GMM, Indirect Inference and Bootstrap

Instrumental variables

Willi Mutschler

TU Dortmund

Winter 2015/2016

TO IMPROVE

- Slide with inconsistency and forecastablity? Do you really need it?
- Make the structure and dimensions of W, $W'u$, W_t, $W_t'u_t$ more clear
- Interpretations and structures of M_X, P_W

- OLS is not consistent and biased if $E\left(u_t|X_t\right) \neq 0$
- Define an information set Ω_t (a σ-algebra), such that

$$E\left(u_t|\Omega_t\right) = 0$$

- This rationality conditions or moment conditions can be used for estimation
- Variables in Ω_t are called **instrumental variables** (or instruments)
- We denote the instrument vector by W_t

Errors in variables (e.g. in questionnaires)

- Consider the model

$$y_t = \alpha + \beta x_t^* + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

- The exogenous variable x_t^* is unobservable
- We can only observe

$$x_t = x_t^* + v_t$$

  where $v_t \sim iid(0, \sigma_v^2)$ are independent of everything else
- Estimators of $y_t = \alpha + \beta x_t + u_t$ are inconsistent [P]

Omitted variables bias

- Let

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$$

- If x_2 is unobservable, one estimates

$$y_t = \alpha + \beta_1 x_{1t} + u_t$$

where $u_t = \beta_2 x_{2t} + \varepsilon_t$

- If x_{2t} and x_{1t} are correlated then so are u_t and x_{1t}

Endogeneity

- Standard example: supply and demand curves determine both price and quantity

$$
\begin{aligned}
q_t &= \gamma_d p_t + X_t^d \beta_d + u_t^d \\
q_t &= \gamma_s p_t + X_t^s \beta_s + u_t^s
\end{aligned}
$$

- Solve for q_t and p_t

$$
\begin{bmatrix} q_t \\ p_t \end{bmatrix} = \begin{bmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{bmatrix}^{-1} \left(\begin{bmatrix} X_t^d \beta_d \\ X_t^s \beta_s \end{bmatrix} + \begin{pmatrix} u_t^d \\ u_t^s \end{pmatrix} \right)
$$

Instrumental variables

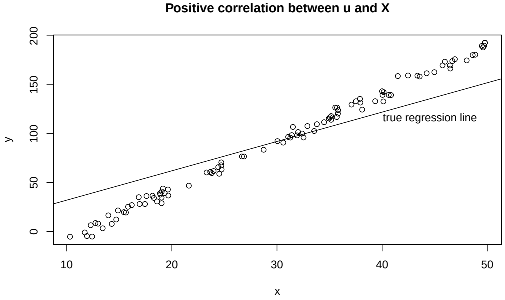
Correlation between errors and disturbances (III)

- Since q_t and p_t depend on both u_t^d and u_t^ssingle equation OLS estimation of

$$
\begin{aligned}
q_t &= \gamma_d p_t + X_t^d \beta_d + u_t^d \\
q_t &= \gamma_s p_t + X_t^s \beta_s + u_t^s
\end{aligned}
$$

  is inconsistent
- The right hand side variable p_t is correlated with the error term
- The condition $E\left(u_t|\Omega_t\right) = 0$ is violated if p_t is in Ω_t

- **Warning!** Inconsistency is not always a problem
- If we simply want to forecast, we can use inconsistent estimators
- Trivial example:



- Let W denote the $T \times K$ matrix of instruments
- All columns of X with $X_t \in \Omega_t$ should be included in W
- Then $E\left(u_t | W_t\right) = 0$ implies the moment condition

$$E\left(W'u\right) = E\left(W'\left(y - X\beta\right)\right) = 0$$

- The IV estimator is a method of moment estimator
- The solution is

$$\hat{\beta}_{IV} = \left(W'X\right)^{-1} W'y$$

- The simple IV estimator is consistent if

$$plim\frac{1}{n}W'X = S_{WX}$$

is deterministic and nonsingular [P]

- The simple IV estimator is asymptotically normal,

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) \to U \sim N\left(0, \sigma^2\left(S_{WX}\right)^{-1} S_{WW}\left(S'_{WX}\right)^{-1}\right)$$

where $S_{WW} = plim\frac{1}{n}W'W$ [P]

- Instruments must be
  1. exogenous, i.e. $plim \frac{1}{n} W'u = 0$
  2. valid, i.e. $plim \frac{1}{n} W'X = S_{WX}$ non-singular
- Natural experiments (weather, earthquakes, ...)
- Angrist and Pischke (2009):

Good instruments come from a combination of institutional knowledge and ideas about the processes determining the variable of interest.

Examples

Natural experiments

1. Brückner and Ciccone: Rain and the demographic window of opportunity, Econometrica 79 (2011) 923-947
2. Angrist and Evans: Children and their parents' labor supply: Evidence from exogenous variation in family size, American Economic Review 88 (1998) 450-77.

Examples

Natural experiments

1. Brückner and Ciccone: Rain and the demographic window of opportunity, Econometrica 79 (2011) 923-947
2. Angrist and Evans: Children and their parents' labor supply: Evidence from exogenous variation in family size, American Economic Review 88 (1998) 450-77.

- Negative konomische EK-Schocks knnen fr demokratische Verbesserung Instrumente liefern (konomische Rezessionen), Sub-Sahara Afrikanische Lnder, INstrument: Negative Regenschocks
- Endogenitt von Fertilitt, Arbeitsentscheidung nach Geburt. Effekt von Kindergeburt auf Arbeitsangebot. Instrument: Prferenz fr mixed Kinder-Komposition, Eltern mit 2 Mdels probieren noch mal fr Jungen.

Examples

Institutional arrangements

1. Angrist and Krueger: Does Compulsory School Attendance Affect Schooling and Earnings?, Quarterly Journal of Economics 106 (1991) 979-1014.
2. Levitt: The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation, Quarterly Journal of Economics 111 (1996) 319-351.

Examples

Institutional arrangements

1. Angrist and Krueger: Does Compulsory School Attendance Affect Schooling and Earnings?, Quarterly Journal of Economics 106 (1991) 979-1014.
2. Levitt: The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation, Quarterly Journal of Economics 111 (1996) 319-351.

- Welcher Monat man geboren ist: Keine Korrelation mit Fhigkeit und Earnings, keine determinante von Fhigkeiten, aber korreliert mit years of schooling. Mnner die gezwungen werden zur Schule zu gehen, verdienen meht
- Gerichtsverfahren. Simultanitt/Endogenitt zwischen Gefngnisbevlkerung un Verbrechensraten. Ideee: Gefngnisbevlkerung steigt impliziert sinkende Verbrechensrate. Messfehler. IV: Korreliert mit Verndung Gef.Bevlkerung aber sonst unrelated to crime rates, state prison overcroding litigation.

- In a time series context, one can sometimes use lagged endogenous regressors as instrumental variables
- Example:
$$y_t = \alpha + \beta x_t + u_t$$
with $E(u_t|x_t) \neq 0$
- If $Cov\,(x_t, x_{t-1}) \neq 0$ but $Cov\,(u_t, x_{t-1}) = 0$, then x_{t-1} can be used as instrumental variable
- Attention: $Cov\,(u_t, x_{t-1}) = 0$ is not always obvious

Example (Measurement error in time series)

Consider the model

$$
\begin{aligned}
y_t &= \alpha + \beta x_t^* + u_t \\
x_t^* &= \rho x_{t-1}^* + \varepsilon_t \\
x_t &= x_t^* + v_t.
\end{aligned}
$$

Then x_{t-1} is a valid instrument for a regression of y_t on x_t, and α and β will be estimated consistently.

[IVLags.R]

Example (Omitted variable bias in time series)

Consider the model

$$
\begin{aligned}
y_t &= \alpha + \beta_1 x_{1t} + \beta_2 x_{t2} + u_t \\
x_{1t} &= \rho_{11} x_{1,t-1} + \rho_{12} x_{2,t-1} + \varepsilon_{1t} \\
x_{2t} &= \rho_{21} x_{1,t-1} + \rho_{22} x_{2,t-1} + \varepsilon_{2t}
\end{aligned}
$$

Then $x_{1,t-1}$ is **not** a valid instrument for a regression of y_t on x_{1t}, and α and β_1 will **not** be estimated consistently.

[IVLags.R]

Example (Endogeneity in time series)

Consider the model

$$
\begin{aligned}
y_t &= \alpha + \beta_1 x_t + \beta_2 y_{t-1} + u_t \\
x_t &= \gamma + \delta_1 y_t + \delta_2 x_{t-1} + v_t
\end{aligned}
$$

Then $x_{1,t-1}$ is a valid instrument for a regression of y_t on x_t and y_{t-1}, and α, β_1 and β_2 will be estimated consistently.

[`IVLags.R`]

- If the number of instruments L is larger than the number of parameters K, the model is **overidentified**
- Right-multiply the $T \times L$ matrix W by an $L \times K$ matrix J to obtain an $T \times K$ instrument matrix WJ
- Linear combinations of the instruments in W
- One can show that the asymptotically optimal matrix is $J = (W'W)^{-1} W'X$, see Davidson and MacKinnon.

- The generalized IV estimator is

$$\begin{aligned}\hat{\beta}_{GIV} &= \left((WJ)'X\right)^{-1}(WJ)'y\\ &= \left(X'W\left(W'W\right)^{-1}W'X\right)^{-1}X'W\left(W'W\right)^{-1}W'y\\ &= \left(X'P_WX\right)^{-1}X'P_Wy\end{aligned}$$

  with $P_W = W\left(W'W\right)^{-1}W'$
- Consistency and asymptotic normality still hold

- The two-stage-least-squares (2SLS) interpretation
- The matrix J is similar to $\hat{\beta}$ in the standard OLS model,

$$J = \left(W'W\right)^{-1} W'X$$

- Hence, WJ is similar to $X\hat{\beta}$
- The optimal instruments are obtained if we regress the endogenous regressors on the instruments (1st stage), and then use the fitted values as regressors (2nd stage)

- The finite sample properties of IV estimators are complex
- In the overidentified case, the first $L - K$ moments exist, but higher moments do not
- If the expectation exists, IV estimators are in general biased
- The simple IV estimator has **very** heavy tails, even the first moment does not exist!
- The estimator can be extremely far off the true value

[`ivfinite.R`]

- Exact hypothesis tests are usually not feasible
- Asymptotic tests are based on the asymptotic normality
- An estimator of the covariance matrix of $\hat{\beta}_{IV}$ is

$$\widehat{Cov}\left(\hat{\beta}_{IV}\right) = \hat{\sigma}^2 \left(X' P_W X\right)^{-1}$$

with

$$\begin{aligned} P_W &= W \left(W'W\right)^{-1} W' \\ \hat{\sigma}^2 &= \frac{1}{n} \left(y - X\hat{\beta}_{IV}\right)' \left(y - X\hat{\beta}_{IV}\right) \end{aligned}$$

- Asymptotic t-test

$$
\begin{aligned}
H_0 &: \beta_i = \beta_{i0} \\
H_1 &: \beta_i \neq \beta_{i0}
\end{aligned}
$$

- Under the null hypothesis, the test statistic

$$
t = \frac{\hat{\beta}_i - \beta_{i0}}{\sqrt{\widehat{Var}\left(\hat{\beta}_i\right)}}
$$

is asymptotically $N(0,1)$

- Asymptotic Wald test (similiar to an F-test)

$$H_0 : \beta_2 = \beta_{20}, \quad H_1 : \beta_2 \neq \beta_{20}$$

where β_2 is a length L subvector of β

- Under the null hypothesis, the test statistic

$$W = \left(\hat{\beta}_2 - \beta_{20}\right)' \left[\widehat{Cov}\left(\hat{\beta}_2\right)\right]^{-1} \left(\hat{\beta}_2 - \beta_{20}\right)$$

is asymptotically χ^2 with L degrees of freedom

- Testing overidentifying restrictions
- The identifying restrictions are

$$\begin{aligned} E(u_t|W_t) &= 0 \\ \text{or} \quad E\left(W'u\right) &= 0 \end{aligned}$$

- If the model is just identified the validity of the restriction cannot be tested
- If the model is overidentified, one can test if the overidentifying restrictions hold, i.e. if the instruments are valid and exogenous

- Basic test idea: Check if the IV residuals can be explained by the full set of instruments
- Compute the IV residuals $\hat{u}$
- Regress the residuals on all instruments W
- Under the null hypothesis, the test statistic

$$nR^2 \sim \chi^2_m$$

where m is the degree of overidentification

Davidson and MacKinnon (2004, p. 338):

> *Even if we do not know quite how to interpret a significant value of the overidentification test statistic, it is always a good idea to compute it. If it is significantly larger than it should be by chance under the null hypothesis, one should be extremely cautious in interpreting the estimates, because it is quite likely either that the model is specified incorrectly or that some of the instruments are invalid.*

- Durbin-Wu-Hausman test

$$
\begin{aligned}
H_0 &: \quad E\left(X'u\right) = 0 \\
H_1 &: \quad E\left(W'u\right) = 0
\end{aligned}
$$

- Test if IV estimation is really necessary or if OLS would do
- Under H_0, OLS is consistent and efficient, but IV is just consistent
- Under H_1, OLS is inconsistent, but IV is still consistent
- Basic test idea: Compare $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$. If they are 'too different', reject H_0

- The difference between the estimators is

$$
\begin{aligned}
& \hat{\beta}_{IV} - \hat{\beta}_{OLS} \\
= \;& \left(X' P_W X\right)^{-1} X' P_W y - \left(X'X\right)^{-1} X'y \\
= \;& \left(X' P_W X\right)^{-1} \left(X' P_W y - \left(X' P_W X\right) \left(X'X\right)^{-1} X'y\right) \\
= \;& \left(X' P_W X\right)^{-1} \left(X' P_W \left(I - X \left(X'X\right)^{-1} X'\right) y\right) \\
= \;& \left(X' P_W X\right)^{-1} \left(X' P_W M_X \, y\right)
\end{aligned}
$$

- We need to test if $X' P_W M_X y$ is significantly different from 0
- This term is identically equal to zero for all variables in X that are instruments (i.e. that are also in W)
- Denote by $\tilde{X}$ all possibly endogenous regressors
- To test if $\tilde{X}' P_W M_X y$ is significantly different from zero, perform a Wald test of $\delta = 0$ in the regression

$$y = X\beta + P_W \tilde{X} \delta + u$$