Advanced Econometrics

University of Warsaw

Jerzy Mycielski

Doctoral School of Social Sciences, 2024

Resampling methods

- Derivation of the final sample distributions is often practically impossible
- It also happens that even the derivation of the asymptotic distributions of the statistics is cumbersome
- In this cases resampling methods are often used as a tool to:
  - obtain in a simple way asymptotically valid approximation of the distributions of statistics
  - improve the quality of approximations relative to standard methods (asymptotic refinement)
- There are two main resampling methods used:
  - jackknife
  - bootstrap
- Jackknife is a simpler method but bootstrap is more general

Jackknife

- Jackknife sample is constructed by omitting observation i
- In this way we can obtain n samples
- For every one these samples we obtain calculate estimator $\widehat{\theta}_{(-i)}$
- The Tukey's jackknife estimator of variance

$$\widehat{\mathbf{V}}_{\widehat{\theta}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}}_{(-i)} \right) \left(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}}_{(-i)} \right)'$$

- This method can also be used for estimation of the variance of transformations of estimators $g\left(\widehat{\theta}\right)$
- When used for clustered samples, we omit one cluster rather than one observation
- It can be proven that under quite general conditions jackknife estimator of variance is equivalent (but not better) to one obtained with delta method but does not require calculation of derivatives

Bootstrap

- This method is based on drawing randomly *with replacement* the bootstrap sample from original sample.
- Notice that bootstrap sample contains some duplicates of original observations
- Number of bootstrapped samples B can be made arbitrary large
- Bootstrap can be used to estimate variance of the estimator:

$$\widehat{\mathbf{V}}^{boot}_{\widehat{\theta}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\widehat{\theta}^{*}(b) - \overline{\widehat{\theta}}^{*} \right) \left(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}}_{(-i)} \right)'$$

- It can be proven that $\widehat{\theta}^{*}(b)$ is converging in probability, and also in distribution to $\widehat{\theta}$.
- It was also proven that $\widehat{\mathbf{V}}^{boot}_{\widehat{\theta}} \xrightarrow{p^{*}} \mathbf{V}_{\widehat{\theta}}$
- It is suggested that trimmed (with extreme $\widehat{\theta}^{*}(b)$ deleted from the sample) bootstrap estimator of variance has better properties

Bootstrap confidence intervals

- Bootstrap can also be used to construct confidence intervals e.g. percentile t-interval

$$T^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{se\left(\widehat{\theta}^*\right)}$$

$$C^{pt} = \left[\widehat{\theta} - se\left(\widehat{\theta}\right)\widehat{q}^*_{1-\frac{\alpha}{2}}, \widehat{\theta} - se\left(\widehat{\theta}\right)\widehat{q}^*_{\frac{\alpha}{2}}\right]$$

where $\widehat{q}^*_{1-\frac{\alpha}{2}}$, $\widehat{q}^*_{\frac{\alpha}{2}}$ are quantiles of the bootstrap distribution.

- Percentile t-interval achieves an asymptotic refinement (converges to true values with the rate n rather than $\sqrt{n}$)

Bootstrap t-test

- Bootstrap p-values

$$p^* = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I} \left(\left| T^* \left(b \right) \right| > \left| T \right| \right)$$

where T is value of static, T^* is bootstrapped value of statistics, B is the number of bootstrap draws and p^* is the percentage of bootstrapped samples for which H_0 was rejected.

- We reject H_0 is p^* smaller that significance level
- Under suitable assumptions bootstrap test achieve asymptotic refinement
- Other statistics like Wald, LM etc. can also be bootstrapped

Endogeneity problem

- Structural equation
$$y = X'\beta + e$$
but the regressors exogeneity condition not satisfied:
$$\mathbb{E}\left[Xe\right] \neq 0$$
- In such a case we say that *regressors are endogenous*.
- If endogeneity is present OLS estimator is biased and inconsistent as linear projection coefficient estimated by OLS is not equal to structural parameter:

$$\beta^* = \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[XY\right] = \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[X\left(X'\beta + e\right)\right]$$
$$= \beta + \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[Xe\right] \neq \beta$$

- **Omitted variable problem** :
  - structural model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \beta_q q + u$$

  - say that q is not observable and therefore us omitted
  - as q is omitted from regression it has to be included in error term

$$u^* = \beta_q q + u$$

  - **But**: if q and X are correlated than X and u^* are correlated, $\mathbb{E}\left[Xu^*\right] \neq 0$ and X is endogenous

- **Measurement error**. If instead of Z we use approximate value X, then measurement error will become part of error term and u will generally be correlated with X
  - structural model
    $$y = Z'\beta + e$$
    assume that regressors are exogenous: $\mathbb{E}[Ze] = 0$
  - X measured value of Z
    $$X = Z + u$$
    assume that u is not correlated with e and Z: $\mathbb{E}[eu'] = 0$, $\mathbb{E}[Zu'] = 0$
  - then the relationship is linear with coefficient β
    $$y = (X - u)'\beta + e = X'\beta + v$$
    $$v = e - u'\beta$$
  - unfortunately regressors in this equation are endogenous
    $$\mathbb{E}[Xv] = \mathbb{E}[(Z + u)(e - u'\beta)] = -\mathbb{E}(uu')\beta \neq 0$$
  - generally speaking this kind of bias results in shrinking the estimates of β towards 0 (attenuation bias)

Choice variables as regressors.

- wages and education model

$$log\left(wage\right)=\beta_0+\beta_1 education+e$$

- more able workers (e.g. with respect to IQ, interpersonal skills) self select to be better educated, they are also more productive
- ability measures are not included in the regression,
- as a consequence of self selection in education the variable education and error term are correlated
- if both dependent variable and regressor are dependent on choices of economic agent then regressor should be treated as endogenous

Causes of endogeneity: simultaneity

Simultaneity. Simultaneity arises if dependent variable influences (e.g. with feedback) one of the explanatory variables

- supply and demand model

$$Q_D = \alpha_0 - \alpha_1 P + \varepsilon_D$$

$$Q_S = \beta_0 + \beta_1 P + \varepsilon_S$$

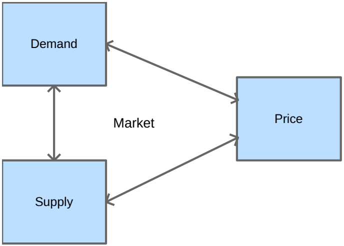
$$Q_S = Q_D$$

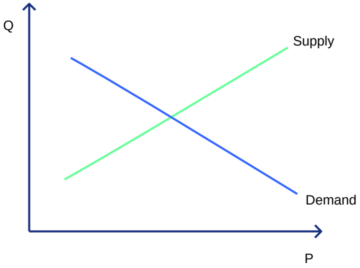
- solving this system of equations for price in equilibrium we obtain:

$$P = \frac{1}{\alpha_1 + \beta_1} \left(\alpha_0 - \beta_0 + \varepsilon_D - \varepsilon_S \right)$$

- but then the regressor P in demand and supply equation is correlated with error terms in this equations
- if both dependent variables and regressors are determined simultaneously in equilibrium then regressors should be treated as endogenous

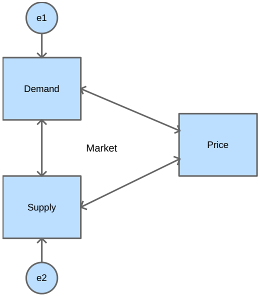
Supply and demand model, deterministic model

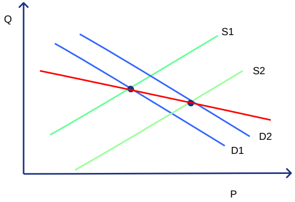




- This model explains the quantity demanded, quantity supplied and equilibrium price is determined
- Prices and quantities are set simultaneously
- According to macroeconomic theory, price elasticity of demand is negative, price elasticity of supply positive
- The values of elasticities should be estimated from data

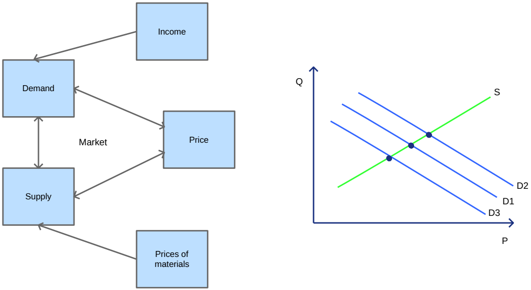
Supply and demand statistical model and identification problem





- Demand and supply are influenced by random factors e_1, e_2
- It is not possible to find the slopes of the demand/supply curves on the basis of the data on prices and quantities only!
- This case is related to so called identification problem (demand and supply curves are not identified)

Supply and demand model: instrumental variables



- Economic theory suggests that income of consumers is only influencing demand
- Prices of raw materials are only influencing supply
- Say that we are able select observations for which only incomes varied but prices of raw materials are stable
- We are able to approximate supply curve!

Instrumental variable method

- One of the most popular methods used to solve the endogeneity problem is the instrumental variable method.
- This method is based on using in the processes of estimation additional variables, so called instruments.
- Structural equation

$$Y = X'\beta + e$$

$$\mathbb{E}[Xe] \neq 0$$

- Random vector of l instrumental variables Z has following properties:
  1. $\mathbb{E}[Ze] = 0$ (instruments uncorrelated with structural error)
  2. $\mathbb{E}[ZZ'] > 0$ (instruments are not collinear)
  3. $rank(\mathbb{E}[ZX']) = k$ (instruments are correlated with regressors)

Instrumental variable estimator

- Multiplying the structural equation by Z and taking expected values we obtain:

$$\mathbb{E}[ZY] = \mathbb{E}[ZX']\,\beta + \underbrace{\mathbb{E}[Ze]}_{0}$$

and then if $\mathbb{E}[ZX']$ is square matrix ($l = k$, number of instruments equal to number of regressors)

$$\beta = \left(\mathbb{E}[ZX']\right)^{-1}\mathbb{E}[ZY]$$

- We say that β is identified as it can be expressed as a function of data moments
- Using analogy principle we replace expected values with sample moments and obtain IV estimator:

$$\widetilde{\beta}_{iv} = \left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}\sum_{i=1}^{n} Z_i Y_i = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}$$

2 Sage Least Squares Estimator

- If number of instruments is larger than number of regressors ($l > k$) we use 2 Stage Least Squares estimator (2SLS) which is also known as GIV (Generalized Instrumental Variable) estimator
- It is based on obtaining linear projections of X from regressions of X on Z

$$\widehat{X}' = Z'\left(\left(\mathbb{E}\left[ZZ'\right]\right)^{-1}\mathbb{E}\left(ZX\right)\right)$$

- Notice that

$$\mathbb{E}\left(\widehat{X}e\right) = \left(\left(\mathbb{E}\left[ZZ'\right]\right)^{-1}\mathbb{E}\left(ZX\right)\right)'\mathbb{E}\left(Ze\right) = 0$$

so that $\widehat{X}$ is exogenous

- Therefore

$$\beta = \left(\mathbb{E}\left[\widehat{X}\widehat{X}'\right]\right)^{-1}\mathbb{E}\left[\widehat{X}Y\right]$$

and using analogy principle we obtain the 2SLS estimator

$$\widetilde{\beta}_{2SLS} = \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}$$

Card 1995 article - collage proximity as instrument for education

Table 12.1: Instrumental Variable Wage Regressions

	OLS	IV(a)	IV(b)	2SLS(a)	2SLS(b)	LIML
education	0.074	0.132	0.133	0.161	0.160	0.164
	(0.004)	(0.049)	(0.051)	(0.040)	(0.041)	(0.042)
experience	0.084	0.107	0.056	0.119	0.047	0.120
	(0.007)	(0.021)	(0.026)	(0.018)	(0.025)	(0.019)
$\text{experience}^2/100$	−0.224	−0.228	−0.080	−0.231	−0.032	−0.231
	(0.032)	(0.035)	(0.133)	(0.037)	(0.127)	(0.037)
Black	−0.190	−0.131	−0.103	−0.102	−0.064	−0.099
	(0.017)	(0.051)	(0.075)	(0.044)	(0.061)	(0.045)
south	−0.125	−0.105	−0.098	−0.095	−0.086	−0.094
	(0.015)	(0.023)	(0.0284)	(0.022)	(0.026)	(0.022)
urban	0.161	0.131	0.108	0.116	0.083	0.115
	(0.015)	(0.030)	(0.049)	(0.026)	(0.041)	(0.027)
Sargan				0.82	0.52	0.82
p-value				0.37	0.47	0.37

Notes:

1. IV(a) uses *college* as an instrument for *education*.
2. IV(b) uses *college*, *age*, and $age^2/100$ as instruments for *education*, *experience*, and $experience^2/100$.
3. 2SLS(a) uses *public* and *private* as instruments for *education*.
4. 2SLS(b) uses *public*, *private*, *age*, and age^2 as instruments for *education*, *experience*, and $experience^2/100$.
5. LIML uses *public* and *private* as instruments for *education*.

Source: Card (1995)

Difference in differences estimator

- Simplest case
  - year 1 and 2 (no treatment, treatment)
  - control group A (no treatment in year 2)
  - treatment group B (treatment in year 2)
- dummy variable $dB = \begin{cases} 0 & \text{if in group } A \\ 1 & \text{if in group } B \end{cases}$
- dummy variable $d2 = \begin{cases} 0 & \text{if year 1} \\ 1 & \text{if year 2} \end{cases}$
- Simplest equation

$$y = \beta_0 + \delta_0 d2 + \beta_1 dB + \delta_1 d2 \cdot dB + u$$

- δ_0 captures the effect of year for both groups
- β_1 captures the permanent differences between control and treatment group
- δ_1 captures the effect of the treatment (in second year and only for treatment group)

Difference in differences estimator

- It can be proven that the estimator of δ_1 is equal to:

$$\widehat{\delta}_1 = (\overline{y}_{B.2} - \overline{y}_{B.1}) - (\overline{y}_{A.2} - \overline{y}_{A.1})$$

- Interpretation:
  - effect of the treatment is calculated as the change in y observed for the treatment group.
  - effect of the year is controlled for by subtracting the same change calculated for control group.
- Hence the name difference in differences (DID) estimator
- It is possible make this model more complicated by taking into account additional control variables
- Requirement for consistency: treatment not related to factors that affect y and are not observed

Mayer, Viscusi, Durbin 1995 "Workers' Compensation and Injury Duration"

In 1980 Kentucky rised the cap on the weekly earning covered by workers' compensation. This change did not affect low-wage workers (below the old cap). Control group: low-wage workers. Treatment group: high-wage workers. Question: what is the effect of compensation on the duration of stay out of work. Regression:

$$\log(\text{durat}) = \underset{(0.031)}{1.126} + \underset{(.0447)}{.0077}\,\text{afchnge} + \underset{(.047)}{.256}\,\text{highearn} + \underset{(.069)}{.191}\,\text{afchnge}\cdot\text{highearn}$$

$N = 5626 \qquad R^2 = .021$

Result: average duration of the stay increased by $\widehat{\delta}_1 = 19\%$ due to higher cap. No tendency to longer stay for both groups: coefficient of afchnge insignificant. High earners has a tendency to stay on compensation about $\left(\exp\left(\widehat{\beta}_1\right) - 1\right) = 29.2\%$ longer than low earners.

Card, David (1995). “Using geographic variation in college proximity to estimate the return to schooling”. In: *Essays in Honour of John Vanderkamp*. Ed. by E. K. Grant L. N. Christofides and R. Swidinsky. University of Toronto Press.