

Instrumental Variables Notes

Questions

- Is there a more technical, less heuristic way of expressing the idea that a single instrumental variable is "used up" in the estimation of the parameters of the structural equation, and so it is not possible to test for the exogeneity of that variable?
 - In the case of multiple instruments, the test of overidentifying restrictions effectively (implicitly) constructs another linear combination of the instruments that is uncorrelated with the first (that is in the relevant way optimal), and tests if this latter linear combination of the instruments is correlated with the residuals computed using the former (in the relevant way optimal) linear combination.
- Similarly, is there a more technical, less heuristic way of expressing the idea that a rejection of the null in the test of overidentifying restrictions cannot be interpreted as a verification of exogeneity because the instruments may be correlated with the residual but not with its estimate?
 - This test is not testing the null that each instrument is uncorrelated with the residuals, such a test is unfeasible because the residuals are not known. The test of overidentifying restrictions is implicitly a test of the consistency between multiple just-identified models.
 - Estimates of the coefficient of interest using different instruments, if the instruments are exogenous, converge to the same "true" value. But if none are exogenous, then it is still possible that they are equal.

Instrumental Variables Simple Case

- The method of instrumental variables is used to consistently estimate the parameters of a causal model where orthogonality fails. This method applies regardless of the source of the endogeneity.
- Consider the causal model $Y = \beta_0 + \beta_1 X + u$, where X is potentially endogenous, i.e. $\text{cov}(X, u) \neq 0$. Continue to suppose that $\mathbb{E}u = 0$.
 - This causal model is traditionally referred to as the structural equation.
- Suppose that there exists candidate instrumental variable Z . If Z satisfies the following conditions, Z is a valid instrumental variable for X .
 - Relevance: $\text{cov}(X, Z) \neq 0$, i.e. Z is correlated with X .
 - Exogeneity: $\text{cov}(Z, u) = 0$, i.e. Z is uncorrelated with the unmodeled determinants of Y in the causal model above.
 - Exclusion: Z does not appear in the causal model above.
- Exogeneity and Exclusion imply that Z has no direct or indirect causal effect on Y except through X . Relevance implies that Z has some (not necessarily causal) effect on X .
- Consider the population linear regression model of X on Z , $X = \pi_0 + \pi_1 Z + v$, where $\mathbb{E}v = 0$ and $\text{cov}(Z, v) = 0$ by construction.
 - This population regression model is referred to as the first stage regression.
 - Let $X^* := \pi_0 + \pi_1 Z$, then the first stage regression decomposes X into a component X^* that is collinear with Z and a component v that is orthogonal to Z .
- By substitution of the first stage regression into the structural equation,
 $Y = \beta_0 + \beta_1(\pi_0 + \pi_1 Z + v) + u = (\beta_0 + \beta_1\pi_0) + (\beta_1\pi_1)Z + (u + \beta_1v) =: \gamma_0 + \gamma_1 Z + \epsilon$. By supposition, by construction of the first stage regression, and by Exogeneity, $\mathbb{E}(u + \beta_1v) = \mathbb{E}u + \beta_1\mathbb{E}v = 0$ and $\text{cov}(Z, u + \beta_1v) = \text{cov}(Z, u) + \beta_1\text{cov}(Z, v) = 0$. Then $Y = \gamma_0 + \gamma_1 Z + \epsilon$ is a linear regression of Y on Z .
 - This population regression model is the reduced form regression.

Indirect Least Squares

- From the above, $\gamma_1 = \beta_1\pi_1$, then $\beta_1 = \frac{\gamma_1}{\pi_1}$. In words, the causal effect of X on Y is equal to the coefficient on Z in the reduced form regression divided by the coefficient on X in the first stage regression.
 - Intuitively, a one unit increase in Z is associated with a π_1 unit increase in X . A one unit increase in X causes a β_1 unit increase in Y . Given that Z has no direct or indirect effect on Y except through X , a one unit increase in Z is associated with a $\pi_1\beta_1$ unit increase in Y .
- Then $\beta_1 = \frac{\text{cov}(Y, Z)}{\text{var}(Z)} \div \frac{\text{cov}(X, Z)}{\text{var}(Z)} = \frac{\text{cov}(Y, Z)}{\text{cov}(X, Z)}$, and $\hat{\beta}_1 = \frac{\hat{\text{cov}}(Y, Z)}{\hat{\text{cov}}(X, Z)}$ consistently estimates β_1 .

Two-Stage Least Squares

- By substitution of the first stage regression into the structural equation, $Y = \beta_0 + \beta_1(X^* + v) + u = \beta_0 + \beta_1 X^* + \epsilon$.
 - $cov(X^*, \epsilon) = cov(\pi_0 + \pi_1 Z, u + \beta_1 v) = cov(\pi_1 Z, u + \beta_1 v) = 0$, where the final equality holds by Exogeneity and by construction of the first stage regression.
 - Then, $Y = \beta_0 + \beta_1 X^* + \epsilon$ is a population linear regression model of Y on X^* . Hence $\beta_1 = \frac{cov(Y, X^*)}{var(X^*)}$.
- The equations for β_1 from indirect least squares and two-stage least squares are equivalent.
 - $\beta_1 = \frac{cov(Y, X^*)}{var(X^*)} = \frac{cov(Y, \pi_0 + \pi_1 Z)}{var(\pi_0 + \pi_1 Z)} = \frac{\pi_1 cov(Y, Z)}{\pi_1^2 var(Z)} = \frac{cov(Y, Z)}{\pi_1 cov(Z, Z)} = \frac{cov(Y, Z)}{cov(X, Z)}$.
- X^* is consistently estimated by $\hat{X} := \hat{\pi}_0 + \hat{\pi}_1 Z$, where $\hat{\pi}_0, \hat{\pi}_1$ are computed in the familiar way. Then, β_1 is consistently estimated by $\hat{\beta}_1 = \frac{cov(Y, \hat{X})}{var(\hat{X})}$.
 - This is again equivalent to the consistent estimator derived under indirect least squares.
- Where a regressor is endogenous, we do not care whether OLS or 2SLS is more precise, because OLS will be biased. If the regressor is exogenous, then we care which is more precise. With 2SLS, the variability of the regressor (the component of the endogenous variable that is explained by the instrument compared to the endogenous variable itself) decreases. The variability of the residuals remains approximately equal because both OLS and 2SLS estimators are consistent, so the residuals converge in large samples. So 2SLS standard error is weakly greater than the OLS standard error. OLS residual minimises the sum of squared errors by construction. This is unambiguous.
 - 2SLS is more robust (because the instrument can be made random) but could be weaker.

Instrumental Variables General Case

- The important models in instrumental variable regression are the structural equation (casual model), the first stage regression (of the endogenous variable on the instruments and controls), the second stage regression (of the dependent variable on the residual of the first stage and the controls), and the reduced form regression (of the dependent variable directly on the instrument and controls).
- Consider the more general causal model $Y = \beta_0 + \beta_1 X + \beta_2 W_1 + \dots + \beta_{r+1} W_r$, where X is potentially endogenous and each of W_1, \dots, W_r is exogenous, and is either a determinant of Y or a proxy for some determinant of Y . Continue to suppose that the causal effect of interest is β_1 .
- The conditions for instrument validity are as follows.
 - Relevance: in the first stage regression of X on instruments Z_1, \dots, Z_m and controls W_1, \dots, W_r ($X = \pi_0 + \pi_1 Z_1 + \dots + \pi_m Z_m + \pi_{m+1} W_1 + \dots + \pi_{m+r} W_r + v$ where $\mathbb{E}v = 0$ and $cov(Z_1, v), \dots, cov(W_r, v) = 0$ by construction), at least one of the coefficients on the instruments π_1, \dots, π_m is non-zero.
 - Exogeneity: $cov(Z_I, u) = 0$ for all $I \in \{1, \dots, m\}$, i.e. each Z_I is uncorrelated with the unmodelled determinants of Y in the causal model above.
 - Exclusion: each Z_I does not appear in the causal model above.
- Exogeneity and Exclusion are exactly analogous to the simple case.

Two-Stage Least Squares

- The first stage regression is $X = \pi_0 + \pi_1 Z_1 + \dots + \pi_m Z_m + \pi_{m+1} W_1 + \dots + \pi_{m+r} W_r + v =: X^* + v$, where $\mathbb{E}v = 0$ and $cov(Z_1, v) = \dots = cov(W_r, v) = 0$ by construction.
 - The first stage regression is defined exactly analogously to the simple case.
 - Noting that X^* is a linear function of Z_1, \dots, W_r , and that each of these is uncorrelated with v , $cov(X^*, v) = 0$.
- By substitution into the structural equation,
 - $Y = \beta_0 + \beta_1(X^* + v) + \beta_2 W_1 + \dots + \beta_{r+1} W_r + u$
 - $= \beta_0 + \beta_1 X^* + \beta_2 W_1 + \dots + \beta_{r+1} W_r + (\beta_1 v + u)$
 - $= \beta_0 + \beta_1 X^* + \beta_2 W_1 + \dots + \beta_{r+1} W_r + \epsilon$
 - $cov(W_I, \epsilon) = \beta_1 cov(W_I, v) + cov(w_I, u) = 0$ by construction of the first stage regression and by supposition for all $I \in \{1, \dots, r\}$.
 - $cov(X^*, \epsilon) = \beta_1 cov(X^*, v) + cov(X^*, u) = 0$ by construction of the first stage regression and by Exogeneity (given that X^* is a linear function of Z_1, \dots, Z_m).
 - $\mathbb{E}\epsilon = \beta_1 \mathbb{E}v + \mathbb{E}u = 0$ by construction of the first stage regression and by construction of the structural equation.
 - Then, orthogonality holds in the above equation, and the above equation is a population linear regression of Y on X^*, W_1, \dots, W_r . This is the reduced form regression.
- If, additionally, the regressors X^*, W_1, \dots, W_r are not perfectly multicollinear, then the population linear regression problem has a unique solution, which is β_0, β_1, \dots .
 - Supposing that there is no perfect multicollinearity between W_1, \dots, W_r , given that Z_1, \dots, Z_m satisfy Relevance, X^* is not simply a linear function of W_1, \dots, W_r , hence there is no perfect multicollinearity between X^*, W_1, \dots, W_r .

- Then, OLS regression consistently estimates the parameters of the structural equation.
 - X^* is consistently estimated by $\hat{X} := \hat{\pi}_0 + \hat{\pi}_1 Z_1 + \dots + \hat{\pi}_m Z_m$ where $\hat{\pi}_1, \dots, \hat{\pi}_m$ are computed in the familiar way by the Frisch-Waugh-Lovell theorem. Then, $\beta_0, \beta_1, \dots, \beta_{r+1}$ are consistently estimated by the OLS coefficients in the regression of Y on \hat{X}, W_1, \dots, W_r .

Indirect Least Squares

- The indirect least squares reduced form regression is $Y = \beta_0 + \beta_1 \pi_0 + \beta_1 \pi_1 Z_1 + \beta_1 \pi_2 Z_2 + \dots + \beta_1 \pi_m Z_m + \beta_2 W_1 + \dots + \beta_{r+1} W_r + \epsilon$. Which can be verified to be a population regression model of Y on Z_1, \dots, W_r .
- Each $\beta_1 \pi_I$ (for $I \in \{1, \dots, m\}$) can be consistently estimated by OLS regression. Similarly, each π_I can be consistently estimated by OLS regression. Then, each of the ratios $\frac{(\hat{\beta}_1 \pi_I)}{\hat{\pi}_I}$ consistently estimates β_1 .
- In general, each of these estimators yields a distinct estimate, then, it is preferable to estimate β_1 by two-stage least squares, which "optimally utilises the m instruments to produce a single, more efficient, estimate of β_1 .

Inference in Two-Stage Least Squares Regression

- Inference in two-stage least squares regression is largely analogous to that in OLS regression.
- As in OLS regression, in what follows, it is supposed that:
 - (1) each of the regressors W_1, \dots, W_r is exogenous,
 - (2) the sample is an independently and identically distributed random sample,
 - (3) large outliers are unlikely,
 - (4) there is no perfect multicollinearity among any of the regressors W_1, \dots, W_r in the structural equation, or the regressors Z_1, \dots, Z_m in the first stage regression.
- It is additionally supposed that:
 - (5) the instrumental variables Z_1, \dots, Z_m are valid instruments that satisfy Relevance, Exogeneity, and Exclusion.
- Then, the law of large numbers implies that the sample linear regression parameters are consistent estimators of the parameters of the causal model.
- The central limit theorem implies that $\frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \approx \frac{\hat{\beta}_1 - \beta_1}{sd(\hat{\beta}_1)} \rightarrow^d N[0, 1]$.
- Let ω_{β_1} and $\hat{\omega}_{\beta_1}$ be such that $sd(\hat{\beta}_1) = n^{-\frac{1}{2}} \omega_{\beta_1}$ and $se(\hat{\beta}_1) = n^{-\frac{1}{2}} \hat{\omega}_{\beta_1}$.
 - It can be proven that $\omega_{\beta_1}^2 = \frac{\mathbb{E}(X^* - \mathbb{E}X^*)^2 u^2}{[\mathbb{E}(X^* - \mathbb{E}X^*)^2]^2}$. (This is exactly analogous to the equation for OLS regression.)
 - Supposing that u is homoskedastic in the sense that $\mathbb{E}[u^2 | Z_1, \dots, Z_m] = \mathbb{E}u^2$,

$$\begin{aligned} \mathbb{E}(X^* - \mathbb{E}X^*)^2 u^2 &= \mathbb{E}[\mathbb{E}[(X^* - \mathbb{E}X^*)^2 u^2 | Z_1, \dots, Z_m]] \\ &= \mathbb{E}[(X^* - \mathbb{E}X^*)^2 \mathbb{E}[u^2 | Z_1, \dots, Z_m]] \\ &= \mathbb{E}u^2 \mathbb{E}(X^* - \mathbb{E}X^*)^2 \\ &= var(u) var(X^*) \end{aligned}$$
 - where the first equality holds by the law of iterated expectations, the second equality holds by conditioning, the third equality holds by homoskedasticity and linearity of expectation, and the fourth equality holds by definition and the fact that $\mathbb{E}u = 0$.
 - u here refers to the residual of the structural equation, which is presumably estimated by $\hat{u} = Y - \hat{\beta}_0 - \hat{\beta}_1 X - \hat{\beta}_2 W_1 - \dots - \hat{\beta}_{r+1} W_r$, where $\hat{\beta}_0, \dots, \hat{\beta}_{r+1}$ are estimated by two-stage least squares or indirect least squares.
 - Note that the sense of homoskedasticity applied here is different from the sense of homoskedasticity applied in the OLS case. Here, u^2 is mean-independent of the instruments. In the OLS case, u^2 is mean-independent of the regressors X_1, X_2, \dots .
 - Then, $\omega_{\beta_1} = \frac{var(u)}{var(X^*)}$.
- Comparing against the case of OLS regression of Y on exogenous X , we note:
 - Like the OLS case, an increase in the variability of the residual u worsens the precision of the estimator for β_1 .
 - The variability of the residual u of a two stage least squares or indirect least squares regression is weakly greater than that of a OLS regression because the regressions are of the same form, and the OLS parameters minimise the variability of the residual by construction.
 - Unlike the OLS case, an increase in the variability of X does not necessarily improve the precision of the two-stage least squares estimator. Only an increase in the variability of the component of X predicted by the instruments improves the precision of the two-stage least squares estimator.

- The variability of the instrument-predicted component of X is weakly less than the variability of X . This is because by construction, the instrument-predicted component of X is uncorrelated with the residual component of X , so the variance of X is equal to the sum of the variances of the two components. Then the variance of the instrument-predicted component is weakly less than that of X . It is strictly less if the instrument does not perfectly predict X .
- If X is exogenous, the OLS estimator is more efficient than the two-stage least squares estimator.
 - $\text{var}(X) = \text{var}(X^* + v) = \text{var}(X^*) + \text{var}(v) > \text{var}(X^*)$.
 - Then, the asymptotic variance of the OLS estimator is smaller than that of the two-stage least squares estimator.

Testing Instrument Validity

- Exclusion can be verified by inspection of the structural equation (to verify that it does not contain any of the instruments).

Relevance

- One test of Relevance is a F test of the hypothesis:
 - $H_0 : \forall I \in \{1, \dots, m\} : \pi_I = 0$, against
 - $H_1 : \exists I \in \{1, \dots, m\} : \pi_I \neq 0$, in
 - $X = \pi_0 + \pi_1 Z_1 + \dots + \pi_m Z_m + \pi_{m+1} W_1 + \dots + \pi_{m+r} W_r + v$ (the auxiliary regression), where
 - $\mathbb{E}v = \text{cov}(Z_I, v) = \dots = \text{cov}(W_r, v) = 0$ by construction.
- A rejection of the null hypothesis in this test can be unambiguously interpreted as evidence in favour of Relevance.
- Relevance is sufficient for consistent estimation of the causal effects of interest, but is not sufficient for inferences on the basis of two-stage least squares to be reliable. In other words, Relevance is not sufficient for the convergence in distribution of $\frac{\hat{\beta}_1 - \beta_1}{\text{sd}(\hat{\beta}_1)}$ to $N(0, 1)$, hence not sufficient for straightforwardly conducting hypothesis tests and constructing confidence intervals.
 - Even if Relevance is satisfied, if each π_I is very small, the familiar convergence in distribution to the standard normal does not hold, even in large samples. This is the phenomenon of weak instruments.
- Inferences based on the two-stage least squares estimator and its associated standard error are reliable only if Relevance holds with a sufficiently wide margin.
- One "rule of thumb" test for weak instruments is to conduct the above F test against a critical value of 10.
- If Relevance fails, the first-stage regression estimates of the endogenous variable is simply a function of the controls, then this estimate and the controls are perfectly multicollinear.

Exogeneity

- One test of Exogeneity is a F test of the hypothesis:
 - $H_0 : \forall I \in \{1, \dots, m\} : \delta_I = 0$, against
 - $H_1 : \exists I \in \{1, \dots, m\} : \delta_I \neq 0$, in
 - $u = \delta_0 + \delta_1 Z_1 + \dots + \delta_m Z_m + \delta_{m+1} W_1 + \dots + \delta_{m+r} W_r + \eta$, where
 - $\mathbb{E}\eta = \text{cov}(Z_1, \eta) = \dots = \text{cov}(W_r, \eta) = 0$ by construction.
- The null hypothesis is the hypothesis that the residual in the structural equation is uncorrelated with each of the instruments. The alternative hypothesis is the opposite claim.
- The steps of this test are:
 - Estimate the parameters of structural equation $Y = \beta_0 + \beta_1 X + \beta_2 W_1 + \dots + \beta_{r+1} W_r + u$ by two-stage least squares regression. The first stage regression is $X = \pi_0 + \pi_1 Z_1 + \dots + \pi_m Z_m + \pi_{m+1} W_1 + \dots + \pi_{m+r} W_r + v$. The second stage regression is $Y = \beta_0 + \beta_1 X^* + \beta_2 W_1 + \dots + \beta_{r+1} W_r + \epsilon$, where $X^* = X - v$ and $\epsilon = \beta_1 v + u$.
 - Compute the residuals $\hat{u} := Y - \hat{\beta}_0 - \hat{\beta}_1 X - \hat{\beta}_2 W_1 - \dots - \hat{\beta}_{r+1} W_r$.
 - Perform OLS regression of \hat{u} on $Z_1, \dots, Z_m, W_1, \dots, W_r$.
 - Compute the residuals of this regression, then compute the F statistic, and the adjusted F statistic $F^* = \frac{m}{m-1} F$. Compare F^* against critical values drawn from the $F_{m-1, \infty}$ distribution.
 - The adjustment by multiplying by $\frac{m}{m-1}$ is a degrees of freedom adjustment that is necessary because one degree of freedom is "consumed" in estimating the parameters of the structural equation.
- This test is sometimes referred to as a test of "overidentifying restrictions".
- A failure to reject the null does not constitute unambiguous evidence that Exogeneity is satisfied because it could be the case that each instrument is correlated with the residual u but not with the estimate of the residual that estimation of the

structural equation using every other instrument.

- A rejection of the null constitutes unambiguous evidence that at least one instrument is not exogenous.
- In the case of an endogenous instrument, the bias of the estimator of the coefficient on the instrument in the reduced form regression is inherited by the estimator of the coefficient on the endogenous variable computed by indirect least squares (which is equivalent to that computed by two stage least squares). So for example, if the instrument has positive omitted variable bias, the estimate of the coefficient on the endogenous variable (that the method of instrumental variables is used to correct for) will inherit this positive bias.

Randomised Control Trials with Instrumental Treatment

- In many cases of interest, random assignment (even conditional random assignment) is not possible because the receipt of treatment necessarily involves some element of choice that cannot be ethically eliminated.
 - For example, suppose that researchers are interested in the causal effect of completing a university degree on subsequent earnings. It is not possible to compel completion of a university degree and it is not ethical to compel non-completion of a university degree. Then, the treatment cannot be randomly assigned.
- In such cases, while treatment cannot be randomly assigned, it is possible that some "inducement" or "offer" of treatment can be randomly assigned.
 - Such random assignment implies that this inducement or offer satisfies Exogeneity.
 - Supposing that this inducement or offer affects treatment, it is also Relevant.
 - If it is also plausible that this inducement or offer does not itself directly affect the outcome of interest, then it satisfies Exclusion.
 - Then, randomly assigned inducement or offer is a valid instrument for treatment.

Heterogenous Causal Effects

- Consider the causal model $Y_i = \beta_0 + \beta_{1i}X_i + u_i$ and $X_i = \pi_0 + \pi_{1i}Z_i + v_i$.
 - These equations describe a situation where the causal effect of X on Y varies across observations, and the responsiveness of Z to X varies across observations.
- Suppose that Z_i is randomly assigned such that $Z_{1i} \perp\!\!\!\perp \beta_{1i}, \pi_{1i}, u_i, v_i$, then Exogeneity is satisfied. Suppose further that Relevance is satisfied, i.e. $cov(X_i, Z_i) \neq 0$. By inspection, Exclusion is satisfied, i.e. Z_i does not appear in the structural equation.
- The two-stage least squares estimator $\hat{\beta} := \frac{cov(Y, Z)}{cov(X, Z)}$ is consistent for $\beta = \frac{cov(Y, Z)}{cov(X, Z)}$. It can be proved that $\beta = \mathbb{E}(\beta_{1i} \times \frac{\pi_{1i}}{\mathbb{E}\pi_{1i}})$, which is the local average treatment effect (LATE).
 - LATE is the responsiveness-weighted average treatment effect.
 - Then, LATE is likely to vary between experiments that use different instruments. So even in large samples, the two-stage least squares estimators of the same causal effect with different instruments, in general, do not coincide.