchmark bias for the class of linear GMM estimators considered as it can be seen to be the worst-case benchmark bias under homoskedasticity. Similar in spirit to the Stock-Yogo relative bias results.

Grouped-Data IV Model

The implication is that if $\widehat{F}_r \gg \widehat{F}_{\text{eff}}$ the GMMf estimator will be better behaved than the 2sls estimator in terms of bias.

This is the case with the Andrews (2018) example, which is a grouped-data IV specification:

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

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

Flexibility by group:

$$\begin{pmatrix} u_i \\ v_{2,i} \end{pmatrix} \sim \left(0, \Sigma_g = \begin{bmatrix} \sigma^2_{u,g} & \sigma_{uv_2,g} \\ \sigma_{uv_2,g} & \sigma^2_{v_2,g} \end{bmatrix}\right).$$

Grouped-Data IV Model

The group-specific IV estimators for β are given by

$$\widehat{\beta}_g = \frac{\overline{y}_g}{\overline{x}_g}.$$

The group-specific F-statistics are

$$\widehat{F}_g = \frac{n_g \overline{x}_g^2}{\widehat{\sigma}^2_{v_2,g}}.$$

A large value of $\widehat{F}_g$ indicates that $\widehat{\beta}_g$ is well behaved.

$\widehat{\beta}_{2sls}$ is a weighted average of the $\widehat{\beta}_g$,

$$\widehat{\beta}_{2sls} = \sum_{g=1}^{G} w_{2sls,g} \widehat{\beta}_g,$$

$$w_{2sls,g} = \frac{n_g \overline{x}_g^2}{\sum_{l=1}^{G} n_l \overline{x}_l^2} = \frac{\widehat{\sigma}^2_{v_2,g} \widehat{F}_g}{\sum_{l=1}^{G} \widehat{\sigma}^2_{v_2,l} \widehat{F}_l},$$

so when $\widehat{\sigma}^2_{v,g}$ is relatively small, an informative group may be given a small weight.

Grouped-Data IV Model

The non-robust F-statistic is given by

$$\widehat{F} = \frac{1}{G}\sum_{g=1}^{G} \frac{\widehat{\sigma}^2_{v_2,g}}{\left(\sum_{l=1}^{G} \frac{n_l}{n}\widehat{\sigma}^2_{v_2,l}\right)}\widehat{F}_g.$$

The Effective F-statistic is given by

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

and so $\widehat{F}_{\text{eff}} = \widehat{F}$ if $n_g = n/G$ for $g = 1, \ldots, G$, (which is the case in expectation below).

The robust F-statistic is given by

$$\widehat{F}_r = \frac{1}{G}\sum_{g=1}^{G} \widehat{F}_g.$$

Grouped-Data IV Model

With equal group sizes and first-stage homoskedasticity, $\sigma^2_{v,g} = \sigma^2_v$ for $g = 1, \ldots, G$, we have that

$$\widehat{F} = \widehat{F}_{\text{eff}} \approx \frac{1}{G} \sum_{g=1}^{G} \widehat{F}_g$$

and

$$w_{2sls,g} \approx \frac{\widehat{F}_g}{G\widehat{F}}.$$

For the GMMf estimator and the robust F-statistic we have that, independent of the values of the variances,

$$\widehat{F}_r = \frac{1}{G} \sum_{g=1}^{G} \widehat{F}_g$$

and

$$\widehat{\beta}_{gmmf} = \sum_{g=1}^{G} w_{gmmf,g} \widehat{\beta}_g; \;\; w_{gmmf,g} = \frac{\widehat{F}_g}{G\widehat{F}_r}.$$

Grouped Data IV Model

Andrews (2018) design, $G = 10$, $\beta = 0$, $n = 10,000$, group sizes 1000 in expectation.

Table 1: Estimation results

	$\widehat{F}$	$\widehat{F}_{\text{eff}}$	$\widehat{F}_r$	$\widehat{\beta}_{ols}$	$\widehat{\beta}_{2sls}$	$\widehat{\beta}_{gmmf}$	$Wald_{2sls}$	$Wald_{gmmf}$
ME	1.411	1.411	80.23	-0.608	-0.424	-0.001	0.534	0.049
		[17.09, 0; 17.09, 0]	[12.22, 1; 13.45, 1]	(0.011)	(0.257)	(0.563)		
HE	0.993	0.993	80.12	0.747	0.742	0.007	0.999	0.065
		[17.12, 0; 17.12, 0]	[12.31, 1; 12.26, 0]	(0.001)	(0.057)	(0.029)		

Notes: $\beta = 0$. Means and (st.dev.), [mean of critical values, rej.freq., $\alpha = 0.05$, $\tau = 0.10$, LS benchmark in second row], of 10,000 replications. Rej.freq. of robust Wald tests at 5% level.

Grouped Data IV Model

Table 2: Group information and estimator weights, ME

s	1	2	3	4	5	6	7	8	9	10
π_g	0.058	-0.023	0.049	0.015	0.022	0.008	-0.017	0.011	-0.036	-0.040
$\sigma^2_{v,g}$	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
$w_{2sls,g}$	0.126	0.098	0.178	0.035	0.031	0.180	0.049	0.015	0.096	0.192
$w_{gmmf,g}$	0.984	0.002	0.002	0.002	0.003	0.001	0.002	0.002	0.002	0.002

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

Grouped Data IV Model

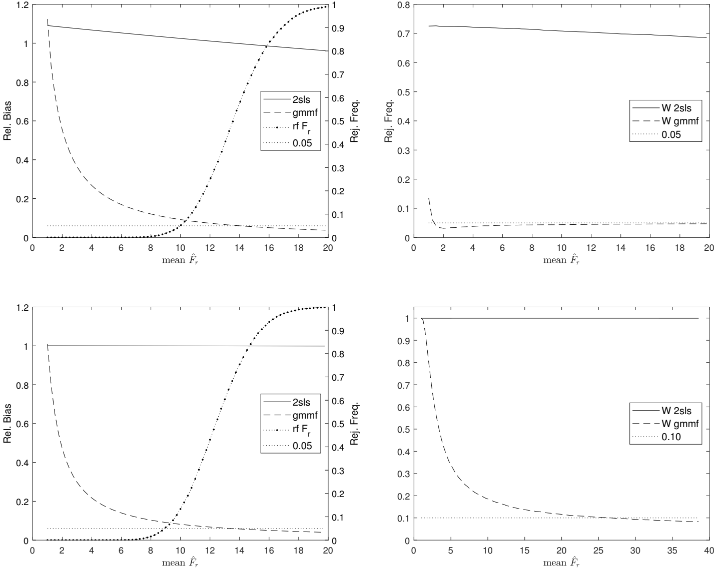

Figure 1: Top: Moderate Endogeneity. Bottom: High Endogeneity. Left: Bias of 2SLS and GMMf estimators relative to OLS bias, and $\widehat{F}_r$-based weak-instrument test rejection frequencies, $\alpha = 0.05$, $\tau = 0.10$, least-squares benchmark bias. The “0.05” line refers to the rej. freq. Right: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

In Practice

Compute both $\widehat{F}_r$ and $\widehat{F}_{\text{eff}}$ and their critical values. Large differences favour one estimator over the other, $\widehat{\beta}_{2sls}$ vs $\widehat{\beta}_{gmmf}$.

I have adapted "weakivtest" in Stata to do this.

One of the AER examples in Andrews et al. (2019) is Stevens and Yang (2014), "Compulsory Education and the Benefits of Schooling", estimates of return of number of years of education on log weekly wages, using three indicators RS7, RS8 and RS9 for being required to attend 7, 8 and nine or more years of schooling.

Estimated by 2sls, with reported cluster-robust first-stage F-statistics (state of birth/year of birth).

In Practice

Table 3: The Effect of Schooling on log Weekly Wages

	2sls	GMMf	$\widehat{F}$	$\widehat{F}_{\text{eff}}$	$\widehat{F}_r$	cv_{eff}^{LS}	cv_r^{LS}
White males 40-49	0.095 (0.016)	0.095 (0.016)	108.6	42.85	42.75	9.21	8.64
+region*yob	−0.020 (0.041)	−0.014 (0.040)	16.06	8.11	8.22	10.31	8.73
White males 25-54	0.0973 (0.0096)	0.0998 (0.0095)	548.0	64.17	81.37	13.50	8.74
+region*yob	−0.014 (0.021)	−0.012 (0.021)	89.06	24.38	23.63	10.36	8.74
All whites 20–54	0.105 (0.011)	0.111 (0.011)	870.0	62.98	91.73	14.65	9.62
+region*yob	−0.0025 (0.016)	0.0035 (0.016)	197.0	42.40	40.57	11.10	8.65
Whites 25-54, Non-South	−0.0086 (0.012)	0.015 (0.011)	603.6	34.40	67.25	16.30	8.96
South	0.019 (0.043)	0.022 (0.044)	20.89	6.13	6.34	9.49	8.82

Dynamic Panel Data Model

Consider the dynamic AR(1) panel data specification

$$y_{it} = \gamma y_{i,t-1} + \eta_i + u_{it}.$$

for $i = 1, \ldots, n$ and $t = 1, \ldots T$.

Generalized effective F-statistic as test for weak instruments applies directly to Arellano-Bond first-differenced model one-step GMM estimator,

$$\widehat{\gamma}_{FD} = \frac{\Delta y'_{-1} Z \Omega_n^{FD} Z' \Delta y}{\Delta y'_{-1} Z \Omega_n^{FD} Z' \Delta y_{-1}},$$

$$\Omega_n^{FD} = \left(\frac{1}{n} \sum_{i=1}^{n} Z_i' D Z_i \right)^{-1},$$

with instruments Z_i the matrix of sequential lagged levels.

Then $\Delta y_{i,-1} = Z_i \pi + \xi_i$, $\widehat{\xi}_i = \Delta y_{i,-1} - Z_i \widehat{\pi}$, $\widehat{W}_2 = \frac{1}{n} \sum_{i=1}^{n} Z_i' \widehat{\xi}_i \widehat{\xi}_i' Z_i$, and

$$\widehat{F}_{\text{geff}} = \frac{\Delta y'_{-1} Z \Omega_n^{FD} Z' \Delta y_{-1}}{n \text{tr} \left(\widehat{W}_2 \Omega_n^{FD} \right)}.$$

Conclusions

- Montiel Olea and Pflueger methods apply to general class of linear GMM estimators with associated class of generalized effective F-statistics.
- Robust F-statistic informative about Nagar bias of new GMMf estimator.
- When estimating by 2sls, compute both effective F and robust F and their weak-instrument critical values, and also the GMMf estimator (“gfweakivtest” for Stata).
- To do: Extend to behaviour of Wald test.
- To do: Extend to multiple endogenous variables. Lewis and Mertens (2022) have done so for the 2sls effective F.