Advanced Econometrics

University of Warsaw

Jerzy Mycielski

Doctoral School of Social Sciences, 2026

Random sample.

- Sample is the set $\{(Y_i, X_i) : i = 1, ..., n\}$ of n realizations of the random variables (Y, X)
- The variables (Y_i, X_i) are a **sample** from the distribution F if they are identically distributed with distribution F
- The variables (Y_i, X_i) are a **random sample** if they are mutually independent and identically distributed (i.i.d.) across $i = 1, ..., n$.
- The sample have this properties if the process of selecting the sample from the population satisfies the following conditions:
  - every member of the population have the same probability of being drawn to the sample
  - the probabilities of being drawn from the population are independent: observations are independent

Moment estimators

- Assume that elements of the sample are identically distributed and are draws from common distribution F (random sample)
- We refer to the underlying distribution F as the population distribution or Data Generating Process (DGP)
- The simplest estimators are based on moments, we obtain them by replacing population moments by sample moments
- E.g. expected value μ and variance σ^2 of Y can be estimated as follows:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{n}, \ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{\mu})^2}{n}$$

  as $var(Y) = \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2$
- Alternative derivation of OLS estimator is based on replacing $\mathbf{Q}_{XX} = \mathbb{E}(XX')$ and $\mathbf{Q}_{XY} = \mathbb{E}(XY)$ by $\widehat{\mathbf{Q}}_{XX} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i'$ and $\widehat{\mathbf{Q}}_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$ so that linear projection coefficient is estimated as

$$\widehat{\beta} = \widehat{\mathbf{Q}}_{XX}^{-1}\widehat{\mathbf{Q}}_{XY}$$

Method of Moments (MM)

- Moment equations
$$\mathbb{E}\left[g_i\left(\beta\right)\right] = 0$$
for $i = 1, \ldots, n$
- The dimension of β (number of parameters) is k
- The dimension of $g_i\left(\beta\right)$ is m.
- If $k = m$ parameter β is just identified
- If $k < m$ parameters β is not identified: β cannot be estimated
- If $k > m$ is overidentified: GMM estimator should be used

Generalised Method of Moments (GMM): just identified case

- Define

$$\overline{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)$$

- For just identified model MM estimator sets this equation to zero

$$\overline{g}_n\left(\widehat{\beta}_{MM}\right) = \frac{1}{n}\sum_{i=1}^{n} g_i\left(\widehat{\beta}_{MM}\right) = 0$$

- Examples of MM estimators

1. Mean $g_i = Y_i - \mu$
2. OLS $g_i = X_i\varepsilon_i = X_i(y_i - X_i\beta)$
3. Simple IV $g_i = Z_i\varepsilon_i = Z_i(y_i - X_i\beta)$

Generalized Method of Moments: overidentified case

- In overidentified case it is not possible to find the exact solution of the moment equations in the sample
- Define the following objective function

$$J(\beta) = n\overline{g}_n(\beta)' \mathbf{W} \overline{g}_n(\beta)$$

where $\mathbf{W}$ is some positive definite weighting matrix.

- $J(\beta)$ can be interpreted as a distance between $\overline{g}_n(\beta)$ and 0
- GMM estimator is defined as

$$\beta_{GMM} = \underset{\beta}{\operatorname{argmin}} J(\beta)$$

- Asymptotic distribution of GMM estimator

$$\sqrt{n}\left(\widehat{\beta}_{GMM} - \beta\right) \xrightarrow{d} N(0, \mathbf{V}_\beta)$$

$$\mathbf{V}_\beta = (\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1}(\mathbf{Q}'\mathbf{W}\Omega\mathbf{W}\mathbf{Q})(\mathbf{Q}'\mathbf{W}\mathbf{Q})^{-1}$$

with $\Omega = \mathbb{E}\left[g_i(\beta) g_i(\beta)'\right]$, $\mathbf{Q} = \mathbb{E}\left[\frac{\partial}{\partial\beta'} g_i(\beta)\right]$ estimated as $\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^n g_i\left(\widehat{\beta}\right) g_i\left(\widehat{\beta}\right)'$, $\widehat{\mathbf{Q}} = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\beta'} g_i\left(\widehat{\beta}\right)$.

Efficient GMM estimator

- In some cases weighting matrix is treated as known
- It can be proven that the variance of the GMM estimator is asymptotically the smallest if $\mathbf{W} = \Omega^{-1}$ (optimal weighting matrix)
- In such a case $\mathbf{V}_\beta = \left(\mathbf{Q}'\Omega^{-1}\mathbf{Q}\right)^{-1}$
- It can be proven as well that in this case GMM estimator is also efficient in the sense that in the class of semiparametric estimators based only on moment equations it has the smallest variance
- In most of the cases matrix Ω is a function on unknown parameters
- In such cases we use the 2 step GMM estimator:
  1. Estimate of β using some arbitrary weighting matrix (e.g. $\mathbf{W} = \mathbf{I}$)
  2. Obtain the estimate of $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$ and the efficient GMM using $J(\beta) = n\bar{g}_n(\beta)' \widehat{\mathbf{W}} \bar{g}_n(\beta)$

Testing in GMM

- GMM estimator for linear model is equivalent 2SLS (IV) estimator in homoscedastic case
- In heteroscedastic case GMM estimator is more efficient that 2SLS estimator
- The weighting and covariance matrix of GMM can be modified to take into account clustered dependence
- In the context of GMM we can also define the analogues of the Wald, LR (distance) and LM statistics
- The test which is specific for GMM is the Sargan test for validity of overidentifing restrictions
- It can be proven that if moment equations are valid then

$$J\left(\widehat{\beta}_{GMM}\right) \xrightarrow{d} \chi_{k-m}$$

- The null $H_0 : \mathbb{E}\left[g_i\left(\beta\right)\right] = 0$ and $H_1 : \mathbb{E}\left[g_i\left(\beta\right)\right] \neq 0$
- The test can only be used if model is overidentified $k > m$.
- Rejection of H_0 suggests that some of the moment equations are false

Example: Ordinary Least Squares Estimator (IV)

- Model

$$y = X'\beta + \varepsilon$$

and $\mathbb{E}\left(X\varepsilon\right) = 0$ (exogeneity)

- Moment equations

$$\mathbb{E}\left[X\varepsilon\right] = \mathbb{E}\left[X\left(y - X'\beta\right)\right] = 0$$

- Sample analogue

$$\frac{1}{n}\sum_{i=1}^{n} X\left(y - X'\beta\right) = \frac{1}{n}\left(\mathbf{X}'y + \mathbf{X}'\mathbf{X}\beta\right) = 0$$

- Solution: OLS estimator

$$\widehat{\beta}_{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'y$$

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

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

- supply and demand model

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

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

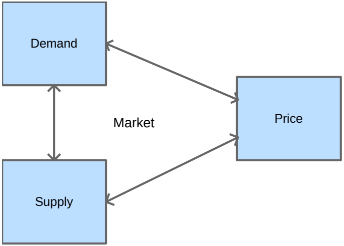
$$Q_S = Q_D$$

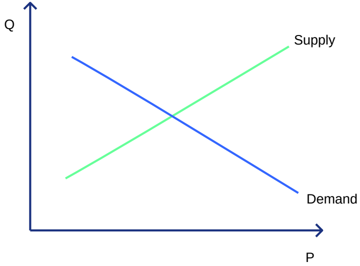
- solving this system of equations for price in equilibrium we obtain:

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

- 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

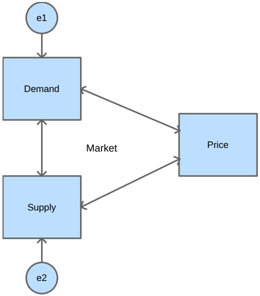
Supply and demand model, deterministic model

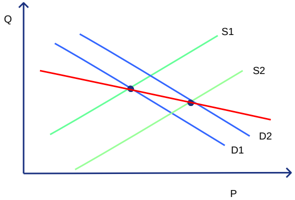




- This model explains the quantity demanded, quantity supplied and equilibrium price is determined
- Prices and quantities are set simultaneously
- According to macroeconomic theory, price elasticity of demand is negative, price elasticity of supply positive
- The values of elasticities should be estimated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

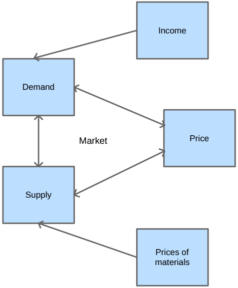
Supply and demand statistical model and identification problem

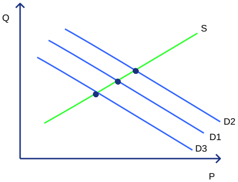




- Demand and supply are influenced by random factors e_1, e_2
- It is not possible to find the slopes of the demand/supply curves on the basis of the data on prices and quantities only!
- This case is related to so called identification problem (demand and supply curves are not identified)

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:

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

Special case: Wald (1940) estimator

- Consider the case of the model of Y with one edogenous explanatory variable X and one binary instrument Z

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- It can be shown that in this case the IV estimator is equal to

$$\widehat{\beta} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{X}_1 - \overline{X}_0}$$

  where $\overline{Y}_1$ and $\overline{X}_1$ are means of Y_i and X_i for $Z = 1$ and $\overline{Y}_0$ and $\overline{X}_0$ are means for $Z = 0$
- Intuition:
  - Z causes :
    - directly the change of X from $\overline{X}_0$ to $\overline{X}_1$
    - indirectly (through$X$) the change of Y from Y_0 to Y_1.
- Therefore the causal effect can be calculated as the ratio of the effect of Z on Y to the effect of Z on X

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) 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 $\mathbb{E}\left(\widehat{X}\left(Y - X'\beta\right)\right) = 0$ and solving for β gives:

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

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)
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)

Exogeneity testing

- Structural model

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

- Reduced form model

$$X = \Gamma' Z + u$$

- Linear projection of e on u:

$$e = u'\alpha + \nu$$

- Structural Model with control function

$$y = X'\beta + u'\alpha + \nu$$

and potentially, X is endogenous

IV estimation with control function

- β can be consistently estimated as follows:

1. obtain residuals $\widehat{u}$ from regression of X on Z
2. estimate β by regressing y on X and $\widehat{u}$

- In this regressing we are controlling the correlation between X and e with control function $\widehat{u}$
- It can be shown that resulting estimator is consistent and algebraically equivalent to IV estimator
- If α is signifcant then X is endogenous

Hausman-Wu test for endogeneity

- Consider following hypotheses
  - $H_0 : \mathbb{E}[Xe] = 0$
  - $H_1 : \mathbb{E}[Xe] \neq 0$
- Notice that $\mathbb{E}[Xe] = 0$ only if $\alpha = 0$
- Therefore, we can for exogeneity, by testing the parametric hypothesis
  - $H_0 : \alpha = 0$
  - $H_1 : \alpha \neq 0$
- If e, u have normal distribution then the F statistic for this hypothesis have in small sample F distribution.
- Otherwise we can use Wald test with asymptotic $\chi^2_{K_1}$ distribution, where K_1 is the number of of explanatory variables tested for endogeneity.
- Traditional way of testing this hypothesis is based on comparing the estimates obtained with OLS and IV

$$\left(\widehat{\beta}_{IV} - \widehat{\beta}_{OLS}\right)\left(\widehat{\Sigma}_{IV} - \widehat{\Sigma}_{OLS}\right)^{-1}\left(\widehat{\beta}_{IV} - \widehat{\beta}_{OLS}\right) \xrightarrow{d} \chi^2_{K_1}$$

- If instruments are weak (small correlation with endogenous variable) than small sample distributions of test statistics can be heavily distorted
- Consider simplest model with one endogenous variable and one instrument ($k_2 = 1$, $\ell_2 = 1$)

$$\begin{aligned} Y &= X\beta + e \\ X &= Z\Gamma + u_2 \end{aligned}, \quad (e, u_2) \sim N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

- It was shown that the t-statistic for hypothesis $H_0 : \beta = 0$ has the asymptotic distribution:

$$T = \frac{\widehat{\beta}_{IV}}{\widehat{se}\left(\widehat{\beta}_{IV}\right)} \xrightarrow{d} \frac{\xi_1}{\sqrt{1 - 2\rho\frac{\xi_1}{\mu+\xi_2} + \left(\frac{\xi_1}{\mu+\xi_2}\right)^2}} \overset{def}{=} S$$

where $\Gamma = n^{-1/2}\mu$ and (ξ_1, ξ_2) is bivariate normal. Distribution S is obviously not gaussian except for $\mu \to \infty$.
- Notice that here small μ is small in proportion to $n^{-1/2}$.

Testing for weak instruments

- We choose the value of ρ for which the distortion is the most serious ($\rho = 1$)
- For this case formula for distribution S simplifies to

$$S = \xi \left| 1 + \frac{\xi}{\mu} \right|, \ \xi \sim N(0, 1)$$

- We choose such $\mu^2 = \tau^2$ that for $\mu^2 > \tau^2$

$$Pr\left(|S| \geq t_{\alpha}\right) \leq \alpha^*$$

  where α is a nominal significance level for $H_0 : \beta = 0$ and α^* is the worst case significance level when testing the $H_0 : \beta = 0$.
- If this condition is satisfied even in the worst case the actual size of the test of $H_0 : \beta = 0$ does not exceed α^*.

Testing for weak instruments

- The idea of Stock and Yogo (2005) is to test $H_0 : \mu^2 = \tau^2$ against $H_1 : \mu^2 > \tau^2$ using F-test statistic

$$F = \frac{\widehat{\gamma}^2}{se\left(\widehat{\gamma}^2\right)} \xrightarrow{d} \chi_1^2\left(\tau^2\right)$$

where $\chi_1^2\left(\tau^2\right)$ is non-central χ^2 distribution with non-centrality parameter τ^2 . Critical value c selected so that

$$Pr\left(F > c|\, \mu^2 = \tau^2\right) \rightarrow P\left(\chi^2\left(\tau^2\right) > c\right) = \alpha$$

- The values of c for If we rejecting $H_0 : \mu^2 = \tau^2$ then maximum size distortion of the test of hypothesis $H_0 : \beta = 0$ on nominal significance level α is smaller or equal to $\alpha^* + \alpha$ (Bonferroni correction)
- The values of c for general case $(k_2 \geq 1,\ \ell_2 \geq k_2)$ can be found in Stock and Yogo (2005)

Critical values of Stock and Yogo (2005) test

Table 12.5: 5% Critical Value for Weak Instruments, $k_2 = 2$

	Maximal Size r							
	2SLS				LIML			
ℓ_2	0.10	0.15	0.20	0.25	0.10	0.15	0.20	0.25
2	7.0	4.6	3.9	3.6	7.0	4.6	3.9	3.6
3	13.4	8.2	6.4	5.4	5.4	3.8	3.3	3.1
4	16.9	9.9	7.5	6.3	4.7	3.4	3.0	2.8
5	19.4	11.2	8.4	6.9	4.3	3.1	2.8	2.6
6	21.7	12.3	9.1	7.4	4.1	2.9	2.6	2.5
7	23.7	13.3	9.8	7.9	3.9	2.8	2.5	2.4
8	25.6	14.3	10.4	8.4	3.8	2.7	2.4	2.3
9	27.5	15.2	11.0	8.8	3.7	2.7	2.4	2.2
10	29.3	16.2	11.6	9.3	3.6	2.6	2.3	2.1
15	38.0	20.6	14.6	11.6	3.5	2.4	2.1	2.0
20	46.6	25.0	17.6	13.8	3.6	2.4	2.0	1.9
25	55.1	29.3	20.6	16.1	3.6	2.4	1.97	1.8
30	63.5	33.6	23.5	18.3	4.1	2.4	1.95	1.7

Source: Hansen (2022)

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.

Stock, James and Motohiro Yogo (2005). “Asymptotic distributions of instrumental variables statistics with many instruments”. In: *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* 6, pp. 109–120.

Wald, Abraham (1940). “The Fitting of Straight Lines if Both Variables are Subject to Error”. In: *The Annals of Mathematical Statistics* 11.3, pp. 284–300. ISSN: 00034851. URL: http://www.jstor.org/stable/2235677 (visited on 03/30/2026).