ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 79/80 – 2005

Estimation with Valid and Invalid Instruments

Jinyong HAHN† and Jerry HAUSMAN‡

ABSTRACT. – We demonstrate analytically that for the widely used simultaneous equation model with one jointly endogenous variable and valid instruments, 2SLS has smaller MSE error, up to second order, than OLS unless the R^2, or the F statistic of the reduced form equation is extremely low. We also consider the relative bias of estimators when the instruments are invalid, i.e. the instruments are correlated with the stochastic disturbance. Here, both 2SLS and OLS are biased in finite samples and inconsistent. We investigate conditions under which the approximate finite sample bias or the MSE of 2SLS is smaller than the corresponding statistics for the OLS estimator. We again find that 2SLS does better than OLS under a wide range of conditions, which we characterize as functions of observable statistics and one unobservable statistic.

Estimation avec instruments valides et invalides

RÉSUMÉ. – Nous démontrons analytiquement que, pour le modèle largement utilisé à équations simultanées avec une variable conjointement endogène et des instruments valides, les DMC ont une plus petite erreur quadratique (MSE), jusqu'au second ordre, que les MCO à moins que la statistique de R^2 ou de Fischer de l'équation en forme réduite soit extrêmement faible. Nous examinons aussi le comportement des estimateurs lorsque les instruments ne sont pas valides, c'est-à-dire lorsqu'ils sont corrélés avec la perturbation stochastique. Dans ce cas, à la fois les DMC et les MCO sont biaisés et non convergents à distance finie. Nous recherchons les conditions pour lesquelles le biais à distance finie ou l'erreur quadratique moyenne est plus faible pour les DMC que pour les MCO. Nous trouvons encore que les DMC se comportent mieux que les MCO sous un large ensemble de conditions, caractérisées par des fonctions des statistiques observables et d'une statistique non observable.

† J. Hahn: University of California, Los Angeles.
‡ J. Hausman: Massachusetts Institute of Technology (MIT).

Introduction

While 2SLS is the most widely used estimator for simultaneous equation models, OLS may do better in finite samples. Econometricians have recognized this possibility, and many Monte Carlo studies were undertaken in the early years of econometrics to attempt to determine condition when OLS might do better than 2SLS. Here we demonstrate analytically that for the widely used simultaneous equation model with one jointly endogenous variable and valid instruments, 2SLS has smaller MSE error, up to second order, than OLS unless the R^2, or the F statistic of the reduced form equation is extremely low.

We then consider the relative bias of estimators when the instruments are invalid, i.e. the instruments are correlated with the stochastic disturbance. Here, both 2SLS and OLS are biased in finite samples and inconsistent. We investigate conditions under which the approximate finite sample bias or the MSE of 2SLS is smaller than the corresponding statistics for the OLS estimator. We again find that 2SLS does better than OLS under a wide range of conditions, which we characterize as functions of observable statistics and one unobservable statistic.

We then present a method of sensitivity analysis, which calculates the maximal asymptotic bias of 2SLS under small violations of the exclusion restrictions. For a given correlation between invalid instruments and the error term, we derive the maximal asymptotic bias. We demonstrate how the maximal asymptotic bias can be estimated in practice.

Next, we turn to inference. In the "weak instruments" situation the bias in the 2SLS estimator creates a problem, since it is biased towards the OLS estimator, which is also biased. The other problem that arises is that the estimated standard errors of the 2SLS estimator are often much too small to signal the problem of imprecise estimates. Here we derive the bias in the estimated standard errors for the first time, which turns out to cause the problem. This derivation also has implications for the test of over-identifying restrictions.

We do not survey the weak instruments literature. For recent surveys see Stock *et. al.* [2002] and Hahn and Hausman [2003].

1 Model specification

We begin with the model specification with one right hand side (RHS) jointly endogenous variable so that the left hand side (LHS) variable depends only on the single jointly endogenous RHS variable. This model specification accounts for other RHS strictly exogenous variables, which have been "partialled out" of the specification. We will assume that[1]

1 Without loss of generality we normalize the data such that y_2 has zero mean.

(1.1) $$y_1 = \beta y_2 + \varepsilon_1$$

(1.2) $$y_2 = z\pi_2 + \nu_2,$$

where dim $(\pi_2) = K$. Thus, the matrix z is the matrix of all strictly exogenous variables, and equation (1.1) is the reduced form equation for y_2 with coefficient vector π_2. We also assume homoscedasticity:

(1.3) $$\begin{pmatrix} \varepsilon_{1i} \\ \nu_{2i} \end{pmatrix} \sim N(0, \Sigma) \sim N\left(0, \begin{bmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\nu} \\ \sigma_{\varepsilon\nu} & \sigma_{\nu\nu} \end{bmatrix}\right).$$

We use the following notation:

$$y \equiv \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad z \equiv \begin{pmatrix} z_1' \\ \vdots \\ z_n' \end{pmatrix}, \quad \sigma_{\varepsilon\varepsilon} \equiv \mathrm{Var}(\varepsilon_{1i}),\ \sigma_{\nu\nu} = Var(\nu_{2i}),\ \sigma_{\varepsilon\nu} \equiv \mathrm{Cov}(\varepsilon_{1i}, \nu_{2i}).$$

We initially assume the presence of valid instruments, $E[z'\varepsilon/n] = 0$ and $\pi_2 \neq 0$. Throughout this paper, we assume that $\Theta \equiv \pi_2' z' z \pi_2 / n$ is fixed as in Bekker [1994]. We also assume:

Condition 1: $K \to 0$ *as* $n \to \infty$ *such that* $K/\sqrt{n} = \mu + o(1)$ *for some* $\mu \neq 0$.

Bekker [1994] introduced an alternative asymptotics, where $K = O(n)$. Bekker asymptotics are a simpler version of a higher (third) order asymptotic approximation[2]. The approximation $K = O(\sqrt{n})$ adopted here is a simpler version of a second order asymptotics, which will highlight the role of the second order bias in various instrumental variable estimators.

2 Estimation with valid instruments

In this section, we calculate the second order properties of 2SLS and OLS, and demonstrate analytically that 2SLS has smaller bias and mean squared error (MSE), up to second order, than OLS unless the R^2, or the F statistic of the reduced form equation is extremely low.

2 See, e.g., Hahn and Hausman [2002a].

We first characterize asymptotic properties of 2SLS under Condition 1. As a special case of Theorem 3 in Section 3, we obtain that:

THEOREM 1: $\sqrt{n}\left(b_{2SLS}-\beta\right)\Rightarrow N\left(\frac{\sigma_{\varepsilon v}\mu}{\Theta}, V_{2SLS}\right)$, *where* $V_{2SLS}=\sigma_{\varepsilon\varepsilon}/\Theta$.

Note that Theorem 1 predicts that the asymptotic variance of $\sqrt{n}\left(b_{2SLS}-\beta\right)$ is equal to the usual 2SLS first order asymptotic variance. Theorem 1 also predicts that the approximate bias of 2SLS:

$$E\left[b_{2SLS}\right]-\beta\approx\frac{K\sigma_{\varepsilon v}}{n\Theta}=\frac{K\sigma_{\varepsilon v}}{nR^2\,\mathrm{var}(y_2)}=B_2, \tag{2.1}$$

where $R^2\equiv\Theta/\mathrm{var}(y_2)$ is the theoretical value from the second (reduced form) equation[3]. As a consequence, we obtain the approximate mean squared error (MSE) of 2SLS:

$$MSE(2SLS)\approx\frac{\sigma_{\varepsilon v}{}^2\mu^2}{n\Theta^2}+\frac{V_{2SLS}}{n}=\frac{K^2\sigma^2_{\varepsilon v}}{R^4(y_2'y_2)^2}+\frac{\sigma_{\varepsilon\varepsilon}}{R^2(y_2'y_2)}\equiv M_{2SLS}. \tag{2.2}$$

Note that both terms in equation (2.2) approach zero as $(y_2'\,y_2)$ increases with increasing sample size. The first term, bias squared also approach zero more quickly, as expected, since 2SLS is $\sqrt{n}$ -consistent.
We now characterize the asymptotic properties of OLS under Condition 1. As a special case of Theorem 4 in Section 3, we obtain the distribution for the OLS estimator:

THEOREM 2: $\sqrt{n}\left(b_{OLS}-\left(\beta+\frac{\sigma_{\varepsilon v}}{\Theta+\sigma_{vv}}\right)\right)\rightarrow N(0,V_{OLS})$,

where $$V_{OLS}\equiv\frac{\sigma_{\varepsilon\varepsilon}}{\Theta+\sigma_{vv}}-\frac{\sigma^2_{\varepsilon v}}{(\Theta+\sigma_{vv})^2}-\frac{2\sigma^2_{\varepsilon v}\Theta^2}{(\Theta+\sigma_{vv})^4}.$$

Theorem 2 predicts the approximate bias and approximate variance as:

$$E\left[b_{OLS}\right]-\beta\approx\frac{\mathrm{cov}\left(y_2,\varepsilon\right)}{\mathrm{var}\left(y_2\right)}=\frac{\sigma_{\varepsilon v}}{\Theta+\sigma_{vv}}=B_{OLS}, \tag{2.3}$$

$$V_{OLS}=\frac{\sigma_{\varepsilon\varepsilon}}{\Theta+\sigma_{vv}}-\frac{\sigma^2_{\varepsilon v}}{(\Theta+\sigma_{vv})^2}-\frac{2\sigma^2_{\varepsilon v}\Theta^2}{(\Theta+\sigma_{vv})^4}. \tag{2.4}$$

---

3 Note that the approximate bias of 2SLS in equation (2.1) is identical to the well-known result for the second order bias of 2SLS. See, e.g., ROTHENBERG [1983] or HAHN and HAUSMAN [2002a, 2002b].

Thus, the approximate MSE of OLS is

$$MSE(OLS) \approx \frac{\sigma_{\varepsilon v}^2}{(\Theta + \sigma_{vv})^2} + \frac{V_{OLS}}{n} \equiv M_{OLS}. \tag{2.5}$$

The inconsistency of OLS is evident from equation (2.5) because while the second term goes to zero as *n* becomes large, the first term is not a function of *n*.

We now compare the approximate finite sample properties of 2SLS and OLS. We first compare biases:

$$\frac{B_{2SLS}}{B_{OLS}} = \frac{K}{nR^2} = \frac{1}{1-R^2}\frac{1}{F}, \tag{2.6}$$

where $F \equiv \frac{R^2/K}{\left(1-R^2\right)/n}$ can be interpreted as the "theoretical" *F*-statistic from the first-stage reduced form. Thus, if $F \gg 1$, 2SLS has less bias. However the OLS variance is less than the 2SLS variance so we compare the MSEs below.

Before leaving the bias comparisons, we also consider what happens when we are close to being unidentified so that $\pi_2 = a/\sqrt{n}$, where the vector *a* has dimension *K*. Thus, the reduced form coefficients are "local to zero". With $\pi_2 = a/\sqrt{n}$, equation (2.1) predicts the bias of 2SLS to be

$$E\left[b_{2SLS}\right] - \beta \approx \frac{K\sigma_{\varepsilon v_2}}{\Psi} = \frac{\sigma_{\varepsilon v_2}}{\frac{1}{K}\Psi} \equiv B_{L2SLS}, \tag{2.7}$$

where $\Psi \equiv a'z'za$. On the other hand, equation (2.3) predicts the approximate bias for OLS to be:

$$E\left[b_{OLS}\right] - \beta \approx \frac{\sigma_{\varepsilon v}}{\frac{1}{n}\Psi + \sigma_{vv}} \equiv B_{LOLS}. \tag{2.8}$$

Taking the ratio of the biases under local to zero asymptotics:

$$\frac{B_{L2SLS}}{B_{LOLS}} = \frac{\frac{1}{n}\Psi + \sigma_{vv}}{\frac{1}{K}\Psi}. \tag{2.9}$$

From equation (2.9), it follows that the bias of 2SLS is smaller than OLS as long as $K \ll n$, a condition which will usually be satisfied in practice.

We next compare the MSE of 2SLS to the MSE of OLS. It is convenient to introduce a normalization, which will simplify the M_{2SLS} and M_{OLS} expressions. Without loss of generality we rescale the units of variables $\sigma_{\varepsilon\varepsilon} = \sigma_{\nu\nu} = 1$ so that $\text{var}(y_2) = 1/(1\text{-}R^2)$ and $\sigma_{\varepsilon\nu} = \rho$.[4] Using this normalizations we find:

$$(2.10)\qquad M_{2SLS} = \left(\frac{K^2\rho^2\left(1-R^2\right)+nR^2}{n^2R^2}\right)\left(\frac{1-R^2}{R^2}\right),$$

$$(2.11)\qquad \begin{aligned} M_{OLS} &= \frac{\sigma_{\varepsilon\nu}^2}{\text{var}^2(y_2)} + \frac{\sigma_{\varepsilon\varepsilon}}{n\,\text{var}(y_2)} - \frac{\sigma_{\varepsilon\nu}^2}{n\,\text{var}^2(y_2)} - \frac{2\sigma_{\varepsilon\nu}^2 R^4}{n\,\text{var}^2(y_2)} \\ &= \left(\frac{\rho^2(n-1-2R^4)(1-R^2)+1}{n}\right)\left(1-R^2\right) \end{aligned}$$

and

$$(2.12)\qquad \frac{M_{2SLS}}{M_{OLS}} = \frac{K^2\rho^2(1-R^2)+nR^2}{(nR^4)[(n-1-2R^4)\rho^2(1-R^2)+1]}.$$

Which estimator to use will depend on whether equation (2.12) is less than or greater than unity. We can solve for the "critical value" of ρ^2 which causes the MSE of the 2 estimators to be equal.[5] The solution for this "critical value" has a remarkably simple form:

$$(2.13)\qquad \rho^2 = \frac{nR^2}{nR^4(n-1-2R^4)-K^2}.$$

As n becomes large the "critical value" of ρ^2 goes to zero. In any particular sample R^2 and F can typically be accurately estimated from the unbiased estimates of the reduced form so that only ρ^2 is unknown. While this parameter value is typically unknown, the applied econometrician will often have a good (*a priori*) knowledge of the possible values of ρ so that she will be able to determine whether the critical value is below the square of the correlation coefficient.[6] As we now demonstrate, the critical value is often so low that 2SLS will have a lower MSE than OLS, even for situation with relatively "weak instruments" or a low F statistic.

In Table 1 we calculate the critical value of ρ (using the absolute value) for a range of values of R^2 for K of 5, 10, and 30 and for sample sizes of $n = 500$ and $n = 1{,}000$. Here we find that if $R^2 \geq 0.1$ that 2SLS typically will have a lower MSE. Thus, except in the case of weak instruments, which can arise when both R^2 is low and the number of instruments is high, 2SLS is typically the preferred estimator based on an approximate finite sample comparison of MSEs.

---

4 Structural equations are homogeneous of degree zero. Since we rescale the variance of the stochastic disturbance we have to either rescale the coefficients or rescale the units of the variables. We adopt the latter convention here, although we can do either.

5 The correlation parameter ρ is the key parameter in simultaneous equation analysis because if it is zero the OLS estimator is the unbiased Gauss-Markov estimator and the ratio of MSEs in equation (2.12) equals $1/R^2 > 1$, but OLS is biased and inconsistent if the parameter value of ρ is not zero.

6 The parameter ρ is also estimated from the 2SLS estimation, but a good estimate may be difficult to achieve in a "weak instrument" situation.

TABLE 1
Critical Values of ρ

R^2	0.01	0.1	0.2	0.3	0.5	0.7	0.9
$K = 5$							
100	**	0.3677	0.2323	0.1863	0.1432	0.1210	0.1070
500	**	0.1423	0.1002	0.0818	0.0634	0.0536	0.0473
1,000	0.3654	0.1002	0.0708	0.0578	0.0448	0.0378	0.0334
$K = 10$							
100	**	**	0.2601	0.1949	0.1455	0.1220	0.1075
500	**	0.1445	0.1006	0.0819	0.0634	0.0536	0.0473
1,000	**	0.1006	0.0708	0.0578	0.0448	0.0378	0.0334
$K = 10$							
100	**	**	**	**	0.1789	0.1339	0.1135
500	**	0.1771	0.1050	0.0834	0.0638	0.0538	0.0474
1,000	**	0.1049	0.0716	0.0581	0.0448	0.0379	0.0334

*Notes : ** denotes no critical value of* ρ *less than 1.0 exists.*

3 Estimation with invalid instruments

Up to this point we have assumed that the instruments are valid so that they are orthogonal to the stochastic disturbance ε_1. However, the econometrician may not be certain that the instruments satisfy the orthogonality condition. We now consider the situation where the orthogonality condition on the instruments fails so that $E[z'\varepsilon_1/n] \neq 0$. We first consider the traditional "large sample bias" of 2SLS:

(3.1) $$\text{plim}\left[b_{2SLS}\right] - \beta \approx \frac{\sigma_{W'\varepsilon}}{R^2 \sigma_{y_2 y_2}},$$

where $W = z\pi_2$. When we compare this expression with the analogous expression for OLS

(3.2) $$\text{plim}\left[b_{OLS}\right] - \beta \approx \frac{\sigma_{\sigma_2'\varepsilon}}{\sigma_{y_2 y_2}}.$$

In general either estimator may be preferred on this criterion depending on circumstances. The numerator of equation (3.1) would likely be smaller ("less correlation" in the instrument) than the numerator of equation (3.2), but the denominator of equation (3.1) is always smaller since $R^2 < 1$. Indeed, if R^2 is very small, the OLS estimator may do better in terms of inconsistency.

In order to gain a better insight, we adopt an asymptotic approximation similar to the one in the previous section, and investigate conditions under which the approximate finite sample bias or the MSE of 2SLS is smaller than the corresponding statistics for the OLS estimator. We again find that 2SLS does better than OLS under a wide range of conditions, which we characterize as functions of observable statistics and one unobservable statistic.

To do asymptotic approximations we need to specify the correlation of the instrument with the stochastic disturbance in the structural equation (1.1). We use a local specification:

$$\varepsilon_1 = z(\gamma/\sqrt{n}) + e_1 \; for \; \gamma \neq 0. \tag{3.3}$$

We assume that (e_1, ν_2) is homoscedastic and zero mean normally distributed with covariance matrix:

$$\begin{pmatrix} e_{1i} \\ \nu_{2i} \end{pmatrix} \sim N(0,\Omega) \sim N\left(0, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right).$$

Throughout this section, we also assume that $\Xi \equiv \pi_2' z' z\gamma/n$ is fixed.

First, we derive the asymptotic distribution of the 2SLS estimator with locally invalid instruments[7]:

THEOREM 3: *Under Condition 1,* $\sqrt{n}(b_{2SLS} - \beta) \Rightarrow N\left(\frac{\Xi + \mu\sigma_{12}}{\Theta}, V_{2SLS}\right)$.

The first term Ξ in the numerator of the mean arises from failure of the orthogonality condition. The second term is the usual finite sample bias term and it decreases with the sample size. The variance continues to be V_{2SLS} under instrument invalidity because of the local departure in equation (3.3) similar to HAUSMAN ([1978], p. 1256).

We use Theorem 3 to calculate the approximate bias of the 2SLS estimator with invalid instruments is:

$$B_{I2SLS} \equiv \frac{\Xi/\sqrt{n} + K\sigma_{12}/n}{\Theta} = \frac{1-R^2}{R^2}\left(\frac{1}{\sqrt{n}}\alpha\rho + \frac{1}{n}K\rho\right), \tag{3.4}$$

where we use the previous normalizations and set $\sigma_{W'\varepsilon} = \Xi/\sqrt{n} \equiv \alpha\rho/\sqrt{n}$. Using Theorem 3 we find the approximate MSE of 2SLS to be:

$$\begin{aligned} MSE_{I2SLS} &\equiv \frac{(\Xi + \mu\sigma_{12})^2}{\Theta^2}\frac{1}{n} + \frac{V_{2SLS}}{n} \\ &= \left(\frac{1-R^2}{R^2}\right)^2 \left(\frac{1}{\sqrt{n}}\alpha\rho + \frac{1}{n}K\rho\right)^2 + \frac{1}{n}\left(\frac{1-R^2}{R^2}\right) \end{aligned} \tag{3.5}$$

7 See Appendix for proof of Theorem 3.

We now derive the asymptotic distribution of the OLS estimator with locally invalid instruments[8]:

THEOREM 4: *Under Condition 1,*

$$\sqrt{n}\left(b_{OLS}-\left(\beta+\frac{\sigma_{12}}{\Theta+\sigma_{22}}\right)\right)\Rightarrow N\left(\frac{\Xi}{\Theta+\sigma_{22}},V_{OLS}\right).$$

The distribution is centered around the usual OLS bias, as before, and the numerator of the mean of the distribution arises from the instrument invalidity. Again, the variance continues to be V_{OLS} under instrument invalidity because of the local departure in equation (3.3). Using Theorem 4 and the previous normalizations, we find the approximate MSE of OLS to be:

$$\begin{aligned} MSE_{IOLS} &\equiv \left(\frac{\sigma_{12}}{\Theta+\sigma_{22}}+\frac{1}{\sqrt{n}}\frac{\Xi+\mu\sigma_{12}}{\Theta+\sigma_{22}}\right)^2+\frac{V_{OLS}}{n} \\ &= (1-R^2)^2\left(\frac{1}{\sqrt{n}}\alpha\rho+\rho\right)^2+\left(\frac{-\rho^2(1+2R^4)(1-R^2)+1}{n}\right)\left(1-R^2\right) \end{aligned} \tag{3.6}$$

The first term in parentheses is the "usual" simultaneous equation bias of OLS that does not decrease with the sample size.

We now compare the bias of 2SLS under instrument invalidity with the bias of OLS given similar circumstances. We re-write the bias of OLS using the normalization:

$$B_{IOLS}=(1-R^2)\left(\frac{1}{\sqrt{n}}\alpha\rho+\rho\right). \tag{3.7}$$

As before, we take the ratio of (3.4) and (3.7):

$$\frac{B_{I2SLS}}{B_{IOLS}}=\frac{\alpha\rho/\sqrt{n}+K\rho/n}{R^2(\alpha\rho/\sqrt{n}+\rho)}. \tag{3.8}$$

The ratio of the biases is homogeneous of degree zero in the correlation coefficient ρ, so we can simplify terms. We plot the ratio of the biases in Figure 1 for the case of $n = 100$ and $K = 5$ and $\alpha = 0.1$.

We find that the 2SLS bias is less than the OLS bias if:

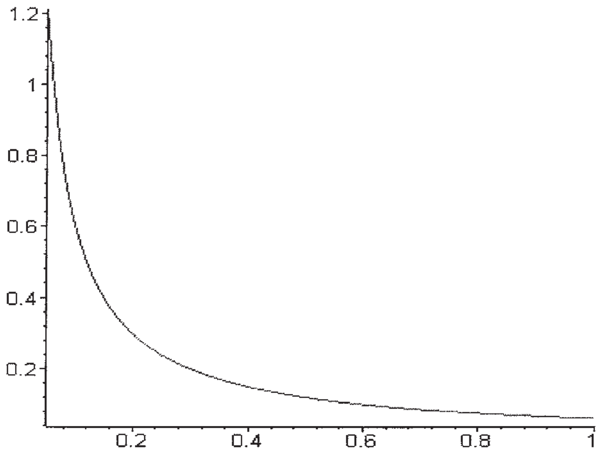
$$\alpha<\frac{1}{\sqrt{n}}\frac{nR^2-K}{1-R^2}. \tag{3.9}$$

---

8 See Appendix for proof of Theorem 4.

FIGURE 1

Ratio of 2SLS Bias to OLS Bias with Invalid Instruments
$N = 100$, $K = 5$, $\alpha = 0.1$



Equation (3.9) is very easy to interpret. We calculate a "critical α" in Figure 2, and note that it increase quite rapidly, so that the bias of 2SLS with invalid instruments remains less than the bias of OLS so long as F exceeds 1.0 by a small amount. The straightforward relationship of equation (3.9) allows for an easy interpretation on which the econometrician may well have some a priori knowledge.

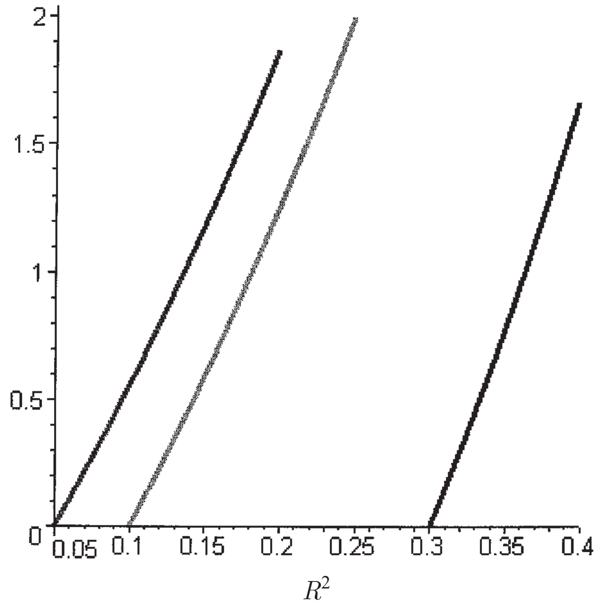
Note that the common empirical finding that the 2SLS coefficient is larger than the OLS coefficient can arise because of the OLS bias when the instruments are valid or because of an improper instrument. Thus, even if the instrument is "almost uncorrelated" so that $\sigma_{W'\varepsilon} \approx 0$ substantial bias can still arise because R^2 is often quite small in the weak instruments situation. Thus, comparing equation (3.4) to the bias of OLS in equation (3.7), the empirical finding that the 2SLS estimate increases compared to the OLS estimate may indicate that the instrument is not orthogonal to the stochastic disturbance. The resulting bias can be substantial. Indeed, it could exceed the OLS bias, leading to an increase in the estimated 2SLS coefficient over the estimated OLS coefficient.

Returning to the general situation and using the normalizations the ratio of the MSEs is

$$(3.10) \qquad \frac{M_{I2SLS}}{M_{IOLS}} = \frac{(1-R^2)\left(\alpha\rho + K\rho/\sqrt{n}\right)^2 / R^4 + 1/R^2}{(1-R^2)[(\alpha\rho+\sqrt{n}\rho)^2 + 1 - (1-R^2)\rho^2 - 2(1-R^2)\rho^2 R^4]}.$$

No straightforward condition can be derived where the ratio is less than one. In order to gain some insight, we calculated the ratio (3.10) for various values of R^2 and ρ fixing $\alpha = 0.1$, $K = 5$, and $n = 100$. The ratio (3.10) is below 1.0 except in the situation where R^2 becomes quite small (as with weak instruments) **and** ρ becomes small (which decreases the OLS bias).

FIGURE 2
Critical Values for Alpha
n = 100 and K = 5, 10, 30



In our comparisons of 2SLS with OLS, two sources of bias arise. The first source of bias is from the use of estimated parameters, $\hat{\pi}_2$ in equation (1.2), in forming the instruments. This source of bias disappears as the sample becomes large. The second source of bias is from the use of invalid instruments, $\gamma \neq 0$ in equation (3.3). This source of bias does not disappear sufficiently fast with the sample size to cause 2SLS to be consistent. An interesting question would be about how the comparison of IV to OLS would change if the first source of bias were eliminated. We can eliminate this source of bias (to second order) by using the Nagar estimator

$$b_N = \frac{y_2' P_z y_1 - \frac{K}{n} y_2' M_z y_1}{y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2}, \quad P_z = z(z'z)^{-1} z', \quad M_z = I - P_z.$$

We derive the asymptotic distribution of the Nagar estimator with locally invalid instruments[9]:

THEOREM 5: *Under Condition 1,* $\sqrt{n}(b_N - \beta) \Rightarrow N\left(\frac{\Xi}{\Theta}, V_{2SLS}\right)$.

9 See Appendix for proof of Theorem 5.

Thus to compare the MSE of the Nagar estimator to the MSE of the 2SLS estimator with invalid instruments, we see that the variance of the two estimators is the same, but that the bias differs as explained above. However, when we compare the bias square of 2SLS from equation (3.4) with the Nagar estimator we find that

$$(3.11) \qquad \left(\frac{\Xi/\sqrt{n}+K\sigma_{12}/n}{\Theta}\right)^2-\left(\frac{\Xi/\sqrt{n}}{\Theta}\right)^2=\left(\frac{2K\sigma_{12}\Xi/n^{3/2}+K^2\sigma_{12}^2/n^2}{\Theta^2}\right).$$

Equation (3.11) can be less than or greater than zero. Thus, we cannot conclude that using the Nagar estimator to compare with OLS would make the comparison more favorable to an IV estimator.

4 Sensitivity analysis

CARD [2001] discusses possible concerns that the instruments may be invalid in discussing the empirical literature that estimates the return to additional education. The use of instrumental variables in this situation began with GRILICHES' [1977] well known paper. To investigate the possible effect of invalid instruments, we consider the specification:

$$(4.1) \qquad \begin{aligned} y_1(\theta) &= \beta y_2 + z\theta + \varepsilon \\ \varepsilon^* &= z\theta + \varepsilon \end{aligned}$$

Note that we have added $z\theta$ to the error ε which causes the instruments to be invalid.[10] We derive the maximal asymptotic bias for a small violation of the exclusion restriction, where ψ is the correlation between $z_i\pi$ and ε_i^* so that ψ^2 is the R^2 of between $z_i\pi$ and ε_i^*. We find the maximal asymptotic bias[11] to be:

THEOREM 6: $\max\left|\text{plim } bias\ \hat{\beta}_{2SLS}(\theta)\right| = \left(\frac{1}{R^2}\frac{\text{plim } n^{-1}\sum\varepsilon_i^2}{\text{plim } n^{-1}\sum y_{2i}^2}\right)^{1/2}\left(\frac{\psi^2}{1-\psi^2}\right)^{1/2}.$

Note that the maximal asymptotic bias can be consistently estimated by

$$(4.2) \qquad \left(\frac{1}{R^2}\frac{n^{-1}\sum\hat{\varepsilon}_i^2}{n^{-1}\sum y_{2i}^2}\right)^{1/2}\left(\frac{\psi^2}{1-\psi^2}\right)^{1/2}.$$

10 IMBENS [2003] considers the question of sensitivity analysis, but not in the context of instrumental variables.

11 See Appendix for proof of Theorem 6.

The maximal asymptotic bias in (4.2) can be used to conduct a "bound analysis". The bound can often be quite large in "weak instrument" situations. This bound can sometimes conflict with the bound produced by MANSKI'S [1990, 2003] non-parametric approach, since the Manski approach does not allow for errors in variables.

5 Bias in estimated standard errors

We have previously discussed the biases in the 2SLS estimator in equation (2.1) and Theorem 1. In the "weak instruments" situation this bias may be quite large. A further problem arises in that the 2SLS estimator is biased in the same direction as the OLS estimator as equation (2.4) and Theorem 2 demonstrate. Thus, HAUSMAN [1978] specification type test will be biased towards not rejecting the null hypothesis of lack of orthogonality between ε_1 and ν_2 in equations (1.1) and (1.2). However, another problem has been recognized in the weak instruments situation. The estimated standard errors for the 2SLS estimator are downward biased, sometimes leading to the mistaken inference that the 2SLS estimate are much more precise than they actually are. From analysis based on first order asymptotics the usual conclusion would be that with "weak instruments" that the reported standard error of the 2SLS estimator would be sufficiently large to signal the finding that so much uncertainty exists with the estimate that it would not be of much use. However, researchers have found that, to the contrary, often the 2SLS estimator in the presence of weak instruments leads to a reasonably small standard error. Thus, the researcher may be unaware of the weak instruments problem. The source of the problem of small reported standard errors of the 2SLS estimator has not been discussed in the literature. Here we derive the source of the problem and offer a possible approach to fixing it.

The variance of 2SLS is derived in Theorem 1 and takes the usual form of $V_{2SLS} = \sigma_{\varepsilon\varepsilon}\Theta^{-1}$ where $\Theta = \pi' z' z\pi/n$ is assumed to be fixed. Now $\hat{\Theta}$ is not difficult to estimate since unbiased estimated of π follow from OLS on equation (1.2). Thus, the downward bias in the estimated 2SLS standard errors must arise from a downward biased estimate of $\sigma_{\varepsilon\varepsilon}$. We now derive the bias. The intuition follows from the fact that 2SLS is biased towards the OLS estimator, which minimizes $\hat{\sigma}_{\varepsilon\varepsilon}$. Thus, we find that the bias of the 2SLS estimator of β creates a bias in the 2SLS estimate of $\sigma_{\varepsilon\varepsilon}$. We find the bias to be:

THEOREM 7: $E[\hat{\sigma}^2_{2SLS}] \approx \sigma_{\varepsilon\varepsilon} - \frac{2}{n}\frac{(K-2)\sigma^2_{\varepsilon\nu}}{\Theta} - \frac{1}{n}\sigma^2_{\varepsilon\nu} + \frac{1}{n}\frac{\sigma_{\varepsilon\varepsilon}\sigma_{\nu\nu}}{\Theta}.$

Note that the leading term in the bias calculation of Theorem 5 is 2 times the bias of the 2SLS estimator from equation (2.1). As either the number of instruments grows or the covariance between the structural and reduced term stochastic disturbances becomes large, the bias in the estimation of $\sigma_{\varepsilon\varepsilon}$ will also become large. We now apply the normalization that we used above to find:

$$E[\hat{\sigma}^2_{2SLS}] \approx 1 - \frac{2}{n}\frac{(K-2)\rho^2(1-R^2)}{R^2} - \frac{1}{n}\rho^2 + \frac{1}{n}\frac{(1-R^2)}{R^2}$$
$$= 1 - \frac{1}{n}\frac{[(2K-4)\rho^2 - 1](1-R^2)}{R^2} - \frac{1}{n}\rho^2 \quad (5.1)$$

The bias can be quite substantial as demonstrated by equation (5.1). The final term in equation (4.2) will typically be small so that it can be ignored. Equation (5.1) demonstrates that the downward bias can be substantial; in Monte-Carlo[12] results reported in Table 2, we find that for $R^2 = .01$ and $\rho = 0.9$ that the mean bias of the 2SLS estimate of the variance varies from -70% to -80% as K, the number of instruments, increases from 5 to 30. Thus, we note that the bias in the estimation even when $K = 5$ can be quite large. This finding explains the result that when weak instruments are present, the estimated standard errors of 2SLS can appear to be near those of OLS and small enough to allow the researcher to make conclusions about the likely true parameter value. However, with weak instruments these conclusions could be erroneous because of the substantial bias in the estimated standard error of the 2SLS estimator[13].

We now consider the finding that the often used test of over identifying restrictions (OID test) rejects "too often" when weak instruments are present, i.e. the actual size of the test is considerably larger than the nominal size. The OID test can be quite important since it tests the economic theory embodied in the model as discuss by e.g. HAUSMAN [1983]. In the weak instrument situation it may have increased importance given the substantial bias in the 2SLS estimator and the large MSE that we calculation in equations (3.4) and (3.5). We may write the OID test as[14]:

$$W = \frac{\hat{\varepsilon}' P_Z \hat{\varepsilon}}{\hat{\sigma}_{\varepsilon\varepsilon}}. \quad (5.2)$$

W is distributed as chi-square with K - 1 degrees of freedom under conventional asymptotics. From equation (5.2), we see that a downward biased of $\sigma_{\varepsilon\varepsilon}$ can lead to substantial over-rejection and an upward biased size of the OID test. Thus, correcting for this problem can have an important effect on test results.

12 The Monte-Carlo design is the same as in HAHN-HAUSMAN [2002a].

13 We note recent development on the correctly sized confidence intervals of β, including KLEIBERGEN [2002], may be of importance in detecting these problems. The new confidence intervals may be subject to a power problem. See, e.g., ANDREWS, MOREIRA and STOCK [2004].

14 See, e.g., HAUSMAN [1983].

Table 2
Bias of $\hat{\sigma}^2$

			n = 100								n = 500								n = 1,000				
K	R^2	ρ	Mean Bias	Median Bias	MAE	RMSE	IQR	K	R^2	ρ	Mean Bias	Median Bias	MAE	RMSE	IQR	K	R^2	ρ	Mean Bias	Median Bias	MAE	RMSE	IQR
5	0.01	0	0.25	0.10	0.31	0.81	0.33	5	0.01	0	0.14	0.07	0.16	0.34	0.16	5	0.01	0	0.09	0.04	0.10	0.21	0.11
5	0.01	0.5	-0.04	-0.16	0.28	0.56	0.26	5	0.01	0.5	-0.05	-0.15	0.22	0.35	0.22	5	0.01	0.5	-0.04	-0.12	0.20	0.33	0.23
5	0.01	0.9	-0.70	-0.76	0.72	0.73	0.12	5	0.01	0.9	-0.48	-0.58	0.55	0.61	0.28	5	0.01	0.9	-0.32	-0.41	0.43	0.51	0.34
10	0.01	0	0.09	0.06	0.17	0.25	0.25	10	0.01	0	0.08	0.05	0.10	0.17	0.13	10	0.01	0	0.06	0.04	0.07	0.12	0.09
10	0.01	0.5	-0.17	-0.20	0.21	0.25	0.19	10	0.01	0.5	-0.15	-0.19	0.19	0.22	0.13	10	0.01	0.5	-0.13	-0.18	0.18	0.21	0.14
10	0.01	0.9	-0.77	-0.78	0.77	0.77	0.06	10	0.01	0.9	-0.68	-0.71	0.68	0.69	0.13	10	0.01	0.9	-0.56	-0.59	0.56	0.58	0.18
30	0.01	0	0.01	0.00	0.12	0.15	0.20	30	0.01	0	0.03	0.02	0.06	0.08	0.10	30	0.01	0	0.03	0.02	0.05	0.06	0.07
30	0.01	0.5	-0.24	-0.24	0.24	0.26	0.15	30	0.01	0.5	-0.22	-0.23	0.22	0.23	0.08	30	0.01	0.5	-0.21	-0.22	0.21	0.22	0.07
30	0.01	0.9	-0.80	-0.80	0.80	0.80	0.04	30	0.01	0.9	-0.78	-0.78	0.78	0.78	0.03	30	0.01	0.9	-0.74	-0.75	0.74	0.75	0.06
5	0.1	0	0.06	0.03	0.15	0.23	0.22	5	0.1	0	0.02	0.01	0.06	0.07	0.09	5	0.1	0	0.01	0.01	0.04	0.05	0.06
5	0.1	0.5	-0.05	-0.11	0.21	0.30	0.27	5	0.1	0.5	-0.01	-0.03	0.11	0.14	0.18	5	0.1	0.5	-0.01	-0.02	0.08	0.10	0.13
5	0.1	0.9	-0.31	-0.39	0.41	0.47	0.34	5	0.1	0.9	-0.07	-0.11	0.19	0.24	0.28	5	0.1	0.9	-0.04	-0.06	0.14	0.17	0.21
10	0.1	0	0.04	0.03	0.13	0.18	0.21	10	0.1	0	0.01	0.01	0.05	0.07	0.09	10	0.1	0	0.01	0.01	0.04	0.05	0.06
10	0.1	0.5	-0.13	-0.16	0.19	0.23	0.21	10	0.1	0.5	-0.05	-0.06	0.11	0.14	0.16	10	0.1	0.5	-0.03	-0.04	0.08	0.10	0.12
10	0.1	0.9	-0.53	-0.56	0.53	0.55	0.19	10	0.1	0.9	-0.19	-0.21	0.23	0.26	0.23	10	0.1	0.9	-0.10	-0.12	0.15	0.18	0.19
30	0.1	0	0.00	0.00	0.12	0.15	0.20	30	0.1	0	0.01	0.01	0.05	0.07	0.09	30	0.1	0	0.01	0.01	0.04	0.05	0.06
30	0.1	0.5	-0.21	-0.22	0.22	0.24	0.16	30	0.1	0.5	-0.13	-0.14	0.14	0.16	0.11	30	0.1	0.5	-0.08	-0.09	0.10	0.11	0.10
30	0.1	0.9	-0.70	-0.70	0.70	0.70	0.07	30	0.1	0.9	-0.44	-0.45	0.44	0.45	0.12	30	0.1	0.9	-0.29	-0.30	0.29	0.31	0.13
5	0.3	0	0.01	0.00	0.12	0.15	0.20	5	0.3	0	0.00	0.00	0.05	0.06	0.09	5	0.3	0	0.00	0.00	0.04	0.04	0.06
5	0.3	0.5	-0.03	-0.06	0.16	0.20	0.25	5	0.3	0.5	0.00	-0.01	0.07	0.09	0.12	5	0.3	0.5	0.00	-0.01	0.05	0.06	0.09
5	0.3	0.9	-0.10	-0.15	0.24	0.29	0.32	5	0.3	0.9	-0.02	-0.03	0.11	0.13	0.17	5	0.3	0.9	-0.01	-0.02	0.08	0.10	0.12
10	0.3	0	0.01	0.00	0.12	0.15	0.20	10	0.3	0	0.00	0.00	0.05	0.06	0.09	10	0.3	0	0.00	0.00	0.04	0.04	0.06
10	0.3	0.5	-0.06	-0.08	0.15	0.19	0.22	10	0.3	0.5	-0.01	-0.02	0.07	0.09	0.12	10	0.3	0.5	-0.01	-0.01	0.05	0.07	0.09
10	0.3	0.9	-0.23	-0.26	0.27	0.31	0.25	10	0.3	0.9	-0.06	-0.06	0.11	0.14	0.17	10	0.3	0.9	-0.03	-0.04	0.08	0.10	0.12
30	0.3	0	0.00	-0.01	0.11	0.14	0.19	30	0.3	0	0.00	0.00	0.05	0.06	0.08	30	0.3	0	0.00	0.00	0.04	0.04	0.06
30	0.3	0.5	-0.15	-0.16	0.17	0.20	0.18	30	0.3	0.5	-0.05	-0.06	0.08	0.09	0.10	30	0.3	0.5	-0.03	-0.03	0.05	0.07	0.08
30	0.3	0.9	-0.47	-0.48	0.47	0.48	0.12	30	0.3	0.9	-0.17	-0.18	0.18	0.20	0.13	30	0.3	0.9	-0.09	-0.10	0.11	0.12	0.11

Note: Results are based on 5000 Monte Carlo runs. True σ^2 *was set equal to 1.*

6 Conclusions

We derive second order approximations for the bias and MSE of 2SLS (and the Nagar estimator) with both valid and invalid instruments. The derivation for invalid instruments is new, to the best of our knowledge. We find that substantial finite sample bias can occur when weak instruments exist which arises when the R^2 of the reduced form regression is low, the number of instruments is high, or the correlation between the structural and reduced form stochastic terms ρ is high.

We then compare the bias and MSE of 2SLS with OLS. The OLS estimator is biased and inconsistent, but its smaller variance may make it preferable to 2SLS in a weak instruments situation. We determine straightforward and easily checked conditions under which 2SLS has smaller bias than OLS. These bias conditions carry over, in large part, to the MSE comparisons because changes in the bias term are quite important in changes in the MSE term given typical sample sizes of $n = 100$ or larger. We find that 2SLS is generally the preferred estimator. However, the econometrician can use our formulae to check the expected performance of 2SLS and OLS in a given situation given some *a priori* knowledge about likely parameter values.

We also find that the estimated standard errors for the 2SLS estimator are downward biased, sometimes leading to the mistaken inference that the 2SLS estimate are much more precise than they actually are. Such bias explains why the actual size of the often used test of over identifying restrictions (OID test) is considerably larger than the nominal size. ■

References

ANDREWS D.W.K., MOREIRA M.J. and STOCK J.H. (2004). – « Optimal Invariant Similar Tests for Instrumental Variables Regression », *unpublished manuscript*.

ANGRIST J. and KRUEGER A. (1991). – « Does Compulsory School Attendance Affect Schooling and Earnings », *Quarterly Journal of Economics*, 106, pp. 979-1014.

BEKKER P.A. (1994). – « Alternative Approximations to the Distributions of Instrumental Variable Estimators », *Econometrica*, 92, pp. 657-681.

CARD D. (2001). – « Estimating the Return to Schooling », *Econometrica*, pp. 1127-1152.

GRILICHES Z. (1957). – « Specification Bias in Estimates of Production Function », *Journal of Farm Economics*, 38, pp. 8-20.

GRILICHES Z. (1977). – « Estimating the Returns to Schooling: Some Econometric Problems », *Econometrica*, 45, pp. 1-22.

HAHN J. and HAUSMAN J. (2002a). – « A New Specification Test for the Validity of Instrumental Variables », *Econometrica*, 70, pp. 163-189.

HAHN J. and HAUSMAN J. (2002b). – « Notes on Bias in Estimators for Simultaneous Equation Models », *Economics Letters*, 75, pp. 237-241.

HAHN J. and HAUSMAN J. (2003). – « Weak Instruments: Diagnosis and Cures in Empirical Econometrics », *American Economic Review*.

HAHN J., HAUSMAN J., KUIERSTEINER G. (2004). – « Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations », *Econometrics Journal*, 7, pp. 272-306.

HAUSMAN J.A. (1978). – « Specification Tests in Econometrics », *Econometrica*, 46, pp. 1251-1271.

HAUSMAN J.A. (1983). – « Specification and Estimation of Simultaneous Equation Models », in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Vol. 1, Amsterdam: North Holland.

HAUSMAN J. (2001). – « Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left », *Journal of Economic Perspectives*, 2001.

IMBENS G. (2003). – « Sensivity to Exogeneity Assumptions in Program Evaluation », *American Economic Review*, 93, pp. 126-132.

KLEIBERGEN F. (2002). – « Pivotal Statistics for Testing Structural Parameters in IV Regression », *Econometrica*, 70, pp. 1781-1803.

MANSKI C. (1990). – « Nonparametric Bounds on Treatment Effects », *American Economic Review Papers and Proceedings*, 80, pp. 319-323.

MANSKI C. (2003). – « Partial Identification of Probability Distributions », New York: *Springer-Verlag*.

ROTHENBERG T.J. (1983). – « Asymptotic Properties of Some Estimators in Structural Models », in *Studies in Econometrics, Time Series, and Multivariate Statistics*.

STAIGER D., STOCK J.H. (1997). – « IV Regression with Weak Instruments », *Econometrica*, 65, pp. 557-586.

STOCK J.H., WRIGHT J., YOGO M. (2002). – « A Survey of Weak Instruments and Weak Identification in GMM », *Journal of Business and Economic Statistics*, 20, pp. 518-529.

APPENDIX

Bekker asymptotic distribution of 2SLS, OLS, and nagar under misspecification

Suppose that

$$y_{1i}^{*} = y_{2i}\beta + e_1 = \left(z_i'\pi_2\right)\beta + u_{1i}$$
$$y_{2i} = z_i'\pi_2 + \nu_{2i}$$

where

$$\begin{pmatrix} u_{1i} \\ \nu_{2i} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \omega_{1,1} & \omega_{1,2} \\ \omega_{1,2} & \omega_{2,2} \end{bmatrix}\right)$$

Following is the Lemma reproduced from HAHN and HAUSMAN [2001]:

LEMMA: *Let* $U \equiv \left[y_1^* \quad y_2\right]$. *Assume that* $\frac{K}{n} \to \alpha + o\left(n^{-1/2}\right)$, *and that* $\pi_2' z' z \pi_2 / n$ *is fixed* at Θ. *Let* $\overline{S} \equiv U' P_z U$ *and* $S^{\perp} \equiv U' M_z U$. *We then have*

$$\sqrt{n}\left(\begin{pmatrix} n^{-1}\overline{S}_{11} \\ n^{-1}\overline{S}_{12} \\ n^{-1}\overline{S}_{22} \\ n^{-1}S_{11}^{\perp} \\ n^{-1}S_{12}^{\perp} \\ n^{-1}S_{22}^{\perp} \end{pmatrix} - \begin{pmatrix} \Theta\cdot\beta^2 + \frac{K}{n}\cdot\omega_{1,1} \\ \Theta\cdot\beta + \frac{K}{n}\cdot\omega_{1,2} \\ \Theta + \frac{K}{n}\cdot\omega_{2,2} \\ \left(1-\frac{K}{n}\right)\cdot\omega_{1,1} \\ \left(1-\frac{K}{n}\right)\cdot\omega_{1,2} \\ \left(1-\frac{K}{n}\right)\cdot\omega_{2,2} \end{pmatrix}\right) \Rightarrow \mathcal{N}\left(0, \begin{bmatrix} \overline{\Lambda} & 0 \\ 0 & \Lambda^{\perp} \end{bmatrix}\right),$$

where $\overline{\Lambda}$ *and* $\Lambda^{\perp}$ *denote symmetric* 3×3 *matrices such that*

$$\overline{\Lambda}_{1,1} = 4\omega_{1,1}\Theta\beta^2 + 2\alpha\omega_{1,1}^2$$
$$\overline{\Lambda}_{1,2} = 2\omega_{1,1}\Theta\beta + 2\beta^2\Theta\omega_{1,2} + 2\alpha\omega_{1,1}\omega_{1,2}$$
$$\overline{\Lambda}_{1,3} = 4\beta\Theta\omega_{1,2} + 2\alpha\omega_{1,2}^2$$
$$\overline{\Lambda}_{2,2} = \omega_{1,1}\Theta + \beta^2\Theta\omega_{2,2} + 2\Theta\omega_{1,2}\beta + \alpha\omega_{1,1}\omega_{2,2} + \alpha\omega_{1,2}^2$$
$$\overline{\Lambda}_{2,3} = 2\omega_{2,2}\Theta\beta + 2\Theta\omega_{1,2} + 2\alpha\omega_{2,2}\omega_{1,2}$$
$$\overline{\Lambda}_{3,3} = 4\omega_{2,2}\Theta + 2\alpha\omega_{2,2}^2$$

and

$$\Lambda_{1,1}^{\perp} = 2(1-\alpha)\omega_{1,1}^{2}$$
$$\Lambda_{1,2}^{\perp} = 2(1-\alpha)\omega_{1,1}\omega_{1,2}$$
$$\Lambda_{1,3}^{\perp} = 2(1-\alpha)\omega_{1,2}^{2}$$
$$\Lambda_{2,2}^{\perp} = (1-\alpha)\omega_{1,1}\omega_{2,2} + (1-\alpha)\omega_{1,2}^{2}$$
$$\Lambda_{2,3}^{\perp} = 2(1-\alpha)\omega_{2,2}\omega_{1,2}$$
$$\Lambda_{3,3}^{\perp} = 2(1-\alpha)\omega_{2,2}^{2}$$

Remark: The ωs in the Lemma correspond to the "reduced form". It would be convenient to rewrite the above with structural form parameters. Because

$$u = \varepsilon + \beta\nu$$

we can see that

$$\omega_{1,1} = \sigma_{1,1} + 2\beta\sigma_{1,2} + \beta^2\sigma_{2,2}$$
$$\omega_{1,2} = \sigma_{1,2} + \beta\sigma_{2,2}$$
$$\omega_{2,2} = \sigma_{2,2}$$

Lemma: *Suppose that* $\frac{K}{\sqrt{n}} = \mu + o(1)$. *Then we have*

$$\sqrt{n}\left(\frac{y_2' P_z y_1^*}{y_2' P_z y_2} - \frac{\Theta\cdot\beta}{\Theta}\right) \Rightarrow N\left(\frac{\mu\sigma_{1,2}}{\Theta}, V_{2SLS}\right)$$
$$\sqrt{n}\left(\frac{y_2' P_z y_1^* - \frac{K}{n} y_2' M_z y_1^*}{y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2} - \beta\right) \Rightarrow N(0, V_{2SLS})$$
$$\sqrt{n}\left(\frac{y_2' y_1^*}{y_2' y_2} - \left(\beta + \frac{\sigma_{1,2}}{\Theta + \omega_{2,2}}\right)\right) \Rightarrow N(0, V_{OLS})$$

Proof: Suppose that α = 0. Using the previous Lemma, we obtain

$$\sqrt{n}\left(\begin{pmatrix} n^{-1}\overline{S}_{12} \\ n^{-1}\overline{S}_{22} \end{pmatrix} - \begin{pmatrix} \Theta\cdot\beta + \frac{K}{n}\cdot\omega_{1,2} \\ \Theta + \frac{K}{n}\cdot\omega_{2,2} \end{pmatrix}\right) \Rightarrow \mathcal{N}\left(0, \begin{bmatrix} (\beta^2\omega_{2,2} + 2\omega_{1,2}\beta + \omega_{1,1})\Theta & 2(\omega_{2,2}\beta + \omega_{1,2})\Theta \\ 2(\omega_{2,2}\beta + \omega_{1,2})\Theta & 4\omega_{2,2}\Theta \end{bmatrix}\right)$$
$$\sqrt{n}\left(\begin{pmatrix} n^{-1}\left(\overline{S}_{12} - \frac{K}{n} S_{12}^{\perp}\right) \\ n^{-1}\left(\overline{S}_{22} - \frac{K}{n} S_{22}^{\perp}\right) \end{pmatrix} - \begin{pmatrix} \Theta\cdot\beta \\ \Theta \end{pmatrix}\right) \Rightarrow \mathcal{N}\left(0, \begin{bmatrix} (\beta^2\omega_{2,2} + 2\omega_{1,2}\beta + \omega_{1,1})\Theta & 2(\omega_{2,2}\beta + \omega_{1,2})\Theta \\ 2(\omega_{2,2}\beta + \omega_{1,2})\Theta & 4\omega_{2,2}\Theta \end{bmatrix}\right)$$

and

$$\sqrt{n}\left(\begin{pmatrix} n^{-1}\left(\overline{S}_{12}+S_{12}^{\perp}\right) \\ n^{-1}\left(\overline{S}_{22}+S_{22}^{\perp}\right)\end{pmatrix} - \begin{pmatrix}\Theta\cdot\beta+\omega_{1,2} \\ \Theta+\omega_{2,2}\end{pmatrix}\right)$$

$$\Rightarrow \mathcal{N}\left(0,\begin{bmatrix}\left(\beta^2\omega_{2,2}+2\omega_{1,2}\beta+\omega_{1,1}\right)\Theta+\omega_{1,2}^2+\omega_{1,1}\omega_{2,2} & 2\left(\omega_{2,2}\beta+\omega_{1,2}\right)\Theta+2\omega_{2,2}\omega_{1,2} \\ 2\left(\omega_{2,2}\beta+\omega_{1,2}\right)\Theta+2\omega_{2,2}\omega_{1,2} & 4\omega_{2,2}\Theta+2\omega_{2,2}^2\end{bmatrix}\right)$$

Therefore, using Delta method, we obtain the following:

$$\sqrt{n}\left(\frac{y_2'P_zy_1^*}{y_2'P_zy_2}-\frac{\Theta\cdot\beta+\frac{K}{n}\cdot\omega_{1,2}}{\Theta+\frac{K}{n}\cdot\omega_{2,2}}\right)\Rightarrow N\left(0,\frac{\sigma_{1,1}}{\Theta}\right)$$

$$\sqrt{n}\left(\frac{y_2'P_zy_1^*-\frac{K}{n}y_2'M_zy_1^*}{y_2'P_zy_2-\frac{K}{n}y_2'M_zy_2}-\frac{\Theta\cdot\beta}{\Theta}\right)\Rightarrow N\left(0,\frac{\sigma_{1,1}}{\Theta}\right)$$

$$\sqrt{n}\left(\frac{y_2'y_1^*}{y_2'y_2}-\frac{\Theta\cdot\beta+\omega_{1,2}}{\Theta+\omega_{2,2}}\right)\Rightarrow N\left(0,\frac{\sigma_{1,1}}{\left(\Theta+\omega_{2,2}\right)}-\frac{\sigma_{1,2}^2}{\left(\Theta+\omega_{2,2}\right)^2}-2\frac{\sigma_{1,2}^2\Theta^2}{\left(\Theta+\omega_{2,2}\right)^4}\right)$$

where we used the fact that

$$\begin{aligned}\omega_{1,1}&=\sigma_{1,1}+2\beta\sigma_{1,2}+\beta^2\sigma_{2,2}\\ \omega_{1,2}&=\sigma_{1,2}+\beta\sigma_{2,2}\\ \omega_{2,2}&=\sigma_{2,2}\end{aligned}$$

Because $\frac{K}{\sqrt{n}}=\mu+o(1)$, we can see that

$$\sqrt{n}\left(\frac{\Theta\cdot\beta+\frac{K}{n}\cdot\omega_{1,2}}{\Theta+\frac{K}{n}\cdot\omega_{2,2}}-\frac{\Theta\cdot\beta}{\Theta}\right)=\frac{\frac{K}{\sqrt{n}}\sigma_{1,2}}{\Theta+\frac{1}{\sqrt{n}}\frac{K}{\sqrt{n}}\sigma_{2,2}}=\frac{\mu\sigma_{1,2}}{\Theta}+o(1)$$

and

$$\begin{aligned}\sqrt{n}\left(\frac{y_2'P_zy_1^*}{y_2'P_zy_2}-\frac{\Theta\cdot\beta}{\Theta}\right)&=\sqrt{n}\left(\frac{y_2'P_zy_1^*}{y_2'P_zy_2}-\frac{\Theta\cdot\beta+\frac{K}{n}\cdot\omega_{1,2}}{\Theta+\frac{K}{n}\cdot\omega_{2,2}}\right)+\sqrt{n}\left(\frac{\Theta\cdot\beta+\frac{K}{n}\cdot\omega_{1,2}}{\Theta+\frac{K}{n}\cdot\omega_{2,2}}-\frac{\Theta\cdot\beta}{\Theta}\right)\\ &=\sqrt{n}\left(\frac{y_2'P_zy_1^*}{y_2'P_zy_2}-\frac{\Theta\cdot\beta+\frac{K}{n}\cdot\omega_{1,2}}{\Theta+\frac{K}{n}\cdot\omega_{2,2}}\right)+\frac{\mu\sigma_{1,2}}{\Theta}+o(1)\\ &\Rightarrow N\left(\frac{\mu\sigma_{1,2}}{\Theta},\frac{\sigma_{1,1}}{\Theta}\right)\end{aligned}$$

Asymptotic distribution of 2SLS under misspecification

Note that

$$
\begin{aligned}
b_{2SLS} &= \frac{y_2' P_z y_1}{y_2' P_z y_2} = \frac{y_2' P_z \left(y_1^* + \frac{1}{\sqrt{n}} z\gamma \right)}{y_2' P_z y_2} \\
&= \frac{y_2' P_z y_1^*}{y_2' P_z y_2} + \frac{1}{\sqrt{n}} \frac{n^{-1} \left(z\pi_2 + \nu_2 \right)' z\gamma}{n^{-1} y_2' P_z y_2}
\end{aligned}
$$

But

$$
\begin{aligned}
n^{-1} \left(z\pi_2 + \nu_2 \right)' z\gamma &= \Xi + n^{-1} \nu_2' z\gamma = \Xi + o_p(1) \\
n^{-1} y_2' P_z y_2 &= n^{-1} \overline{S}_{22} = \Theta + o_p(1)
\end{aligned}
$$

so that

$$
\frac{n^{-1} \left(z\pi_2 + \nu_2 \right)' z\gamma}{n^{-1} y_2' P_z y_2} = \frac{\Xi}{\Theta} + o_p(1)
$$

It follows that

$$
\begin{aligned}
\sqrt{n} \left(b_{2SLS} - \beta \right) &= \sqrt{n} \left(\frac{y_2' P_z y_1^*}{y_2' P_z y_2} - \beta \right) + \frac{n^{-1} \left(z\pi_2 + \nu_2 \right)' z\gamma}{n^{-1} y_2' P_z y_2} \\
&= \sqrt{n} \left(\frac{y_2' P_z y_1^*}{y_2' P_z y_2} - \beta \right) + \frac{\Xi}{\Theta} + o_p(1) \\
&\Rightarrow N \left(\frac{\Xi + \mu\sigma_{1,2}}{\Theta}, V_{2SLS} \right)
\end{aligned}
$$

Asymptotic distribution of OLS under misspecification

Note that

$$
\begin{aligned}
b_{OLS} &= \frac{y_2' y_1}{y_2' y_2} = \frac{y_2' \left(y_1^* + \frac{1}{\sqrt{n}} z\gamma \right)}{y_2' y_2} \\
&= \frac{y_2' y_1^*}{y_2' y_2} + \frac{1}{\sqrt{n}} \frac{n^{-1} \left(z\pi_2 + \nu_2 \right)' z\gamma}{n^{-1} y_2' y_2}
\end{aligned}
$$

But

$$n^{-1}\left(z\pi_2+\nu_2\right)' z\gamma = \Xi + n^{-1}\nu_2' z\gamma = \Xi + o_p\left(1\right)$$
$$n^{-1}y_2' y_2 = n^{-1}\overline{S}_{22} + n^{-1}S_{22}^{\perp} = \Theta + \sigma_{2,2} + o_p\left(1\right)$$

so that

$$\frac{n^{-1}\left(z\pi_2+\nu_2\right)' z\gamma}{n^{-1}y_2' P_z y_2} = \frac{\Xi}{\Theta+\sigma_{2,2}} + o_p\left(1\right)$$

It follows that

$$\begin{aligned}\sqrt{n}\left(b_{OLS} - \left(\beta + \frac{\sigma_{1,2}}{\Theta+\omega_{2,2}}\right)\right) &= \sqrt{n}\left(\frac{y_2' y_1^*}{y_2' y_2} - \left(\beta + \frac{\sigma_{1,2}}{\Theta+\omega_{2,2}}\right)\right) + \frac{n^{-1}\left(z\pi_2+\nu_2\right)' z\gamma}{n^{-1}y_2' y_2} \\ &= \sqrt{n}\left(\frac{y_2' y_1^*}{y_2' y_2} - \left(\beta + \frac{\sigma_{1,2}}{\Theta+\omega_{2,2}}\right)\right) + \frac{\Xi}{\Theta+\sigma_{2,2}} + o_p\left(1\right) \\ &\Rightarrow N\left(\frac{\Xi}{\Theta+\sigma_{2,2}}, V_{OLS}\right)\end{aligned}$$

Asymptotic distribution of nagar under misspecification

Note that

$$\begin{aligned} b_N &= \frac{y_2' P_z y_1 - \frac{K}{n} y_2' M_z y_1}{y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2} = \frac{y_2' P_z\left(y_1^* + \frac{1}{\sqrt{n}} z\gamma\right) - \frac{K}{n} y_2' M_z\left(y_1^* + \frac{1}{\sqrt{n}} z\gamma\right)}{y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2} \\ &= \frac{y_2' P_z y_1^* - \frac{K}{n} y_2' M_z y_1^*}{y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2} + \frac{1}{\sqrt{n}} \frac{n^{-1}\left(z\pi_2+\nu_2\right)' z\gamma}{n^{-1}\left(y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2\right)}\end{aligned}$$

But

$$n^{-1}\left(z\pi_2+\nu_2\right)' z\gamma = \Xi + o_p\left(1\right)$$
$$n^{-1}\left(y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2\right) = \Theta + o_p\left(1\right)$$

so that

$$\frac{n^{-1}\left(z\pi_2+\nu_2\right)' z\gamma}{n^{-1}\left(y_2' P_z y_2 - \frac{K}{n} y_2' M_z y_2\right)} = \frac{\Xi}{\Theta} + o_p\left(1\right)$$

It follows that

$$\sqrt{n}\left(b_N-\beta\right)=\sqrt{n}\left(\frac{y_2'P_zy_1^*-\frac{K}{n}y_2'M_zy_1^*}{y_2'P_zy_2-\frac{K}{n}y_2'M_zy_2}-\beta\right)+\frac{n^{-1}\left(z\pi_2+\nu_2\right)'z\gamma}{n^{-1}\left(y_2'P_zy_2-\frac{K}{n}y_2'M_zy_2\right)}$$

$$=\sqrt{n}\left(\frac{y_2'P_zy_1^*-\frac{K}{n}y_2'M_zy_1^*}{y_2'P_zy_2-\frac{K}{n}y_2'M_zy_2}-\beta\right)+\frac{\Xi}{\Theta}+o_p\left(1\right)$$

$$\Rightarrow N\left(\frac{\Xi}{\Theta},V_{2SLS}\right)$$

Sensitivity analysis

Consider a model with one endogenous regressor where other included exogenous variables are partialled out. The model takes the form where

$$y_i=x_i\beta+\varepsilon_i,\quad i=1,\dots,n.$$

Denote the available instrument as z_i, and write the first stage regression as

$$x_i=z_i'\pi+\nu_i.$$

2SLS estimator is obviously given by

$$\hat{\beta}_{2SLS}=\left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1}X'Z\left(Z'Z\right)^{-1}Z'Y,$$

where

$$X=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix},\quad Y=\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix},\quad Z=\begin{bmatrix}z_1'\\ \vdots\\ z_n'\end{bmatrix}.$$

What is the property of *b* if the exclusion restriction is in fact violated? In order to implement violation exclusion restriction, we add a little noise to ε_i, and consider a new model

$$y_i^*\left(\theta\right)=x_i\beta+z_i'\theta+\varepsilon_i$$

where

$$\varepsilon_i^*=z_i'\theta+\varepsilon_i$$

Let

$$\hat{\beta}^*_{2SLS}(\theta) = \left[X'Z(Z'Z)^{-1} Z'X \right]^{-1} X'Z(Z'Z)^{-1} Z'Y^*(\theta)$$

and

$$b_{2SLS}(\theta) \equiv \text{plim}\, \hat{\beta}^*_{2SLS}(\theta) - \beta$$

We would like to examine the maximal asymptotic bias $|b_{2SLS}(\theta)|$ for a small violation of exclusion restriction, i.e., the violation such that the correlation between $z_i'\theta$ and ε_i^* is some small number ψ. We argue that

$$\sqrt{\frac{n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n^{-1}\sum_{i=1}^{n} x_i^2} \frac{1}{\widehat{\mathbb{R}}_f^2}} \sqrt{\frac{\psi^2}{1-\psi^2}}$$

provides such measure of sensitivity. Here, $\widehat{\mathbb{R}}_f^2$ denotes the R^2 in the first stage.

It can be shown that

$$b_{2SLS}(\theta) = (\pi'\Phi\pi)^{-1} \pi'\Phi\theta$$

where

$$\Phi = \text{plim}\, n^{-1} Z'Z$$

Note that

$$\frac{\partial b_{2SLS}(\theta)}{\partial \theta'} = (\pi'\Phi\pi)^{-1} \pi'\Phi$$

which is maximized when $\theta \propto \pi$. We therefore focus on the type of violation such that $\theta = \xi \bullet \pi$ for some scalar ξ. Without loss of generality, we will write

$$b_{2SLS}(\theta) = b_{2SLS}(\xi \cdot \pi) = b_{2SLS}(\xi)$$

Note that the population R^2 in the regression of ε^* on z, which is equal to the square of the correlation ψ between ε_i^* and $z_i'\pi$, is equal to

$$\psi^2 = \frac{\theta'\Phi\theta}{\theta'\Phi\theta + E\left[\varepsilon_i^2\right]} = \frac{\xi^2 \cdot \pi'\Phi\pi}{\xi^2 \cdot \pi'\Phi\pi + E\left[\varepsilon_i^2\right]}$$

and

$$b_{2SLS}(\xi) = (\pi'\Phi\pi)^{-1}\pi'\Phi(\xi\cdot\pi) = \xi$$

We can solve ξ^2 as a function of ψ^2 ,and obtain

$$\xi^2 = \frac{E[\varepsilon_i^2]}{\pi'\Phi\pi}\frac{\psi^2}{1-\psi^2}$$

Now, note that the population R^2 in the first stage $\mathbb{R}_f^2$ is equal to

$$\mathbb{R}_f^2 = \frac{\pi'\Phi\pi}{E[x_i^2]}$$

which can be solved for $\pi'\Phi\pi$ as

$$\pi'\Phi\pi = \mathbb{R}_f^2 \cdot E[x_i^2]$$

We therefore obtain

$$\xi^2 = \frac{E[\varepsilon_i^2]}{E[x_i^2]}\frac{1}{\mathbb{R}_f^2}\frac{\psi^2}{1-\psi^2}$$

or

$$|\xi| = |b_{2SLS}(\xi)| = \sqrt{\frac{E[\varepsilon_i^2]}{E[x_i^2]}\frac{1}{\mathbb{R}_f^2}}\sqrt{\frac{\psi^2}{1-\psi^2}}$$

We note that $|\xi|$ can be approximated by the empirical counterpart

$$|\xi| = |b_{2SLS}(\xi)| \approx \sqrt{\frac{n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n^{-1}\sum_{i=1}^{n}x_i^2}\frac{1}{\widehat{\mathbb{R}}_f^2}}\sqrt{\frac{\psi^2}{1-\psi^2}}$$

Digression: robustness of 2SLS

In general, we estimate β by

$$\hat{\beta}_A = \left[(ZA)'X\right]^{-1}(ZA)'Y$$

and the counterpart under small misspecification is

$$\hat{\beta}_A^*(\theta) = \left[(ZA)' X\right]^{-1} (ZA)' Y^*(\theta)$$

so that

$$\begin{aligned}
b_A(\theta) &\equiv \text{plim}\,\hat{\beta}_A^*(\theta) - \beta \\
&= \text{plim}\left[(ZA)' X\right]^{-1} (ZA)' Y^*(\theta) - \beta \\
&= \text{plim}\left[(ZA)' X\right]^{-1} (ZA)' (X\beta + Z\theta + \varepsilon) - \beta \\
&= \beta + \text{plim}\left[(ZA)' X\right]^{-1} (ZA)' Z\theta - \beta \\
&= \text{plim}\left[A' Z'X\right]^{-1} A' Z'Z\theta \\
&= \text{plim}\left[A' Z'Z(Z'Z)^{-1} Z'X\right]^{-1} A' Z'Z\theta \\
&= (A'\Phi\pi)^{-1} A'\Phi\theta
\end{aligned}$$

Note that

$$\frac{\partial b_{2SLS}(\theta)}{\partial \theta'} = (\pi'\Phi\pi)^{-1} \pi'\Phi$$

and

$$\frac{\partial b_A(\theta)}{\partial \theta'} = (A'\Phi\pi)^{-1} A'\Phi$$

Instead of dealing with a normalization involving the weight matrix Φ, it is convenient to use assume that $\Phi = I$. We then have

$$\frac{\partial b_{2SLS}(\theta)}{\partial \theta'} = (\pi'\pi)^{-1} \pi'$$

and

$$\frac{\partial b_A(\theta)}{\partial \theta'} = (A'\pi)^{-1} A'$$

Remark: If there is only one instrument, then $\frac{\partial b_{2SLS}(\theta)}{\partial \theta'} = \pi^{-1}$. Therefore, small π indicates that 2SLS is sensitive to misspecification.

Remark: If there are multiple components in π, and if the first component of π is small relative to other components of π, then $\frac{\partial b_{2SLS}(\theta)}{\partial\theta_1}$ would be small, i.e., 2SLS is not very sensitive to the violation of the exclusion restriction in $z_{i,1}$.

Remark: Note that

$$\left\| \frac{\partial b_{2SLS}(\theta)}{\partial\theta'} \right\|^2 = (\pi'\pi)^{-1} = \frac{1}{\|\pi\|^2}$$

and

$$\left\| \frac{\partial b_A(\theta)}{\partial\theta'} \right\|^2 = (A'\pi)^{-1} A'A(A'\pi)^{-1} = \frac{\|A\|^2}{(A'\pi)^2} \geq \frac{\|A\|^2}{\|A\|^2\|\pi\|^2} = \frac{1}{\|\pi\|^2} = \left\| \frac{\partial b_{2SLS}(\theta)}{\partial\theta'} \right\|^2$$

Therefore, 2SLS is the most robust estimator among the class of IV estimators b_A.

Higher order bias of $\hat{\sigma}^2$

Our model is given by

$$\begin{aligned} y_i &= x_i\beta + \varepsilon_i, \\ x_i &= f_i + u_i \equiv z_i'\pi + u_i \qquad i = 1,\ldots,n \end{aligned}$$

where $(\varepsilon_i, u_i)'$ is homoscedastic and normal. We consider the 2SLS

$$\hat{\beta}_{2SLS} \equiv \frac{x'Py}{x'Px}$$

and the related estimator for the variance of ε_i:

$$\hat{\sigma}^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i\hat{\beta}_{2SLS}\right)^2$$

We have the following characterization of $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i - x_i\left(\hat{\beta}_{2SLS} - \beta\right)\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i x_i\right)\left(\hat{\beta}_{2SLS} - \beta\right) + \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)\left(\hat{\beta}_{2SLS} - \beta\right)^2$$
$$= \frac{\varepsilon'\varepsilon}{n}$$
$$-2\left(\frac{\varepsilon' u}{n} + \frac{f'\varepsilon}{n}\right)\left(\hat{\beta}_{2SLS} - \beta\right)$$
$$+\left(H + 2\frac{f'u}{n} + \frac{u'u}{n}\right)\left(\hat{\beta}_{2SLS} - \beta\right)^2$$

where

$$H \equiv \frac{1}{n}f'f = \frac{1}{n}\pi' Z' Z\pi$$

LEMMA:

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta\right) = \frac{1}{H}\sum_{j=1}^{7}T_j + o_p\left(\frac{1}{n}\right)$$

for

$$T_1 = \frac{1}{\sqrt{n}}f'\varepsilon = O_p(1)$$
$$T_2 = \frac{u'P\varepsilon}{\sqrt{n}} = O_p\left(\frac{1}{\sqrt{n}}\right)$$
$$T_3 = -2\left(\frac{f'u}{n}\right)\frac{1}{H}\left(\frac{1}{\sqrt{n}}f'\varepsilon\right) = O_p\left(\frac{1}{\sqrt{n}}\right)$$
$$T_4 = 0$$
$$T_5 = -\left(\frac{u'Pu}{n}\right)\frac{1}{H}\left(\frac{1}{\sqrt{n}}f'\varepsilon\right) = O_p\left(\frac{1}{n}\right)$$
$$T_6 = -2\left(\frac{f'u}{n}\right)\frac{1}{H}\left(\frac{u'P\varepsilon}{\sqrt{n}}\right) = O_p\left(\frac{1}{n}\right)$$
$$T_7 = 2^2\left(\frac{f'u}{n}\right)^2\frac{1}{H^2}\left(\frac{1}{\sqrt{n}}f'\varepsilon\right) = O_p\left(\frac{1}{n}\right)$$

PROOF: Note that 2SLS is a special case of the *k*-class estimator

$$\hat{\beta}_k = \frac{x'Py - k \cdot x'My}{x'Px - k \cdot x'Mx}$$

for

$$k = \frac{a\theta + \frac{b}{n}}{1 - a\theta - \frac{b}{n}}$$

and θ is the "eigenvalue". Note that 2SLS corresponds to $a = 0$ and $b = 0$. The result follows from DONALD and NEWEY [1998].
We therefore obtain

Lemma:

$$\begin{aligned}\hat{\sigma}^2 &= \sigma_\varepsilon^2 \\ &+ \frac{1}{\sqrt{n}}\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n} - \sigma_\varepsilon^2\right) - \frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_1\right) \\ &- \frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_2 + \frac{1}{H}T_3\right) - \frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n} - \sigma_{\varepsilon u}\right)\left(\frac{1}{H}T_1\right) - \frac{1}{n}\frac{T_1^2}{H} + \frac{1}{n}\frac{\sigma_u^2 T_1^2}{H^2} \\ &+ o_p\left(\frac{1}{n}\right)\end{aligned}$$

PROOF: We have

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\varepsilon'\varepsilon}{n} \\ &- \frac{2}{\sqrt{n}}\left(\frac{\varepsilon' u}{n} + \frac{f'\varepsilon}{n}\right)\left(\frac{1}{H}\sum_{j=1}^{7} T_j\right) \\ &+ \frac{1}{n}\left(H + 2\frac{f'u}{n} + \frac{u'u}{n}\right)\left(\frac{1}{H}\sum_{j=1}^{7} T_j\right)^2\end{aligned}$$

Because

$$T_1 = O_p(1)$$

$$T_2 = O_p\left(\frac{1}{\sqrt{n}}\right), \qquad T_3 = O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$T_4 = 0, \qquad T_5 = O_p\left(\frac{1}{n}\right), \qquad T_6 = O_p\left(\frac{1}{n}\right), \qquad T_7 = O_p\left(\frac{1}{n}\right)$$

and

$$\frac{\varepsilon' u}{n} = O_p(1), \qquad \frac{f'\varepsilon}{n} = \frac{1}{\sqrt{n}} T_1 = O_p\left(\frac{1}{\sqrt{n}}\right)$$
$$\frac{f'u}{n} = O_p\left(\frac{1}{\sqrt{n}}\right), \qquad \frac{u'u}{n} = O_p(1)$$

we obtain

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\varepsilon'\varepsilon}{n} \\ &\quad - \frac{2}{\sqrt{n}}\left(\frac{\varepsilon' u}{n}\right)\left(\frac{1}{H}\sum_{j=1}^{3} T_j\right) - \frac{2}{\sqrt{n}}\left(\frac{1}{\sqrt{n}} T_1\right)\left(\frac{1}{H} T_1\right) \\ &\quad + \frac{1}{n}\left(H + \frac{u'u}{n}\right)\left(\frac{1}{H} T_1\right)^2 + o_p\left(\frac{1}{n}\right)\end{aligned}$$

Now, note that

$$\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n} - \sigma_\varepsilon^2\right) = O_p(1), \qquad \sqrt{n}\left(\frac{\varepsilon' u}{n} - \sigma_{\varepsilon u}\right) = O_p(1), \qquad \sqrt{n}\left(\frac{u'u}{n} - \sigma_u^2\right) = O_p(1)$$

We therefore obtain

$$\frac{\varepsilon'\varepsilon}{n} = \sigma_\varepsilon^2 + \frac{1}{\sqrt{n}}\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n} - \sigma_\varepsilon^2\right),$$

$$\begin{aligned}\frac{2}{\sqrt{n}}\left(\frac{\varepsilon' u}{n}\right)\left(\frac{1}{H}\sum_{j=1}^{3} T_j\right) &= \frac{2}{\sqrt{n}}\left(\sigma_{\varepsilon u} + \left(\frac{\varepsilon' u}{n} - \sigma_{\varepsilon u}\right)\right)\left(\frac{1}{H}\sum_{j=1}^{3} T_j\right) \\ &= \frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}\sum_{j=1}^{3} T_j\right) + \frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n} - \sigma_{\varepsilon u}\right)\left(\frac{1}{H} T_1\right) + o_p\left(\frac{1}{n}\right),\end{aligned}$$

and

$$\frac{1}{n}\left(H + \frac{u'u}{n}\right)\left(\frac{1}{H} T_1\right)^2 = \frac{1}{n}\frac{1}{H} T_1^2 + \frac{\sigma_u^2}{n}\frac{1}{H^2} T_1^2 + o_p\left(\frac{1}{n}\right)$$

It follows that

$$\begin{aligned}\hat{\sigma}^2 &= \sigma_\varepsilon^2 + \frac{1}{\sqrt{n}}\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n} - \sigma_\varepsilon^2\right) \\ &\quad - \frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}\sum_{j=1}^{3} T_j\right) - \frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n} - \sigma_{\varepsilon u}\right)\left(\frac{1}{H} T_1\right) - \frac{2}{n}\frac{T_1^2}{H} \\ &\quad + \frac{1}{n}\frac{T_1^2}{H} + \frac{1}{n}\frac{\sigma_u^2 T_1^2}{H^2} + o_p\left(\frac{1}{n}\right)\end{aligned}$$

or

$$\begin{aligned}\hat{\sigma}^2 &= \sigma_\varepsilon^2 \\ &+\frac{1}{\sqrt{n}}\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n}-\sigma_\varepsilon^2\right)-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_1\right) \\ &-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_2+\frac{1}{H}T_3\right)-\frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n}-\sigma_{\varepsilon u}\right)\left(\frac{1}{H}T_1\right)-\frac{1}{n}\frac{T_1^2}{H}+\frac{1}{n}\frac{\sigma_u^2 T_1^2}{H^2} \\ &+o_p\left(\frac{1}{n}\right)\end{aligned}$$

Assume that we can ignore the $o_p\left(\frac{1}{n}\right)$ term in Lemma expansion in calculation of expectation. We then obtain.

$$E\left[\hat{\sigma}^2\right]\approx\sigma_\varepsilon^2-\frac{2}{n}\frac{(K-2)\sigma_{\varepsilon u}^2}{H}-\frac{1}{n}\sigma_\varepsilon^2+\frac{1}{n}\frac{\sigma_u^2\sigma_\varepsilon^2}{H}$$

where

$$H\equiv\frac{1}{n}f'f=\frac{1}{n}\pi' Z' Z\pi$$

This result can be proved in the following way. From the immediately preceding lemma, we have

$$\begin{aligned}\hat{\sigma}^2 &= \sigma_\varepsilon^2 \\ &+\frac{1}{\sqrt{n}}\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n}-\sigma_\varepsilon^2\right)-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_1\right) \\ &-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_2+\frac{1}{H}T_3\right)-\frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n}-\sigma_{\varepsilon u}\right)\left(\frac{1}{H}T_1\right)-\frac{1}{n}\frac{T_1^2}{H}+\frac{1}{n}\frac{\sigma_u^2 T_1^2}{H^2} \\ &+o_p\left(\frac{1}{n}\right)\end{aligned}$$

Because expected values of the $O_p\left(\frac{1}{\sqrt{n}}\right)$ terms in the second line are zero, it suffices to consider the $O_p\left(\frac{1}{n}\right)$ in the third line. First, we note that

$$E[T_2]=E\left[\frac{u' P\varepsilon}{\sqrt{n}}\right]=\frac{1}{\sqrt{n}}K\sigma_{\varepsilon u}$$

$$E[T_3]=E\left[-2\left(\frac{f'u}{n}\right)\frac{1}{H}\left(\frac{1}{\sqrt{n}}f'\varepsilon\right)\right]=-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}$$

from which we obtain

$$E\left[-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_2+\frac{1}{H}T_3\right)\right]=-\frac{2}{n}\frac{(K-2)\sigma_{\varepsilon u}^2}{H}$$

Second, we note that

$$E\left[-\frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n}-\sigma_{\varepsilon u}\right)\left(\frac{1}{H}T_1\right)\right]=0$$

due to symmetry. Third, we note that

$$E\left[T_1^2\right]=H\sigma_\varepsilon^2$$

from which we obtain

$$E\left[-\frac{1}{n}\frac{T_1^2}{H}+\frac{1}{n}\frac{\sigma_u^2 T_1^2}{H^2}\right]=-\frac{1}{n}\sigma_\varepsilon^2+\frac{1}{n}\frac{\sigma_u^2\sigma_\varepsilon^2}{H}$$

We therefore obtain

$$E\left[\hat{\sigma}^2\right]\simeq\sigma_\varepsilon^2-\frac{2}{n}\frac{(K-2)\sigma_{\varepsilon u}^2}{H}-\frac{1}{n}\sigma_\varepsilon^2+\frac{1}{n}\frac{\sigma_u^2\sigma_\varepsilon^2}{H}$$

Remark: In order to understand this result, imagine a counter-factual situation where the first order asymptotic approximation for $\sqrt{n}\left(\hat{\beta}_{2SLS}-\beta\right)$ is exact, i.e., write

$$\sqrt{n}\left(\hat{\beta}_{2SLS}-\beta\right)=\frac{1}{H}T_1$$

We would then have

$$\begin{aligned}\hat{\sigma}^2 &= \sigma_\varepsilon^2\\ &+\frac{1}{\sqrt{n}}\sqrt{n}\left(\frac{\varepsilon'\varepsilon}{n}-\sigma_\varepsilon^2\right)-\frac{2}{\sqrt{n}}\sigma_{\varepsilon u}\left(\frac{1}{H}T_1\right)\\ &-\frac{2}{n}\sqrt{n}\left(\frac{\varepsilon' u}{n}-\sigma_{\varepsilon u}\right)\left(\frac{1}{H}T_1\right)-\frac{1}{n}\frac{T_1^2}{H}+\frac{1}{n}\frac{\sigma_u^2T_1^2}{H^2}\\ &+o_p\left(\frac{1}{n}\right)\end{aligned}$$

and

$$E\left[\hat{\sigma}^2\right] \approx \sigma_\varepsilon^2 - \frac{1}{n}\sigma_\varepsilon^2 + \frac{1}{n}\frac{\sigma_u^2\sigma_\varepsilon^2}{H}$$

Therefore, our result implies that the approximate mean of $\hat{\sigma}^2$ *is smaller by*

$$\frac{2}{n}\frac{(K-2)\sigma_{\varepsilon u}^2}{H}$$

than would be expected out of first order asymptotic approximation.

Remark: our result can be understood from a different perspective. Note that the approximate bias of 2SLS is equal to

$$\frac{1}{\sqrt{n}}E\left[T_2 + T_3\right] = \frac{(K-2)\sigma_{\varepsilon u}}{nH}$$

Roughly speaking, 2SLS is biased toward OLS, which minimizes $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i b\right)^2$ with respect to b. If the 2SLS $\hat{\beta}_{2SLS}$ is close to the OLS $\hat{\beta}_{OLS}$, then we should expect

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i\hat{\beta}_{2SLS}\right)^2 \approx \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i\hat{\beta}_{OLS}\right)^2 \ll \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i\beta\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$$