Overview of Ideas in Weak Instrument Variable[*] 2024-08-19

jessekelighine.com
Jesse C. Chen 陳捷

Definition 1 (Instrument Variable (IV) Model)**.** Data is of the form $\{y_t, \mathbf{x}_t, \mathbf{z}_t, \mathbf{w}_t\}_{t=1}^{T}$ where y_t is a scalar outcome variable, $\mathbf{x}_t$ is a $N \times 1$ vector of potentially endogenous variable, $\mathbf{z}_t$ is a $K \times 1$ vector of IV, $\mathbf{w}_t$ is a $R \times 1$ vector of exogenous control variables. The model assumes the (matrix) form

$$\underset{(T\times 1)}{\mathbf{y}} = \underset{(T\times N)}{\mathbf{X}}\boldsymbol{\beta} + \underset{(T\times R)}{\mathbf{W}}\boldsymbol{\gamma} + \underset{(T\times 1)}{\mathbf{u}} \tag{1}$$

$$\underset{(T\times N)}{\mathbf{X}} = \underset{(T\times K)}{\mathbf{Z}}\boldsymbol{\Pi} + \underset{(T\times R)}{\mathbf{W}}\boldsymbol{\Phi} + \underset{(T\times N)}{\mathbf{v}} \tag{2}$$

$\mathbf{y} = (y_1, ..., y_T)'$, $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_T)'$, $\mathbf{W} = (\mathbf{w}_1, ..., \mathbf{w}_T)'$, $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_T)'$; $\mathbf{u} = (u_1, ..., u_T)'$ and $\mathbf{v} = (\mathbf{v}_1, ..., \mathbf{v}_T)'$ are error terms; $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Pi}$, and $\boldsymbol{\Phi}$ are parameters. Equation 1 is the so-called *structural equation* and Equation 2 is a linear projection, so-called *first stage*.

Remark 1. In most applications that appears on American Economic Review (AER), according to Andrews, Stock, and Sun (2019), there is only one endogenous variable, i.e., $N = 1$. We will focus on the case where $N = 1$ in this short note.

Remark 2. We do not focus on the control variables $\mathbf{w}_t$ in this note. If $\mathbf{z}_t$ is orthogonal to $\mathbf{v}_t$ only after controlling $\mathbf{w}_t$, we can simply invoke Frisch-Waugh-Lovell Theorem and redefined the problem so that $\mathbf{w}_t$ do not appear in the model.

1 A Primer: Concentration Parameter

Let us consider the simplest finite sample analysis possible, derived by Rothenberg (1984). This will provide us with some insight before diving into asymptotics.

Suppose that scalars $\mathbf{u}_t$ and $\mathbf{v}_t$ are drawn from independent and identically distributed (iid) bivariate normal distributions with $\mathbf{Var}(\mathbf{u}_t) = \sigma_{\mathbf{u}}^2$, $\mathbf{Var}(\mathbf{v}_t) = \sigma_{\mathbf{v}}^2$, and $\mathbf{Corr}(\mathbf{u}_t, \mathbf{v}_t) = \rho$. Also suppose that the instruments $\mathbf{z}_t$ are non-stochastic and that there are no controls $\mathbf{w}_t$. The Two Stage Least Square (2SLS) estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{\text{2SLS}} \coloneqq \frac{\mathbf{X}'\mathbf{P_Z}\mathbf{y}}{\mathbf{X}'\mathbf{P_Z}\mathbf{X}}$$

where $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the projection matrix of $\mathbf{Z}$. The 2SLS estimator can be rewritten as

$$\hat{\boldsymbol{\beta}}_{\text{2SLS}} = \frac{\mathbf{X}'\mathbf{P_Z}\mathbf{y}}{\mathbf{X}'\mathbf{P_Z}\mathbf{X}} = \boldsymbol{\beta}_0 + \frac{\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{u} + \mathbf{v}'\mathbf{P_Z}\mathbf{u}}{\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi} + 2\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{v} + \mathbf{v}'\mathbf{P_Z}\mathbf{v}}. \tag{3}$$

We want to standardize the terms that appears in the right-hand side. Define the following variables:

$$\zeta_{\mathbf{u}} \coloneqq \frac{\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{u}}{\sqrt{\sigma_{\mathbf{u}}^2\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi}}}, \quad \zeta_{\mathbf{v}} \coloneqq \frac{\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{v}}{\sqrt{\sigma_{\mathbf{v}}^2\boldsymbol{\Pi}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi}}}, \quad S_{\mathbf{vu}} \coloneqq \frac{\mathbf{v}'\mathbf{P_Z}\mathbf{u}}{\sigma_{\mathbf{v}}\sigma_{\mathbf{u}}}, \quad S_{\mathbf{vv}} \coloneqq \frac{\mathbf{v}'\mathbf{P_Z}\mathbf{v}}{\sigma_{\mathbf{v}}^2}.$$

[*] This note draws heavily from lecture note Shi (2012) and a talk given by Stock (2008). All references are given at the end of this note.

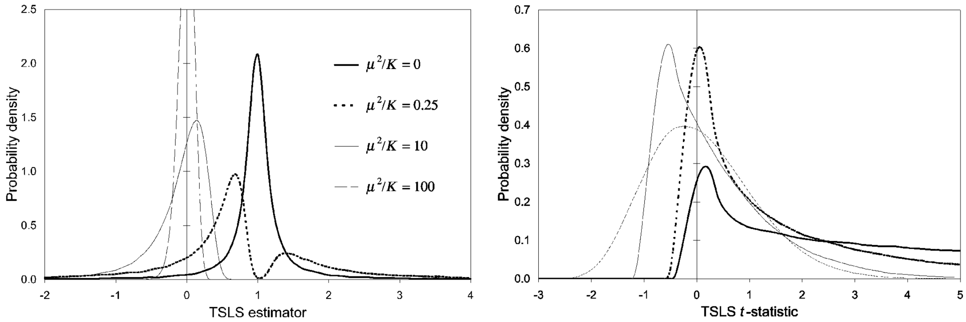

Figure 1: Finite Sample Distributions.

Source: Stock, Wright, and Yogo (2002). More detailed versions of these densities can be found in Nelson and Startz (1990a) and Nelson and Startz (1990b).

The first two random variables, $\zeta_{\mathbf{u}}$ and $\zeta_{\mathbf{v}}$, both follow standard normal distribution with correlation ρ. The random variable $S_{\mathbf{vu}}$ has mean ρ and variance $1+\rho^2$. Similarly, $S_{\mathbf{vv}}$ has mean 1 and variance 2. Note that all these four variables are independent of sample size. We rewrite (3) as

$$\mu(\hat{\boldsymbol{\beta}}_{2\text{SLS}} - \boldsymbol{\beta}) = \left(\frac{\sigma_{\mathbf{u}}}{\sigma_{\mathbf{v}}}\right) \frac{\zeta_{\mathbf{u}} + S_{\mathbf{vu}}/\mu}{1 + \zeta_{\mathbf{v}}/\mu + S_{\mathbf{vv}}/\mu^2} \tag{4}$$

where $\mu^2 := \mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi}/\sigma_{\mathbf{v}}^2$ is called the "concentration parameter."

Remark 3. Note that sample size only enters in the equation via μ^2 and not through other variables. When μ^2 is large, (4) behaves like a standard normal distribution. Hence, one can view μ^2 as the "effective" sample size. It is possible that the "effective" sample size μ^2 is small despite that the sample size T is large.

Remark 4. The concentration parameter μ^2 is closely related to the commonly used first-stage F-statistic. Recall that the first-stage F-statistic, used to test the linear hypothesis $\mathrm{H}_0 : \mathbf{\Pi} = 0$, is given by

$$F = \frac{\hat{\mathbf{\Pi}}'(\mathbf{Z}'\mathbf{Z})\hat{\mathbf{\Pi}}}{K\hat{\sigma}_{\mathbf{v}}^2} \tag{5}$$

where $\hat{\sigma}_{\mathbf{v}}^2$ is an estimator for $\sigma_{\mathbf{v}}^2$, commonly chosen to be $\frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{N-K}$. One immediately sees the connections between μ^2 and (5). Furthermore, if we can calculate the F-statistic with a *known* $\sigma_{\mathbf{v}}^2$, then we have

$$\mathbf{E}\left[\frac{\hat{\mathbf{\Pi}}'(\mathbf{Z}'\mathbf{Z})\hat{\mathbf{\Pi}}}{K\sigma_{\mathbf{v}}^2}\right] = 1 + \frac{\mu^2}{K}.$$

If the first-stage F-statistic is large, then μ^2/K is large, meaning that (4) would be approximately normal. Hence, our usual inference, i.e., t-test, about $\hat{\boldsymbol{\beta}}_{2\text{SLS}}$ can be carried through. (Stock, Wright, and Yogo 2002)

Remark 5. Combining Remark 3 and Remark 4, one can see that a small first-stage F-statistic *cannot* be fixed with more samples. However, in classic IV analysis, one assumes that $\mathbf{\Pi}$ is constant. Then, μ^2 tends to infinity as $T \to \infty$, regardless of how small $\mathbf{\Pi}$ actually is. This is why classical IV asymptotic fails when instruments are weak, i.e., the textbook asymptotic result provides a poor approximation of the finite sample distribution. This problem is really brought to attention of econometricians by numerical simulations by Nelson and Startz (1990a) and Nelson and Startz (1990b), in which they demonstrated that when the concentration parameter is small, then the finite sample distribution 2SLS estimator and t-statistic are highly irregular, as on can see in Figure 1.

2 Weak IV Asymptotics

Classic asymptotic results cannot capture the finite sample distribution of the 2SLS estimator and the test statistic. Staiger and Stock (1997) provided an analysis, now known as "weak IV asymptotics," that successfully approximates the finite sample features using asymptotic methods.[1]

The key to their asymptotic analysis is to model $\mathbf{\Pi}$ to be local to zero by introducing a "drift" in the parameter $\mathbf{\Pi}$ towards 0 as $T \to \infty$, as described in the following assumption presented in Staiger and Stock (1997):[2]

Assumption $L_{\mathbf{\Pi}}$. $\mathbf{\Pi} = \mathbf{\Pi}_T = \mathbf{C}/\sqrt{T}$, where $\mathbf{C}$ is a fixed $K \times N$ matrix.

In the following derivations, we still consider iid errors but relax the normality assumption, since we are in the land of asymptotics. Also, we still assume that $N = 1$ for simplicity in our note. In Staiger and Stock (1997), they consider a general N.

Remark 6. In contrast to classic asymptotic analysis where F-statistic or μ^2 tends to infinity as sample size T tends to infinity, μ^2 is held constant in weak IV asymptotics:

$$\mu^2 = \frac{\mathbf{\Pi}_T' \mathbf{Z}'\mathbf{Z}\mathbf{\Pi}_T}{\sigma_{\mathbf{v}}^2} = \frac{1}{\sigma_{\mathbf{v}}^2}\mathbf{C}'\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)\mathbf{C} \xrightarrow{p} \frac{1}{\sigma_{\mathbf{v}}^2}\mathbf{C}'\mathbf{Q}_{\mathbf{ZZ}}\mathbf{C}$$

where it is (standard to) assumed that $\mathbf{Z}'\mathbf{Z}/T$ converges in probability to a limit $\mathbf{Q}_{\mathbf{ZZ}}$.

Under weak IV asymptotics, we have the limiting distribution[3]

$$\hat{\boldsymbol{\beta}}_{2\text{SLS}} - \boldsymbol{\beta}_0 \xrightarrow{d} \left(\frac{\sigma_{\mathbf{u}}}{\sigma_{\mathbf{v}}}\right)\frac{(\lambda + \zeta_{\mathbf{v}})'\zeta_{\mathbf{u}}}{(\lambda + \zeta_{\mathbf{v}})'(\lambda + \zeta_{\mathbf{v}})} \tag{6}$$

where

$$\begin{bmatrix}\zeta_{\mathbf{u}} \\ \zeta_{\mathbf{v}}\end{bmatrix} \sim \mathcal{N}(0, \bar{\mathbf{\Sigma}} \otimes \mathbf{I}_K), \quad \bar{\mathbf{\Sigma}} = \begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}, \quad \lambda = \frac{1}{\sigma_{\mathbf{v}}}\mathbf{Q}_{\mathbf{ZZ}}^{1/2}\mathbf{C}.$$

Compare (6) to (4), one can clearly see the similarities. Clearly, $\lambda'\lambda$ is simply the asymptotic version of μ^2. Hence, under weak IV asymptotics, one captures the finite sample features derived by Rothenberg (1984). Staiger and Stock (1997) also provided weak IV asymptotics for the Wald statistic with k-class estimators. However, since the asymptotic distributions all depend on λ and λ cannot be consistently estimated under this particular asymptotic setting, the distributional results is of little use in constructing tests.

[1] Other asymptotic analysis of weak instruments have been developed. For example, "many-instrument asymptotics" is an asymptotic analysis that let the number of instruments K goes to infinity. However, these kinds of asymptotics does not capture the non-normality shown in Figure 1. A more detailed survey can be found in Stock, Wright, and Yogo (2002).

[2] This technique belongs to the family of local asymptotics. Specifically, this kind of "drifting" in parameter is sometimes known as *Pitman drift*, a technique often used in power analysis of tests. (Stock, Wright, and Yogo 2002)

[3] A detailed derivation can be found in Staiger and Stock (1997).

Remark 7 (Relative Bias)**.** Under weak IV asymptotics, then the mean of the asymptotic distribution of the 2SLS estimator is the same as the probability limit of the Ordinary Least Squares (OLS) estimator. The probability limit of the OLS estimator can be expressed as

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \xrightarrow{p} \boldsymbol{\beta}_0 + \frac{\sigma_{\mathbf{vu}}}{\sigma_{\mathbf{v}}^2}.$$

Note that this result is identical to standard asymptotics with $\boldsymbol{\Pi} = 0$. We can defined the "relative bias" of the 2SLS estimator as follows and calculate its expectation

$$\underset{T\to\infty}{\text{plim}} \frac{\hat{\boldsymbol{\beta}}_{\text{2SLS}} - \boldsymbol{\beta}_0}{\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}_0} \implies \mathbf{E}\left[\underset{T\to\infty}{\text{plim}} \frac{\hat{\boldsymbol{\beta}}_{\text{2SLS}} - \boldsymbol{\beta}_0}{\hat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}_0}\right] = \mathbf{E}\left[\frac{(\lambda + \zeta_{\mathbf{v}})'\zeta_{\mathbf{v}}}{(\lambda + \zeta_{\mathbf{v}})'(\lambda + \zeta_{\mathbf{v}})}\right].$$

Due to some properties of distributional form (see footnote 4 in Staiger and Stock (1997)), the expected relative bias only depends on $\lambda'\lambda/K$ and K.

Remark 8. Essentially, Staiger and Stock (1997) provided an asymptotic tool, in contrast to previous attempts with finite samples, e.g., Rothenberg (1984), that makes it easier to analyze weak IV. This asymptotic in itself is not very useful in practice, that is, it does not provide any estimators or test.

Remark 9. One critical conclusion to draw from weak IV asymptotics is that $\hat{\boldsymbol{\beta}}_{\text{2SLS}}$ is inconsistent, since it does not converge in probability to anything. We have two ways to proceed:

1. Screen for first-stage F-statistic or some similar tests. If we are convinced that the instruments are "strong," then we can continuous with the classic 2SLS procedure and test $\text{H}_0 : \boldsymbol{\beta} = 0$ via the usual t-test. We need to decide when instruments are considered weak. (Stock and Yogo 2005)
2. Use a fully-robust test that works regardless of whether the instruments are weak. We need to devise a test that is robust to the scenario that $\boldsymbol{\Pi}$ might be arbitrarily close to 0.

In the YouTube video Stock (2008) provided an anecdote that some friend of his never look at first-stage statistic and always uses fully robust inferences. This seems to be the trend nowadays.

3 Fully Robust Inference

Remark 10. Is there a fully robust "estimator"? That is, is there some estimator $\hat{\boldsymbol{\beta}}$ that is consistent regardless of instrument strength? Clearly not, since if $\boldsymbol{\Pi} = 0$, $\boldsymbol{\beta}$ is not identified. Hence, there is no normal t-test that can be performed. However, there are fully robust "test" that tests the hypothesis $\text{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. (from Shi (2012))

There are three approaches in the literature to do fully robust inference (Stock 2008):

A1. Use a statistic that depends on μ^2, but consider the worst case.

A2. Use a statistic that does not depend on μ^2.

A3. Use a statistic that depends on μ^2, but condition it on a computable sufficient statistic so that the conditional distribution does not depend on μ^2.

Essentially, the first approach basically does not work, as the tests devised using this approach does not have any power, i.e., they are too conservative.[4] There are two known procedures that belongs to the second approach: Anderson-Rubin Test and Kleibergen's Lagrangian Multiplier Test. The third approach is proposed by Moreira (2003), which is the most powerful in terms of statistical power. We will focus on A2. and A3. in this note.

The first fully robust inference in structural estimation is credited to Anderson and Rubin (1949), who created the Anderson-Rubin Test (AR).

Definition 2 (Anderson-Rubin Test)**.** Anderson and Rubin (1949) proposed the Anderson-Rubin Test for testing $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. AR is defined by

$$\mathrm{AR}(\boldsymbol{\beta}_0) := \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)'\mathbf{P}_{\mathbf{Z}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)/K}{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)'(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)/(T-K)}$$

Remark 11. The original idea of Anderson and Rubin (1949) is simple: Suppose we have the hypothesis $\mathrm{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, then the residual $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0$ should be uncorrelated with the instrument $\mathbf{Z}$. (see Stock (2008) at around `1:15:50`)

Remark 12. $\mathrm{AR}(\boldsymbol{\beta}_0)$ has an exact $F_{K,T-K}$ distribution under H_0 (of course, only under finite sample normality). Also, under weak IV asymptotics, $\mathrm{AR}(\boldsymbol{\beta}_0)$ converges in distribution to χ^2_K/K under H_0, regardless of the size of μ^2. (Stock, Wright, and Yogo 2002)

Remark 13 (Robust Confidence Region)**.** Using AR, we can construct a robust confidence region of $\boldsymbol{\beta}$ by collecting all $\boldsymbol{\beta}_0$ such that $\mathrm{AR}(\boldsymbol{\beta}_0)$ fails to reject. (Staiger and Stock 1997)

Remark 14. AR is a uniformly most powerful test when $K = 1$ (Moreira 2009). However, the power of AR is low when $K > 1$.

Remark 15. AR can reject either due to $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ or due to non-exogenous variables.

Before moving on to the statistic proposed by Kleibergen (2002), we should take a look at Moreira (2009) to better understand how these "fully-robust" tests are constructed.[5]

Moreira (2009) considered the following setting: The reduced form of the IV model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{Z}\boldsymbol{\Pi}\boldsymbol{\beta} + \tilde{\mathbf{v}}$ where $\tilde{\mathbf{v}} = \mathbf{u} + \mathbf{v}\boldsymbol{\beta}$. Let $(\tilde{\mathbf{v}}_t, \mathbf{v}_t)$ be a bivariate normal distribution with *known*, for now, covariance matrix $\boldsymbol{\Omega}$. We are interested in the hypothesis $\mathrm{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. This probability model belongs to the (curved) exponential family, hence, $\mathbf{Z}'[\mathbf{y}, \mathbf{X}]$, or $\mathbf{Z}'[\mathbf{y}, \mathbf{X}]D$ for any invertible D, is a sufficient statistic for the pair $(\boldsymbol{\beta}, \boldsymbol{\Pi})$. A convenient choice of D is to let $D = [b_0, a_0]$ where $a_0 = (\boldsymbol{\beta}_0, 1)'$ $b_0 = (1, -\boldsymbol{\beta}_0)'$. With some appropriate scaling, we have that

$$\mathcal{S} = \frac{(\mathbf{Z}'\mathbf{Z})^{-1/2}\mathbf{Z}'[\mathbf{y}, \mathbf{X}]b_0}{\sqrt{b_0'\boldsymbol{\Omega}b_0}} \quad \text{and} \quad \mathcal{T} = \frac{(\mathbf{Z}'\mathbf{Z})^{-1/2}\mathbf{Z}'[\mathbf{y}, \mathbf{X}]\boldsymbol{\Omega}^{-1}a_0}{\sqrt{a_0'\boldsymbol{\Omega}a_0}} \tag{7}$$

are sufficient for the pair $(\boldsymbol{\beta}, \boldsymbol{\Pi})$.[6]

[4] One such procedure, Bonferroni Confidence Region, is discussed in Staiger and Stock (1997). However, as mentioned in Stock (2008), no one really pursuits this route further as it is not that useful.

[5] A manuscript of Moreira (2009) appeared as early as 2001.

[6] This particular set of $\mathcal{S}$ and $\mathcal{T}$ appeared in Stock, Wright, and Yogo (2002). Similar definitions can also be found, of course, in Moreira (2009).

Remark 16. Since $\mathcal{S}$ and $\mathcal{T}$ are sufficient for $(\boldsymbol{\beta}, \mathbf{\Pi})$, we can consider only functions of $\mathcal{S}$ and $\mathcal{T}$ when constructing tests. Notice that in $\mathcal{S}$, we have the term $[\mathbf{y}, \mathbf{X}]b_0$, which evaluates to $\mathbf{u}$ under H_0. Therefore, $\mathcal{S}$ does not depend on $\mathbf{\Pi}$ under the null hypothesis H_0. This is the key why we can construct fully robust tests. (Stock, Wright, and Yogo 2002)

Remark 17. Since $\mathcal{S}$ does not depend on $\mathbf{\Pi}$ under H_0, $\mathcal{T}$ is sufficient for $\mathbf{\Pi}$ under H_0. It follows that a test of $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ based on some function $g(\mathcal{S}, \mathcal{T})$ is similar if its critical value is computed from the conditional distribution of $g(\mathcal{S}, \mathcal{T})$ given $\mathcal{T}$. (Stock, Wright, and Yogo 2002)

Remark 18. Furthermore, one can show that $\mathcal{S}$ and $\mathcal{T}$ are independent normal distributions. Hence the conditional distribution of $g(\mathcal{S}, \mathcal{T})$ is obtained simply by replacing $\mathcal{T}$ with the conditioned value. And since the conditional distribution of $g(\mathcal{S}, \mathcal{T})$ does not depend on $\mathbf{\Pi}$ (since $\mathcal{S}$ does not depend on $\mathbf{\Pi}$), it is a pivotal and its distribution is known (or at least can be computed with simulation). (Moreira 2009; Shi 2012)

Remark 19. In practice, the value of $\mathbf{\Omega}$ is not known. However, we can easily estimate it using $\hat{\mathbf{\Omega}} = [\mathbf{y}, \mathbf{X}]'(\mathbf{I} - \mathbf{P_Z})[\mathbf{y}, \mathbf{X}]/(T - K)$. Henceforth, we will denote the "$\hat{\mathbf{\Omega}}$ plugged-in" versions of $\mathcal{S}$ and $\mathcal{T}$ as $\hat{\mathcal{S}}$ and $\hat{\mathcal{T}}$ respectively. (Stock, Wright, and Yogo 2002)

Now we can rewrite Definition 2 as follows

Definition 3 (Anderson-Rubin Test)**.** The Anderson-Rubin Test is defined as

$$\mathrm{AR}(\boldsymbol{\beta}_0) = \frac{1}{K}\hat{\mathcal{S}}'\hat{\mathcal{S}}$$

Remark 20. It is quite clear why AR is robust: $\mathcal{S}$ does not depend on the value of $\mathbf{\Pi}$ under H_0. However, we now see a potential room for improvement: since we need to have both $\mathcal{S}$ and $\mathcal{T}$ to fully extract the information of $\boldsymbol{\beta}$ and $\mathbf{\Pi}$, we must somehow incorporate the set $\mathcal{S}$ and $\mathcal{T}$ in the test statistic.

Definition 4 (Kleibergen's Lagrangian Multiplier Test)**.** Kleibergen (2002) proposed the Kleibergen's Lagrangian Multiplier Test (KLM) as

$$\mathrm{KLM}(\boldsymbol{\beta}_0) = \frac{(\hat{\mathcal{S}}'\hat{\mathcal{T}})^2}{\hat{\mathcal{T}}'\hat{\mathcal{T}}}$$

Remark 21. When $K = 1$, then $\mathrm{KLM}(\boldsymbol{\beta}_0)$ and $\mathrm{AR}(\boldsymbol{\beta}_0)$ are equivalent.

Remark 22. $\mathrm{KLM}(\boldsymbol{\beta}_0)$ has standard/weak IV asymptotic of χ^2_1 limiting distribution under H_0. It is remarkable that the asymptotic distribution of KLM is free of $\mathbf{\Pi}$, even though $\mathcal{T}$ is present. One way to see that the asymptotic distribution of KLM does not depend on $\mathcal{T}$ is to check that for all values of $\mathcal{T}$, KLM has the same asymptotic distribution χ^2_1.

Remark 23. Compared to the asymptotic distribution of AR, which has χ^2_K/K limiting distribution, $\mathrm{KLM}(\boldsymbol{\beta}_0)$ has higher power when K is large since it has constant degree of freedom.

Remark 24. However, one major downside of KLM is that it has weird power qualities. That is, the test statistic is non-monotone in $|\boldsymbol{\beta} - \boldsymbol{\beta}_0|$ in some cases. Also, the power of it is dominated by the statistic proposed by Moreira (2003). Thus in Stock (2008), Stock said that he does not like KLM.

Definition 5 (Moreira Statistic)**.** Moreira (2003) proposed the following Conditional Likelihood Ratio Test (CLR)

$$\mathrm{CLR}(\boldsymbol{\beta}_0) = \frac{1}{2}\left(\hat{\mathcal{S}}'\hat{\mathcal{S}} - \hat{\mathcal{T}}'\hat{\mathcal{T}} + \sqrt{(\hat{\mathcal{S}}'\hat{\mathcal{S}} + \hat{\mathcal{T}}'\hat{\mathcal{T}})^2 - 4 \cdot \left((\hat{\mathcal{S}}'\hat{\mathcal{S}})(\hat{\mathcal{T}}'\hat{\mathcal{T}}) - (\hat{\mathcal{S}}'\hat{\mathcal{T}})^2\right)}\right)$$

conditional on $\hat{\mathcal{T}}$.

Remark 25. The standard/weak IV asymptotic distribution of $\mathrm{CLR}(\boldsymbol{\beta}_0)$, condition of $\hat{\mathcal{T}}$, is non-standard, and depends on $\boldsymbol{\beta}_0$ and $\hat{\mathcal{T}}$. Moreira (2003) suggests to compute the asymptotic distribution with Monte Carlo simulation.

Remark 26. (Stock 2008) The problem of testing $\mathrm{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ when $N = 1$ is essentially solved by CLR in practice. In simulations studies, CLR is practically the uniformly most powerful test, since it closely hugs the theoretical power envelop. It is more powerful than AR and KLM.

4 Further Topics

1. Complete and Similar tests.
2. Limited Information Maximum Likelihood (LIML) and k-class estimators: A more robust form of *estimation* compared to 2SLS.
3. Implications in Generalized Method of Moment (GMM).

Acronyms

2SLS Two Stage Least Square. 1–4, 7
AER American Economic Review. 1
AR Anderson-Rubin Test. 5–7
CLR Conditional Likelihood Ratio Test. 7
GMM Generalized Method of Moment. 7
iid independent and identically distributed. 1, 3
IV Instrument Variable. 1–7
KLM Kleibergen's Lagrangian Multiplier Test. 6, 7
LIML Limited Information Maximum Likelihood. 7
OLS Ordinary Least Squares. 4

References

Anderson, T. W. and Herman Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". In: *The Annals of Mathematical Statistics* 20.1, pp. 46–63. DOI: 10.1214/aoms/1177730090. URL: https://doi.org/10.1214/aoms/1177730090.

Andrews, Isaiah, James H. Stock, and Liyang Sun (2019). “Weak Instruments in Instrumental Variables Regression: Theory and Practice”. In: *Annual Review of Economics* 11.Volume 11, 2019, pp. 727–753. ISSN: 1941-1391. DOI: https://doi.org/10.1146/annurev-economics-080218-025643. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-economics-080218-025643.

Kleibergen, Frank (2002). “Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression”. In: *Econometrica* 70.5, pp. 1781–1803. DOI: https://doi.org/10.1111/1468-0262.00353. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00353. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00353.

Moreira, Marcelo J. (2003). “A Conditional Likelihood Ratio Test for Structural Models”. In: *Econometrica* 71.4, pp. 1027–1048. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1555489 (visited on 08/15/2024).

— (Oct. 2009). “Tests with correct size when instruments can be arbitrarily weak”. In: *Journal of Econometrics* 152.2, pp. 131–140. URL: https://ideas.repec.org/a/eee/econom/v152y2009i2p131-140.html.

Nelson, Charles R. and Richard Startz (1990a). “Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator”. In: *Econometrica* 58.4, pp. 967–976. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2938359 (visited on 08/13/2024).

— (1990b). “The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One”. In: *The Journal of Business* 63.1, S125–S140. ISSN: 00219398, 15375374. URL: http://www.jstor.org/stable/2353264 (visited on 08/13/2024).

Rothenberg, Thomas J. (1984). “Chapter 15: Approximating the distributions of econometric estimators and test statistics”. In: vol. 2. Handbook of Econometrics. Elsevier, pp. 881–935. DOI: https://doi.org/10.1016/S1573-4412(84)02007-9. URL: https://www.sciencedirect.com/science/article/pii/S1573441284020079.

Shi, Xiaoxia (2012). *Lecture 11: Weak IV*. Lecture Note. URL: https://users.ssc.wisc.edu/~xshi/econ715/Lecture_11_WeakIV.pdf.

Staiger, Douglas and James H. Stock (1997). “Instrumental Variables Regression with Weak Instruments”. In: *Econometrica* 65.3, pp. 557–586. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2171753 (visited on 08/13/2024).

Stock, James (2008). *Methods Lecture, James Stock, “Weak Instruments, Weak Identification, and Many Instruments...”* NBER YouTube Video. URL: https://www.youtube.com/watch?v=DvutyWSofG8.

Stock, James, Jonathan Wright, and Motohiro Yogo (2002). “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”. In: *Journal of Business & Economic Statistics* 20.4, pp. 518–29. URL: https://EconPapers.repec.org/RePEc:bes:jnlbes:v:20:y:2002:i:4:p:518-29.

Stock, James and Motohiro Yogo (2005). “Testing for Weak Instruments in Linear IV Regression”. In: *Identification and Inference for Econometric Models*. Ed. by Donald W.K. Andrews. New York: Cambridge University Press, pp. 80–108. URL: http://www.economics.harvard.edu/faculty/stock/files/TestingWeakInstr_Stock%5C%2BYogo.pdf.