UvA ECONOMETRICS

Discussion Paper: 2010/02

Weak instruments and the first stage F-statistic in IV models with a nonscalar error covariance structure

Maurice Bun and Monique de Haan

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Weak instruments and the first stage F-statistic in IV models with a nonscalar error covariance structure

MAURICE J.G. BUN*AND MONIQUE DE HAAN†

this version: 26 January 2010

JEL-code: C13; C23

Keywords: first stage F-statistic, grouped data, instrumental variables, relative bias, serial correlation, weak instruments

Abstract

We analyze the usefulness of the first stage F-statistic for detecting weak instruments in the IV model with a nonscalar error covariance structure. More in particular, we question the validity of the rule of thumb of a first stage F-statistic of 10 or higher for models with correlated errors arising from either a group structure or serial correlation. Using asymptotic expansion techniques we derive bias approximations for IV and OLS estimators in this generalized IV model. We relate these bias approximations to expected values of both standard and robust versions of the first stage F-statistic. Our theoretical and simulation results indicate that the standard first stage F-statistic overestimates the strength of instruments. In addition, there does not seem to be a close correspondence between the robust version of the F-statistic and weak instruments as measured by relative bias of IV with respect to OLS.

*Corresponding author. Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, email: M.J.G.Bun@uva.nl

†Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, email: M.deHaan1@uva.nl.

1. Introduction

Well known in the literature on instrumental variables (IV) estimation is the issue of weak instruments. When instruments are weak, i.e. only weakly correlated with the endogenous regressors, the IV estimator can perform poorly in finite samples, see e.g. Bound et al. (1995), Staiger and Stock (1997) and Stock et al. (2002). With weak instruments, the IV estimator is biased in the direction of the ordinary least squares (OLS) estimator, and its distribution is non-normal which affects inference using Wald testing procedures.

Bound et al. (1995) and Staiger and Stock (1997) advocate use of the first-stage F-statistic to investigate the strength of the instruments. It has been shown that there is a close correspondence between the expected value of the first-stage F-statistic and bias of the IV estimator, relative to the bias of the OLS estimator. Using weak instrument asymptotics, Stock and Yogo (2005) tabulate critical values for the first-stage F-statistic to test whether instruments are weak. They do this separately for the maximum bias of the IV estimator, relative to the bias of the OLS estimator, and for the maximum Wald test size distortion. They conclude that the first stage F statistic should be large, typically exceeding the rule of thumb of 10 proposed by Staiger and Stock (1997).

Applied researchers routinely report OLS estimation results of the first stage as well as the value of the first stage F-statistic to assess the relevance of instruments. For example, Bound et al. (1995) reexamine Angrist and Krueger's (1991) IV estimates of the returns to schooling and report first stage F-statistics for some selected specifications of Angrist and Krueger (1991), who used quarter of birth as instrument for education in wage equations. For most specifications Bound et al. (1995) report rather low values for the first stage F-statistic indicating that instruments based on quarter of birth are weak.

The above mentioned theoretical results on the relation between first stage F-statistic and relative bias/size distortion have been established for models with i.i.d. errors. In applied research, however, typically error terms are allowed to have heteroskedasticity and/or serial correlation. Hence, quite often robust versions of the first stage F-statistic have been reported instead of the standard version. For example, in contrast with the pioneering work of Angrist and Krueger (1991) nowadays typically clustered standard errors are reported when analyzing the returns to schooling. Examples are Oreopoulos (2006) and Pischke and von Wachter (2008), who both analyze the returns to schooling using school leaving age as instrument. The exogenous variation in the instrument typically arises from a change in compulsory schooling laws. Individual data are available and additional exogenous regressors are individual characteristics like age, gender, year of birth, state of residence, etc. Pischke and von Wachter (2008) use individual data and report first stage regression results with standard errors robust to clusters at the state/year of birth level. Oreopoulos (2006) first aggregates the data into cell means at the year of birth, nation, race, sex and survey year level. Next, first stage regressions have been reported with standard errors robust for clusters at the region/birth cohort level.

Several weak instrument robust inference procedures for testing hypotheses about structural parameters have been developed for the IV model with a nonscalar error covariance matrix (Kleibergen, 2007). However, not much is known about the usefulness of the first stage F-statistic for detecting weak instruments in case of IV models with a more general error covariance structure. In this study we will provide both theoretical and simulation results on this relationship. We will make use of asymptotic expansion techniques to approximate the IV and OLS biases. Next, for two specific error covariance structures, i.e. grouped and autoregressive errors, we will obtain analytical formulae for the relative bias and the expected value of both standard and robust versions of the first stage F-statistic.

For the IV model with i.i.d errors Staiger and Stock (1997) propose as rule of thumb for the first stage F-statistic the critical value of 10. Stock and Yogo (2005) conclude that the first stage F-statistic should be large, typically exceeding the rule of thumb of 10. In this paper we want to analyze for the in case of a nonscalar error covariance structure whether 10 is still a sensible critical value for the first stage F-statistic. In addition, we want to verify whether reporting a first stage F-statistic robust for heteroskedasticity and serial correlation makes sense in terms of relative bias. The results in this study are thus complementary to the generalization of weak instruments robust inference methods to IV models with a nonscalar covariance structure (Kleibergen, 2007).

The setup of the paper is as follows. Section 2 discusses relative bias and first stage F-statistic in the IV model with i.i.d. errors. Section 3 introduces the IV model with nonscalar error covariance structure and standard/robust versions of the first stage F-statistic. Also general formulae for approximating IV and OLS biases are derived. Section 4 reports specific analytical results for the case of grouped errors when analyzing cross-section data, while Section 5 deals with the case of serial correlation in time series. Section 6 presents Monte Carlo results for both applications, while Section 7 concludes.

2. Relative bias and first stage F statistic

Consider the simple linear IV regression model with one endogenous regressor x and k_z instruments z

$$\begin{aligned} y_i &= x_i\beta + u_i \\ x_i &= z_i'\pi + v_i, \end{aligned} \tag{2.1}$$

for $i = 1, ..., n$ where the (u_i, v_i) are independent draws from a bivariate normal distribution with zero means, variances σ_u^2 and σ_v^2, and correlation coefficient ρ_{uv}. Stacking the observations we have

$$y = x\beta + u \tag{2.2}$$

$$x = Z\pi + v, \tag{2.3}$$

where y, x and u are the n-vectors $(y_1, ..., y_n)'$, $(x_1, ..., x_n)'$ and $(u_1, ..., u_n)'$ respectively, and Z is the $n \times k_z$ matrix $(z_1, ..., z_n)'$. The parameter β is estimated either by OLS or 2SLS:

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

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

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

It is well known that when instruments are weak, i.e. when they are only weakly correlated with the endogenous regressor, the 2SLS estimator can perform poorly in finite samples, see e.g. Bound et al. (1995) or Staiger and Stock (1997). With weak instruments, the 2SLS estimator is biased in the direction of the OLS estimator, and its distribution is non-normal which affects inference using the Wald testing procedure.

A measure of the strength of the instruments is the concentration parameter, which is defined as

$$\mu = \frac{\pi'Z'Z\pi}{\sigma_v^2}. \tag{2.4}$$

When it is evaluated at the OLS, first stage, estimated parameters

$$\widehat{\mu} = \frac{\widehat{\pi}'Z'Z\widehat{\pi}}{\widehat{\sigma}_v^2},$$

it is clear that $\widehat{\mu}$ is equal to the Wald test for testing the hypothesis $H_0 : \pi = 0$, and $\widehat{\mu}/k_z$ equals the F-test statistic defined as

$$F = \frac{\widehat{\pi}' Z' Z \widehat{\pi} / k_z}{\widehat{\sigma}_v^2}. \tag{2.5}$$

Bound et al. (1995) and Staiger and Stock (1997) advocate use of the first-stage F-statistic to investigate the strength of the instruments. The first stage F statistic has a noncentral F distribution with noncentrality parameter equal to μ. Hence, we have the following approximate relation (Stock et al., 2002) between the expected value of the first stage F statistic and the concentration parameter:

$$E[F] \approx 1 + \frac{\mu}{k_z}. \tag{2.6}$$

From this result it is seen that $F - 1$ can be interpreted as an estimator of $\frac{\mu}{k_z}$.

The concentration parameter μ is a key quantity in describing the finite sample properties of the IV estimator. The approximate bias of the 2SLS estimator can be obtained using higher order asymptotics based on the expansion of the 2SLS estimation error, see Nagar (1959), Buse (1992) and Hahn and Kuersteiner (2002). It follows that the approximate bias of the IV estimator can be expressed as

$$E\left[\widehat{\beta}_{2SLS} - \beta\right] \approx \frac{(k_z - 2)\,\sigma_{uv}}{\pi' Z' Z \pi} = \frac{\sigma_{uv}}{\sigma_v^2} \frac{(k_z - 2)}{\mu}. \tag{2.7}$$

Hence the bias is inversely proportional to the value of the concentration parameter. It does not only depend on the concentration parameter, but also on the number of instruments k_z and the degree of endogeneity embodied in the covariance σ_{uv}. However, the relevance of the concentration parameter for finite sample bias becomes even more pronounced when we consider the bias of the IV estimator, relative to that of the OLS estimator as defined by

$$\text{RelBias} = \frac{E\left[\widehat{\beta}_{2SLS} - \beta\right]}{E\left[\widehat{\beta}_{OLS} - \beta\right]},$$

see e.g. Bound et al. (1995). The bias of the OLS estimator can be approximated by (see e.g. Hahn and Kuersteiner (2002))

$$E\left[\widehat{\beta}_{OLS} - \beta\right] \approx \frac{\sigma_{uv}}{\sigma_v^2} \frac{1}{\frac{\mu}{n} + 1},$$

which is equal to inconsistency of OLS. The relative bias is then approximately given by

$$\text{RelBias} \approx \frac{(k_z - 2)\left(\frac{\mu}{n} + 1\right)}{\mu}, \tag{2.8}$$

i.e. a function of μ, n, and k_z only. Combining now the approximate results in (2.6) and (2.8) we have for moderately large μ that relative bias roughly equals $\frac{1}{F}$. Finally, the concentration parameter is an important element in describing size distortions of t or Wald tests based on the 2SLS estimator, see the discussion in Bun and Windmeijer (2010).

The results discussed above are based on conventional higher-order asymptotics, i.e. assuming strong identification. Hence, these higher-order approximations may not always be informative in case of weak instruments. However, regarding the relevance of the concentration parameter, weak instrument asymptotics as derived by Staiger and Stock (1997) lead to similar conclusions compared with conventional fixed-parameter higher-order asymptotics. Staiger and Stock (1997) develop weak instrument asymptotics by setting $\pi = \pi_n = C/\sqrt{n}$, in which case the concentration parameter converges to a constant. They then show that 2SLS is not consistent and has a nonstandard asymptotic distribution. These results are of course different from conventional asymptotics. However, Staiger and Stock (1997) show that the asymptotic bias of the 2SLS estimator, relative to that of the OLS estimator again only depends on k_z and μ. Furthermore, the distributions of the 2SLS t-ratio and Wald statistic only depend on μ, k_z and ρ_{uv}.

3. IV model with a nonscalar covariance structure

All results in the previous section have been established under the assumption that the error terms in both structural and reduced form equations are i.i.d. We now consider the case where they have a nonscalar covariance matrix structure, i.e.

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim N(0, \Omega). \tag{3.1}$$

We can further decompose Ω as follows

$$\Omega = \begin{pmatrix} \Omega_u & \Omega_{uv} \\ \Omega_{uv} & \Omega_v \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \Delta_u & \sigma_{uv}\Delta_{uv} \\ \sigma_{uv}\Delta_{uv} & \sigma_v^2 \Delta_v \end{pmatrix}. \tag{3.2}$$

The case of i.i.d. disturbances is a special case where $\Delta_u = \Delta_v = \Delta_{uv} = I_n$ and, hence, the covariance matrix attains a Kronecker product form (see e.g.

Kleibergen, 2007):

$$\Omega = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \otimes I_n.$$

Using the reduced form specification (2.3) we can write for the estimation errors

$$\widehat{\beta}_{OLS} - \beta = \frac{\pi' Z' u + v' u}{\pi' Z' Z \pi + 2\pi' Z' v + v' v}, \tag{3.3}$$

$$\widehat{\beta}_{2SLS} - \beta = \frac{\pi' Z' u + v' P_Z u}{\pi' Z' Z \pi + 2\pi' Z' v + v' P_Z v}. \tag{3.4}$$

The approximate bias of the OLS and 2SLS estimators can be obtained using higher order asymptotics, see e.g. Nagar (1959). We will analyze such approximations assuming the nonscalar covariance structure (3.2) and that we have nonstochastic instruments such that $\lim_{n\to\infty} \frac{1}{n} Z'Z$ is finite and nonsingular.

The bias (inconsistency) of the OLS estimator can be approximated by

$$E\left[\widehat{\beta}_{OLS} - \beta\right] \approx \frac{E\left[v'u\right]}{\pi' Q_{ZZ} \pi + E\left[v'v\right]} = \frac{\sigma_{uv} n^{-1} tr(\Delta_{uv})}{\pi' Q_{ZZ} \pi + \sigma_v^2 n^{-1} tr(\Delta_v)}, \tag{3.5}$$

where $Q_{ZZ} = \frac{1}{n} Z'Z$. It is seen that the inconsistency does really only depend explicitly on the contemporaneous correlation between the structural and reduced error terms.

In the Appendix it is shown that we have the following 2SLS bias approximation:

$$E\left[\widehat{\beta}_{2SLS} - \beta\right] = \frac{1}{n}\left(\frac{tr(Q_{ZZ}^{-1} Q_{ZZ}^{*})}{\pi' Q_{ZZ} \pi} - \frac{2\pi' Q_{ZZ}^{*} \pi}{\left(\pi' Q_{ZZ} \pi\right)^2}\right) + o(n^{-1}), \tag{3.6}$$

where $Q_{ZZ}^{*} = \frac{1}{n} Z' \Omega_{uv} Z$. Below we will apply these results to some well known specific covariance structures to assess relative bias.

Next consider first stage $Wald$ statistics for testing $H_0 : \pi = 0$ of the form

$$\hat{\pi}' \hat{V}_{\hat{\pi}}^{-1} \hat{\pi}.$$

F-statistics result when dividing by k_z. When we use

$$\hat{V}_{\hat{\pi}} = \hat{\sigma}_v^2 (Z'Z)^{-1}, \tag{3.7}$$

we have the standard F-statistic, while using

$$\hat{V}_{\hat{\pi}} = (Z'Z)^{-1}\left(Z' \hat{\Omega}_v Z\right)(Z'Z)^{-1}, \tag{3.8}$$

a robust F stastictic results taking into account the nonscalar covariance structure in the reduced form. In some applications (e.g. using grouped data) researchers typically report this robust F-statistic rather the standard version. Below we will analyze below what the consequences are of this choice for population versions, i.e. assuming the covariance structure known.

4. grouped errors

Suppose we have cross-section data with a grouped error structure in both the structural and reduced form. Denoting the individual observation i now with two indices, i.e. g and m, we have for model (2.1)

$$\begin{aligned} y_{gm} &= x_{gm}\beta + u_{gm} \\ x_{gm} &= z'_{gm}\pi + v_{gm}, \end{aligned} \tag{4.1}$$

with $g = 1, ..., G$ and $m = 1, ..., M_g$. For ease of exposition we assume equal group size, hence $M_g = M$ and we have in total $n = G * M$ observations. Also we assume

$$z_{gm} = z_g,$$

hence the value of the instruments is equal within groups (see e.g. Shore-Sheppard, 1996). In this case we have $Z = Z^* \otimes \iota_M$ with $Z^* = (z_1, ..., z_G)'$ and we have

$$Q_{ZZ} = \frac{1}{g} Z^{*\prime} Z^*.$$

Regarding the errors we assume an additive error components structure

$$u_{gm} = \eta_g + \varepsilon_{gm} \tag{4.2}$$

$$v_{gm} = \alpha_g + \xi_{gm}. \tag{4.3}$$

We further assume homoskedasticity and that all within group correlation is modeled by the group effects, hence defining $\sigma_{uv} = \sigma_{\eta\alpha} + \sigma_{\varepsilon\xi}$ and $\sigma_v^2 = \sigma_\alpha^2 + \sigma_\xi^2$ we have

$$\begin{aligned} \Omega_{uv} &= I_G \otimes (\sigma_{\eta\alpha}\iota_M\iota'_M + \sigma_{\varepsilon\xi} I_M) \\ &= \sigma_{uv}\left(I_G \otimes \left(I_M + \frac{\sigma_{\eta\alpha}}{\sigma_{uv}}(\iota_M\iota'_M - I_M)\right)\right), \\ &= \sigma_{uv}\Delta_{uv} \end{aligned}$$

$$\begin{aligned}
\Omega_{vv} &= I_G \otimes \left(\sigma_\alpha^2 \iota_M \iota_M' + \sigma_\xi^2 I_M\right) \\
&= \sigma_v^2 \left(I_G \otimes \left(I_M + \frac{\sigma_\alpha^2}{\sigma_v^2}(\iota_M \iota_M' - I_M)\right)\right) \\
&= \sigma_v^2 \Delta_v .
\end{aligned}$$

Because $tr(\Delta_{uv}) = tr(\Delta_v) = n$ according to (3.5) the bias of the OLS estimator can be approximated by

$$E\left[\widehat{\beta}_{OLS} - \beta\right] \approx \frac{\sigma_{uv}}{\pi' Q_{ZZ}\pi + \sigma_v^2}.$$

Regarding the approximate bias of the 2SLS estimator as described in (3.6) we have

$$\frac{1}{n} Z'\Omega_{uv} Z = (M\sigma_{\eta\alpha} + \sigma_{\varepsilon\xi}) \frac{1}{g} Z^{*\prime} Z^{*},$$

and

$$M\sigma_{\eta\alpha} + \sigma_{\varepsilon\xi} = \sigma_{uv}\left(1 + \frac{(M-1)\sigma_{\eta\alpha}}{\sigma_{uv}}\right).$$

Hence, we have

$$Q_{ZZ}^{*} = \sigma_{uv}\left(1 + \frac{(M-1)\sigma_{\eta\alpha}}{\sigma_{uv}}\right) Q_{ZZ},$$

and

$$E\left[\widehat{\beta}_{2SLS} - \beta\right] = \left(1 + \frac{(M-1)\sigma_{\eta\alpha}}{\sigma_{uv}}\right) \frac{1}{n} \frac{(k_z - 2)\sigma_{uv}}{\pi' Q_{ZZ}\pi} + o(n^{-1}).$$

Now using the definition of μ as in (2.4) the relative bias is approximately given by

$$\text{Relbias} = \frac{E\left[\widehat{\beta}_{2SLS}\right] - \beta}{E\left[\widehat{\beta}_{OLS}\right] - \beta} \approx \left(1 + \frac{(M-1)\sigma_{\eta\alpha}}{\sigma_{uv}}\right) \frac{(k_z - 2)\left(\frac{1}{n}\mu + 1\right)}{\mu}. \tag{4.4}$$

In the i.i.d. case relative bias is a function of μ, n, and k_z only. However, when errors in both structural and reduced form error exhibit within groups correlation and group effects are positively correlated ($\sigma_{\eta\alpha} > 0$) relative bias is always larger keeping the value of the concentration parameter fixed.

Next we consider first stage F-statistics. In this application typically the cluster robust F-statistic (see e.g. Arellano, 1987) using

$$\hat{V}_{\hat{\pi}} = (Z'Z)^{-1}\left(\sum_g Z_g' \hat{v}_g \hat{v}_g' Z_g\right)(Z'Z)^{-1},$$

is reported instead of the standard F-statistic. For the population standard F we have

$$\begin{aligned} E\left[F^{pop}\right] &= \frac{1}{k_z\sigma_v^2}E\left[x'P_Zx\right] \\ &= \frac{1}{k_z\sigma_v^2}\left(\pi'Z'Z\pi + k_z\sigma_v^2\left(1+\frac{(M-1)\sigma_\alpha^2}{\sigma_v^2}\right)\right) \\ &= \frac{\mu}{k_z}+1+\frac{(M-1)\sigma_\alpha^2}{\sigma_v^2}, \end{aligned}$$

hence compared with the i.i.d. case it is overestimating the strength of the instruments for fixed concentration parameter. Regarding the population robust F we have

$$\begin{aligned} E\left[F_r^{pop}\right] &= \frac{1}{k_z}E\left[x'Z\left(Z'\Omega_{vv}Z\right)^{-1}Z'x\right] \\ &= \frac{1}{k_z}\left(\pi'Z'Z\left(Z'\Omega_{vv}Z\right)^{-1}Z'Z\pi + tr((Z'\Omega_{vv}Z)^{-1}Z'E\left[vv'\right]Z)\right) \\ &= \frac{1}{k_z}\left(\frac{\pi'Z'Z\pi}{M\sigma_\alpha^2+\sigma_\xi^2}+k_z\right) \\ &= \frac{1}{\left(1+\frac{(M-1)\sigma_\alpha^2}{\sigma_v^2}\right)}\frac{\mu}{k_z}+1, \end{aligned}$$

where

$$\frac{1}{n}Z'\Omega_{vv}Z = (M\sigma_\alpha^2+\sigma_\xi^2)\frac{1}{g}Z^{*\prime}Z^*,$$

and

$$M\sigma_\alpha^2+\sigma_\xi^2 = \sigma_v^2\left(1+\frac{(M-1)\sigma_\alpha^2}{\sigma_v^2}\right).$$

Hence, compared with the i.i.d. case it is not clear whether it is over- or underestimating the strength of the instruments for a fixed concentration parameter.

5. serial correlation

Suppose we have stationary time-series data with serial correlation in both structural and reduced form errors. Denoting the individual observation i now with a time index t, we have for model (2.1)

$$\begin{aligned} y_t &= x_t\beta + u_t \\ x_t &= z_t'\pi + v_t. \end{aligned} \tag{5.1}$$

where $t = 1, ..., n$. We assume that the instruments follow the following first-order autoregressive process

$$z_t = \rho_z I_{k_z} z_{t-1} + \eta_t,$$
$$\eta_t \sim IIN\left(0, \sigma_\eta^2 I_{k_z}\right).$$

Hence, we have

$$Q_{ZZ} = E\left[z_t z_t'\right] = \sigma_z^2 I_{k_z}$$

Furthermore, we assume that both errors follow a first-order autoregressive process

$$\begin{aligned} u_t &= \rho_u u_{t-1} + \varepsilon_t \\ v_t &= \rho_v v_{t-1} + \xi_t, \end{aligned}$$

where (ε_t, ξ_t) are independent draws from a bivariate normal distribution with zero means, variances σ_ε^2 and σ_ξ^2, and correlation coefficient $\rho_{\varepsilon\xi}$. Hence, defining $\sigma_{uv} = \frac{\sigma_{\varepsilon\xi}}{1-(\rho_u\rho_v)^2}$ and $\sigma_v^2 = \frac{\sigma_\xi^2}{1-\rho_v^2}$ we have

$$\Omega_{uv} = \sigma_{uv} \begin{bmatrix} 1 & \rho_v & \rho_v^2 & & \rho_v^{n-1} \\ \rho_u & 1 & \rho_v & & \\ \rho_u^2 & \rho_u & 1 & & \\ & & & & \rho_v \\ \rho_u^{n-1} & & & \rho_u & 1 \end{bmatrix},$$

$$\Omega_v = \sigma_v^2 \begin{bmatrix} 1 & \rho_v & \rho_v^2 & & \rho_v^{n-1} \\ \rho_v & 1 & \rho_v & & \\ \rho_v^2 & \rho_v & 1 & & \\ & & & & \rho_v \\ \rho_v^{n-1} & & & \rho_v & 1 \end{bmatrix}.$$

The bias of the OLS estimator can be approximated by

$$E\left[\widehat{\beta}_{OLS} - \beta\right] \approx \frac{\sigma_{uv}}{\pi' Q_{ZZ}\pi + \sigma_v^2}.$$

Regarding the approximate bias of the 2SLS estimator as described in (3.6) we have

$$Q_{ZZ}^{*} = \sigma_{uv}\left(\frac{1}{1-(\rho_v\rho_z)^2}+\frac{1}{1-(\rho_u\rho_z)^2}-1\right)Q_{ZZ}+o(1).$$

Substituting this result into (3.6) we get

$$E\left[\widehat{\beta}_{2SLS}-\beta\right] = \left(\frac{1}{1-(\rho_v\rho_z)^2}+\frac{1}{1-(\rho_u\rho_z)^2}-1\right)\frac{(k_z-2)\sigma_{uv}}{\pi' Z' Z\pi}+o(n^{-1}).$$

Now using the definition of μ as in (2.4) the relative bias is approximately given by

$$\text{Relbias} \approx \left(\frac{1}{1-(\rho_v\rho_z)^2}+\frac{1}{1-(\rho_u\rho_z)^2}-1\right)\frac{(k_z-2)\left(\frac{1}{n}\mu+1\right)}{\mu}. \tag{5.2}$$

In the i.i.d. case relative bias is a function of μ, n, and k_z only. However, when both instruments, structural errors and/or reduced form errors exhibit serial correlation relative bias is always larger keeping the value of the concentration parameter fixed.

For the population standard F we have

$$\begin{aligned}
E\left[F^{pop}\right] &= \frac{1}{k_z\sigma_v^2}E\left[x' P_Z x\right] \\
&= \frac{1}{k_z\sigma_v^2}\left(\pi' Z' Z\pi + k_z\sigma_v^2\left(\frac{2}{1-(\rho_v\rho_z)^2}-1\right)\right) \\
&= \frac{\mu}{k_z}+1+2\left(\frac{(\rho_v\rho_z)^2}{1-(\rho_v\rho_z)^2}\right),
\end{aligned}$$

hence compared with the i.i.d. case it is overestimating the strength of the instruments for fixed concentration parameter. Regarding the population robust F we have

$$\begin{aligned}
E\left[F_r^{pop}\right] &= \frac{1}{k_z}E\left[x' Z\left(Z'\Omega_v Z\right)^{-1}Z' x\right] \\
&= \frac{1}{k_z}\left(\pi' Z' Z\left(Z'\Omega_v Z\right)^{-1}Z' Z\pi + tr((Z'\Omega_v Z)^{-1}Z' E\left[vv'\right]Z)\right) \\
&= \frac{1}{k_z}\left(\frac{1-(\rho_v\rho_z)^2}{1+(\rho_v\rho_z)^2}\frac{\pi' Z' Z\pi}{\sigma_v^2}+k_z\right) \\
&= \frac{1-(\rho_v\rho_z)^2}{1+(\rho_v\rho_z)^2}\frac{\mu}{k_z}+1,
\end{aligned}$$

where

$$\frac{1}{n}Z'\Omega_v Z = \sigma_v^2 \frac{1+(\rho_v\rho_z)^2}{1-(\rho_v\rho_z)^2} Q_{ZZ} + o(1).$$

Hence, compared with the i.i.d. case it is unclear whether it is over- or underestimating the strength of the instruments for fixed concentration parameter.

6. Monte Carlo results

In this section we report results from a limited simulation study. We generate data from IV models with either a grouped error structure or autoregressive errors. In order to assess our theoretical findings we have chosen for two different configurations. First, we fix the concentration parameter μ as defined in (2.4) across experiments. We set $\mu/k_z = 9$ which in the i.i.d. case corresponds to a population first stage F-statistic of 10 and relative bias of around 10%. This rule of thumb has been proposed by Staiger and Stock (1997). Second, we fix relative bias at 10% using either (4.4) or (5.2). In this set up we can analyze what kind of values for F-statistics are needed in case of correlated errors.

6.1. grouped data

We generate data voor y and x according to (4.1). Instruments and error terms are drawn according to $z_g \sim IIN(0, I_{k_z})$, $(\eta_g, \alpha_g) \sim IIN(0, \Sigma_{\eta\alpha})$ and $(\varepsilon_{gm}, \xi_{gm}) \sim IIN(0, \Sigma_{\varepsilon\xi})$. We take $\Sigma_{\eta\alpha} = \Sigma_{\varepsilon\xi} = \Sigma$ with

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and $\rho = 0.5$. Hence, we have

$$\begin{aligned} \sigma_\eta^2 &= \sigma_\alpha^2 = \sigma_\varepsilon^2 = \sigma_\xi^2 = 1, \\ \sigma_{\eta\alpha} &= \sigma_{\varepsilon\xi} = 0.5. \end{aligned}$$

We set $n = 500$ while varying the number of groups and choosing $M = n/G$. Choosing $M = 1$ corresponds to the i.i.d. case, while $M > 1$ introduces a grouped error structure. In the latter case we have chosen $M = \{5, 10, 20\}$.

In Table 1 we report simulation results using 2000 Monte Carlo replications. The results in the upper panel follow from fixing the concentration parameter, i.e. we set $\mu/k_z = 9$. This implies for each element in the vector of reduced form coefficients that

$$\pi_j = \sqrt{\frac{\sigma_\alpha^2 + \sigma_\xi^2}{n} \frac{\mu}{k_z}}.$$

The results in the lower panel follow from fixing relative bias as defined in (4.4), i.e. we set $Relbias = 0.1$. Using (4.4) we determine the resulting values for μ depending on the other parameter settings.

Table 1: bias and F statistic in case of grouped errors

M	bias OLS	bias IV	relbias	F	F_r
			$\mu/k_z = 9$		
1	0.459	0.037	0.080	10.04	10.45
5	0.460	0.096	0.210	12.12	4.86
10	0.459	0.162	0.353	15.15	3.90
20	0.460	0.266	0.579	22.72	4.52
			$Relbias = 0.1$		
1	0.470	0.051	0.108	7.42	7.72
5	0.412	0.041	0.099	25.23	10.11
10	0.338	0.035	0.105	57.01	14.71
20	0.191	0.022	0.117	203.37	40.49

Note: average numbers over 2000 replications

Regarding the robust F-statistic we used the robust covariance matrix estimator of Arellano (1987). From Table 1 we see that the pattern of the simulation results follows the theoretical predictions We have that in the i.i.d. case ($M = 1$) results are similar to existing simulation results, i.e. F-statistics of 10 correspond roughly with 10% relative bias. But more importantly for grouped errors ($M > 1$) and for fixed μ we have that relative bias and standard F-statistic increase rapidly. Hence, the standard F-statistic is overestimating relative bias and, hence, strength of instruments. The robust F-statistic decreases considerably, but from the lower panel results it becomes clear that this decrease is not enough to offset the increased relative bias. Fixing relative bias at 10% and allowing for grouped errors leads to much larger F-statistics than in the i.i.d. case. For example, with $M = 20$ observations per group 10% relative bias corresponds to a value of around 200 for the standard F-statistic. Although using a robust F-statistic is a step in the right direction, we still need a value of 40, i.e. considerably larger than 10, to be sure that we have limited relative bias.

6.2. serial correlation

We generate data voor y_t, x_t and z_t according to:

$$\begin{aligned} y_t &= x_t\beta + u_{1t} \\ x_t &= \rho_x x_{t-1} + u_{2t} \\ z_t &= \rho_z I_k z_{t-1} + u_{3t}, \end{aligned}$$

where

$$\begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix} \sim IIN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0' \\ \sigma_{12} & \sigma_{22} & \sigma_{23}\iota_{k_z}' \\ 0 & \sigma_{23}\iota_{k_z} & \sigma_{33}I_{k_z} \end{bmatrix} \right).$$

Exploiting $u_{2t} = \frac{\sigma_{23}}{\sigma_{33}}u_{3t} + \eta_t$ the reduced form for x_t is

$$x_t = \rho_x x_{t-1} + \frac{\sigma_{23}}{\sigma_{33}}z_t - \rho_z\frac{\sigma_{23}}{\sigma_{33}}z_{t-1} + \eta_t.$$

We choose $\rho_x = \rho_z = \rho$, hence we have

$$\begin{aligned} x_t &= \frac{\sigma_{23}}{\sigma_{33}}z_t + v_t, \\ v_t &= \rho v_{t-1} + \eta_t. \end{aligned}$$

If we now use only z_t as instruments we have serial correlation in the reduced form. Of course, we could use x_{t-1} and z_{t-1} as instruments as well to circumvent this problem. We set $n = 500$ while varying the value for ρ. Choosing $\rho = 0$ corresponds to the i.i.d. case, while $0 < \rho < 1$ introduces serial correlation. We set $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$ and choose $\sigma_{12} = 0.5$. This implies $\sigma_{23} = \rho_{23}$ for which we have the following when we fix the concentration parameter μ across experiments:

$$\rho_{23} = \sqrt{\frac{\frac{\mu}{nk_z}}{1 + \frac{\mu}{nk_z}}}.$$

We set again $\mu/k_z = 9$ which in the i.i.d. case corresponds to a population first stage F-statistic of 10 and relative bias of around 10%. Alternatively, we fix relative bias at 10% and let μ vary according to (5.2). The upper and lower panels of Table 2 report the simulation results for fixing μ and relative bias respectively.

Table 2: bias and F statistic in case of AR(1) errors

ρ	bias OLS	bias IV	relbias	F	F_r
			$\mu/k_z = 9$		
0	0.498	0.032	0.064	10.81	11.42
0.3	0.453	0.032	0.070	11.07	6.80
0.6	0.320	0.032	0.099	12.16	4.01
0.9	0.098	0.024	0.244	20.84	3.07
			$Relbias = 0.1$		
0	0.498	0.045	0.090	7.82	8.26
0.3	0.453	0.044	0.097	8.13	4.99
0.6	0.320	0.038	0.119	10.28	3.39
0.9	0.098	0.013	0.130	38.66	5.69

Note: average numbers over 2000 replications

Regarding the robust F-statistic we used the HAC estimator proposed by Newey and West (1987) exploiting a lag length of 10. Indeed we see that the pattern of the simulation results follows the theoretical predictions In the i.i.d. case ($\rho = 0$) results are similar to existing simulation results on relative bias and first stage F-statistic. But more importantly in case of serial correlation ($\rho > 0$) and for fixed μ we have that relative bias and standard F-statistic increase, while the robust F-statistic decreases. Again we need a considerably larger standard F-statistic compared with the i.i.d. case. However, contrary to the grouped error design the decrease in the robust F-statistic is enough to offset the increase in relative bias as can be seen from the lower panel of Table 2. Apparently even values lower than 10 suffice to limit relative IV/OLS bias to 10%.

7. Concluding remarks

In this study we have analyzed the usefulness of the first stage F-statistic for detecting weak instruments in the IV model with a nonscalar error covariance matrix. In particular, we have investigated whether there is like in the case of i.i.d. errors, a close relation between first stage F-statistic and bias of the IV estimator, relative to the bias of the OLS estimator. For two particular well known error covariance structures, i.e. grouped errors and autoregressive errors, we have embarked on this relation in more detail. Employing asymptotic expansion techniques to approximate IV and OLS biases we have shown that in these cases for a fixed concentration parameter more relative bias goes hand in hand with a larger first stage F-statistic compared with the case of i.i.d. errors. In other words, in case of correlated errors the often used rule of thumb of 10 for the first stage F-statistic implying roughly 10% relative bias is too low.

We also analyzed a version of the first stage F-statistic robust for correlated errors. Especially in case of the cross-section IV model with a grouped error structure this statistic is reported routinely by applied researchers. For our two particular examples of correlated errors we have found that there is no close relation between robust first stage F-statistic and relative bias. It depends on the particular model whether it is over- or underestimating the strength of the instruments. Our simulation results show that in case of grouped errors a value of 10 is certainly not always sufficient to warrant limited bias of IV versus OLS. However, in case of autoregressive errors our simulations indicate that the rule of thumb of 10 applied to the robust F-statistic can serve as a conservative lower bound in judging the strength of instruments.

Although the scope of the current analysis has been limited to a few special cases we nevertheless think that two main conclusions emerge for the use of the first stage F-statistic in applied research. First, reporting both standard and robust F-statistics instead of either one can give some insight into the problem of weak instruments. Second, while the use of a robust version of the test statistic seems more conservative there are some potential threats to its use as well. There is no close correspondence with relative bias anymore and, in addition, we do not know the exact covariance structure in practice. For example, in case of grouped data what happens if we cluster at the wrong level? Hence, proper modeling of the reduced form, i.e. avoiding misspecification, seems warranted for successful interpretation of first stage F-statistics for detecting weak instruments.

References

Angrist, J.D. and A.B. Krueger (1991). Does compulsory school attendance affect schooing and earnings? *Quarterly Journal of Economics* 106, 979-1014.

Arellano, M. (1987). Computing robust standard errors for within groups estimators. *Oxford Bulletin of economics and Statistics* 49, 431-434.

Bound, J., D.A. Jaeger and R.M. Baker (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association 90*, 443-450.

Bun, M.J.G., and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. Accepted for publication in *Econometrics Journal*.

Buse, A. (1992). The Bias of Instrumental Variable Estimators. *Econometrica 60*, 173-180.

Hahn, J. and G. Kuersteiner (2002). Discontinuities of Weak Instrument Limiting Distributions. *Economics Letters 75*, 325-331.

Kleibergen, F. (2007). Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics* 139, 181-216.

Nagar, A.L. (1959). The Bias and Moment Matrix of the General k-class Estimators of the parameters in Simultaneous Equations. *Econometrica 27*, 575-595.

Newey, W.K. and K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703-708.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* 96(1), 152-175.

Pischke, J-S and T. von Wachter (2008). Zero returns to compulsory schooling in Germany: evidence and interpretation. *The Review of Economics and Statistics* 90(3), 592-598.

Shore-Sheppard, L. (1996). The precision of instrumental variables estimates with grouped data. Working paper No. 374, Industrial relations section, Princeton University.

Staiger, D. and J.H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65(3), 557-586.

Stock, J.H., J.H. Wright and M. Yogo (2002). A survey of weak instruments and weak identification in Generalized Method of Moments. *Journal of Business & Economic Statistics* 20(4), 518-529.

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

A. Bias approximation

In this appendix we derive the bias approximation of the IV estimator in case of the IV model with nonscalar covariance matrix. We have for the 2SLS estimation error (3.4) the following

$$\widehat{\beta}_{2SLS} - \beta = \frac{c}{d}.$$

We assume a finite number of instruments, hence k_z is $O(1)$. Furthermore, assume that these instruments are not weak, i.e. $\pi' Z' Z \pi$ is $O(n)$. Then we have that $E[c] = O(1)$, $E[d] = O(n)$, $Var[c] = O(n)$, $Var[d] = O(n)$, hence $c = O_p(n^{1/2})$ and $d = O_p(n^{1/2})$. Defining $\bar{d} = E(d)$ we may express the denominator of the estimation error as follows:

$$\begin{aligned} d &= \bar{d} + d - \bar{d} \\ &= \bar{d}(1 + \bar{d}^{-1}(d - \bar{d})), \end{aligned}$$

hence we have

$$d^{-1} = \bar{d}^{-1}(1 + \bar{d}^{-1}(d - \bar{d}))^{-1}.$$

Noting that $\bar{d}^{-1}(d - \bar{d}) = O_p(n^{-1/2})$ the second factor expands as follows

$$(1 + \bar{d}^{-1}(d - \bar{d}))^{-1} = 1 - \bar{d}^{-1}(d - \bar{d}) + O_p(n^{-1}).$$

Using this expansion we write the estimation error as

$$\begin{aligned} \widehat{\beta}_{2SLS} - \beta &= c\bar{d}^{-1}(1 - \bar{d}^{-1}(d - \bar{d})) + O_p(n^{-3/2}) \\ &= \frac{2c}{\bar{d}} - \frac{cd}{\bar{d}^2} + O_p(n^{-3/2}). \end{aligned}$$

Taking expectations we get

$$\begin{aligned} E\left[\widehat{\beta}_{2SLS} - \beta\right] &= c\bar{d}^{-1}(1 - \bar{d}^{-1}(d - \bar{d})) + O_p(n^{-3/2}) \\ &= \frac{2E[c]}{\bar{d}} - \frac{E[cd]}{\bar{d}^2} + o(n^{-1}). \end{aligned}$$

Evaluating both expectations on the right hand side of this expression we get

$$\begin{aligned} E[c] &= E[v' P_Z u] = E[tr(P_Z u v')] \\ &= tr(P_Z \Omega_{uv}) = tr((Z'Z)^{-1} Z' \Omega_{uv} Z) \\ &= tr(Q_{ZZ}^{-1} Q_{ZZ}^{*}). \end{aligned}$$

Furthermore, the product cd consists of six terms of which four have expectation zero because they involve odd moments from the multivariate Normal distribution. Hence, we have

$$\begin{aligned} E[cd] &= E[(\pi' Z' u + v' P_Z u)(\pi' Z' Z\pi + 2\pi' Z' v + v' P_Z v)] \\ &= 2\pi' Z' E[uv'] Z\pi + \pi' Z' Z\pi E[v' P_Z u] \\ &= 2\pi' Z' \Omega_{uv} Z\pi + \pi' Z' Z\pi tr(P_Z \Omega_{uv}) \\ &= 2n\pi' Q^*_{ZZ}\pi + \pi' Z' Z\pi tr(Q^{-1}_{ZZ} Q^*_{ZZ}). \end{aligned}$$

Finally, noting that $\bar{d} = \pi' Z' Z\pi + tr(P_Z \Omega_v)$ and defining $\bar{d}_1 = \pi' Z' Z\pi$ and $\bar{d}_2 = tr(P_Z \Omega_v)$ we have

$$\begin{aligned} \frac{1}{\bar{d}} &= \frac{1}{\bar{d}_1} - \frac{\bar{d}_2}{\bar{d}_1(\bar{d}_1 + \bar{d}_2)} \\ &= \frac{1}{\bar{d}_1} + O(n^{-2}), \end{aligned}$$

$$\begin{aligned} \frac{1}{\bar{d}^2} &= \frac{1}{\bar{d}_1^2} - \frac{\bar{d}_2(2\bar{d}_1 + \bar{d}_2)}{\bar{d}_1^2(\bar{d}_1 + \bar{d}_2)^2} \\ &= \frac{1}{\bar{d}_1^2} + O(n^{-3}), \end{aligned}$$

hence

$$E\left[\widehat{\beta}_{2SLS} - \beta\right] = \frac{2E[c]}{\bar{d}_1} - \frac{E[cd]}{\bar{d}_1^2} + o(n^{-1}).$$

Hence, we have for the approximate bias

$$\begin{aligned} E\left[\widehat{\beta}_{2SLS} - \beta\right] &= \frac{2tr(Q^{-1}_{ZZ} Q^*_{ZZ})}{\pi' Z' Z\pi} - \frac{2n\pi' Q^*_{ZZ}\pi + \pi' Z' Z\pi tr(Q^{-1}_{ZZ} Q^*_{ZZ})}{(\pi' Z' Z\pi)^2} + o(n^{-1}) \\ &= \frac{tr(Q^{-1}_{ZZ} Q^*_{ZZ})}{\pi' Z' Z\pi} - \frac{2n\pi' Q^*_{ZZ}\pi}{(\pi' Z' Z\pi)^2} + o(n^{-1}) \\ &= \frac{1}{n}\left(\frac{tr(Q^{-1}_{ZZ} Q^*_{ZZ})}{\pi' Q_{ZZ}\pi} - \frac{2\pi' Q^*_{ZZ}\pi}{(\pi' Q_{ZZ}\pi)^2}\right) + o(n^{-1}), \end{aligned}$$

which is the result in (3.6). In case of i.i.d. errors we have $Q^*_{ZZ} = \sigma_{uv} Q_{ZZ}$ and the above result further simplifies to the expression in (2.7) as derived by Nagar (1959), Buse (1992) and Hahn and Kuersteiner (2002).