Estimating the effect of financial on economic outcomes using IV

some examples

Rob Alessie

June 28, 2023

Outline

1. IV estimation : a short review
   - Assumptions
   - (Generalized) Method of Moments
   - Hansen (1982) test overidentifying restrictions
   - Two stages least squares
   - Bias of 2SLS (GMM)
   - The C-test for endogeneity
2. Fort, Manaresi, Trucchi (2016)
3. Lewbel (2012)

IV estimation: assumptions

- **Assumption 1 (linearity):** The equation to be estimated is linear:

$$y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \varepsilon_i \tag{1}$$

where $\boldsymbol{x}_i$ is an K-dimensional row vector of **regressors**

- **Assumption 3.2 (Instrument exogeneity):** All the R variables in the **instrument** vector $\boldsymbol{z}_i$ are orthogonal to the current error term:

$$E[\boldsymbol{z}_i' \varepsilon_i] = E[\boldsymbol{z}_i'(y_i - \boldsymbol{x}_i \boldsymbol{\beta})] = \boldsymbol{0} \tag{2}$$

- **Predetermined regressors**: regressors included in $\boldsymbol{z}_i$.
- **Endogenous regressors**: regressors not included in $\boldsymbol{z}_i$.
- **Excluded instruments**: instruments which are no regressors.

IV estimation: general framework (II)

- **Assumption 3 (rank condition for identification):** The $R \times K$ matrix $\boldsymbol{\Sigma}_{zx} = E(\mathbf{z}_i'\mathbf{x}_i)$ is of full column rank.
- The orthogonality conditions (2) can be rewritten as follows:

$$E(\mathbf{z}_i'\varepsilon_i) = E(\mathbf{z}_i'y_i) - E(\mathbf{z}_i'\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{0} \text{ or } \boldsymbol{\Sigma}_{zx}\boldsymbol{\beta} = \boldsymbol{\sigma}_{zy} \tag{3}$$

- (3) is a system of R equations from which one can solve the K unknown element of the parameter vector $\boldsymbol{\beta}$.
- This system has an unique solution if $\boldsymbol{\Sigma}_{zx}$ is of full column rank (cf. assumption 3.4).
- Since rank$(\boldsymbol{\Sigma}_{zx}) < K$ if $R < K$, a *necessary* condition for identification is

$$R((\text{no. predetermined variables}) \geq K((\text{no. regressors}) \tag{4}$$

Order condition for identification

- Since rank$(\boldsymbol{\Sigma}_{zx}) < K$ if $R < K$, a *necessary* condition for identification is

$$R((\text{no. predetermined variables}) \geq K((\text{no. regressors}) \tag{5}$$

- Depending on whether the order condition is satisfied, we say that the equation is
  - **overidentified** if rank condition 3 is satisfied and $R > K$
  - **just identified** if rank condition 3 is satisfied and $R = K$
  - **not identified** if rank condition 3 is NOT satisfied or $R < K$

(Generalized) Methods of Moments (G)MM

- Orthogonality conditions: $E\boldsymbol{z}_i'\varepsilon_i = \boldsymbol{0}$
- **Method of moments**: choose the parameter estimate $\tilde{\beta}$ so that

$$\boldsymbol{g}_n(\tilde{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i'\tilde{\varepsilon}_i = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i'(y_i - \boldsymbol{x}_i\tilde{\beta}) = \boldsymbol{0} \tag{6}$$

- Notice that (6) can be rewritten as follows:

$$\boldsymbol{S}_{zx}\tilde{\beta} = \boldsymbol{s}_{zy} \tag{7}$$

  where $\boldsymbol{S}_{zx} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i'\boldsymbol{x}_i$ and $\boldsymbol{s}_{zy} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i'y_i$

- If the model is just identified, the matrix $\boldsymbol{\Sigma}_{zx}$ (cf. eq. 3) is square and invertible. Assuming a random sample, it holds that

$$plim\ \ \boldsymbol{S}_{zx} = \boldsymbol{\Sigma}_{zx} \text{ and } plim\ \ \boldsymbol{s}_{zy} = \boldsymbol{\sigma}_{zx}$$

- Solving system (7) yields the *Instrumental Variables* (IV) estimator:

$$\hat{\beta}_{IV} = \boldsymbol{S}_{zx}^{-1}\boldsymbol{s}_{zy} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i'\boldsymbol{x}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i'y_i = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y} \tag{8}$$

(Generalized) Methods of moments (II)

- Under assumptions 1 (linearity), assumption 2 (instrument exogeneity), and 3 (rank condition for identification), one can use the law of large number to show that the IV estimator is consistent for $\boldsymbol{\beta}$.
- Extra assumption needed to show that the IV estimator is asymptotically normally distributed
  - **Assumption 4:** $\boldsymbol{S} = E(\boldsymbol{z}_i' \varepsilon_i^2 \boldsymbol{z}_i)$ exists and is finite.

- If the equation is **overidentified** $(R > K)$, we cannot choose an K-dimensional $\tilde{\boldsymbol{\beta}}$ to satisfy the R equations in (7).
- In that case, use GMM (2SLS) which chooses $\tilde{\boldsymbol{\beta}}$ so that $\boldsymbol{g}_n(\tilde{\boldsymbol{\beta}})$ (cf. equation (6)) is as "close" to 0 as possible.

- The GMM estimator minimizes the distance between $\boldsymbol{g}_n(\tilde{\boldsymbol{\beta}})$ and the zero vector:

$$\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{W}}) = \underset{\tilde{\boldsymbol{\beta}}}{argmin} \;\; n \cdot \boldsymbol{g}_n(\tilde{\boldsymbol{\beta}})' \hat{\boldsymbol{W}} g_n(\tilde{\boldsymbol{\beta}}) \tag{9}$$

where $\hat{\boldsymbol{W}}$= weighting matrix (symmetric and positive definite).

Large-Sample Properties of GMM

- Equation (9) defines a set of GMM estimators, depending on $\hat{\boldsymbol{W}}$.
- This section:
  1. Large Sample Properties of GMM for given choice of $\hat{\boldsymbol{W}}$
  2. "Optimal" choice of $\hat{\boldsymbol{W}}$: which GMM estimator is asymptotically efficient?
- Ad 1: One can show that the GMM estimator is consistent and asymptotically normally distributed.
- Optimal choice of of $\hat{\boldsymbol{W}}$: A lower bound for the asymptotic variance of the GMM estimators is reached if

$$\hat{\boldsymbol{W}} = \hat{\boldsymbol{S}}^{-1} = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \boldsymbol{z}_i' \boldsymbol{z}_i \right)^{-1}$$

Large-Sample Properties of GMM

- **Two step estimation procedure**
  1. Choose a 'convenient' matrix $\hat{\boldsymbol{W}}$, e.g $\boldsymbol{W} = \boldsymbol{I}_R$ or $\hat{\boldsymbol{W}} = \boldsymbol{S}_{zz}^{-1}$ (**two stage least squares estimator**, see below)
  2. Use this 'initial estimate of $\boldsymbol{\beta}$ to compute $\hat{\varepsilon}_i$ and $\hat{\boldsymbol{S}}$
  3. Efficient GMM estimator

$$\hat{\beta}(\hat{\boldsymbol{S}}^{-1}) = (\boldsymbol{S}_{zx}'\hat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{zx})^{-1}\boldsymbol{S}_{zx}'\hat{\boldsymbol{S}}^{-1}\boldsymbol{s}_{zy} \tag{10}$$

Hansen (1982) test overidentifying restrictions

- If the equation is exactly identified ($R = K$) and $\boldsymbol{\Sigma}_{zx}$ of full column rank), then distance (cf. equation (9))

  $$J(\tilde{\beta}, \hat{\boldsymbol{W}}) = n \cdot \boldsymbol{g}_n(\tilde{\beta})' \hat{\boldsymbol{W}} \boldsymbol{g}_n(\tilde{\beta})$$

  is equal to zero **irrespective of the choice of $\hat{\boldsymbol{W}}$!**

- If the equation is overidentified, then the distance cannot be set to zero exactly, but we would expect the minimized distance to be close to zero.
  Based on the distance measure, Hansen (1982) has developed a test of "overidentifying restrictions". The null hypothesis is

  $$H_0 : E(\boldsymbol{g}_i) = E(\boldsymbol{z}_i' \varepsilon_i) = 0$$

- If $\hat{\boldsymbol{W}} = \hat{\boldsymbol{S}}^{-1}$, the minimized distance is χ^2 distributed with $R - K$ degrees of freedom.

Two stages least squares

- **Assumption 3.7 (conditional homoskedasticity)**

$$E(\varepsilon_i^2|\boldsymbol{z}_i) = \sigma^2$$

- Under conditional homoskedasticity

$$\boldsymbol{S} = E(\varepsilon_i^2 \boldsymbol{z}_i \boldsymbol{z}_i') = \boldsymbol{\sigma}^2 \boldsymbol{\Sigma}_{zz}$$

- Efficient GMM becomes 2SLS

$$\hat{\beta}_{2SLS} = \hat{\beta}(\hat{\boldsymbol{S}}_{zz}^{-1}) = [\boldsymbol{S}_{zx}'(\boldsymbol{S}_{zz})^{-1}\boldsymbol{S}_{zx}]^{-1}\boldsymbol{S}_{zx}'(\boldsymbol{S}_{zz})^{-1}\boldsymbol{s}_{zy} \tag{11}$$

which does not depend on $\hat{\sigma}^2$.

Two stages least squares (II)

- "Justification" for the name **Two-stage least squares (2SLS)** Without loss of generality I assume that there is only 1 endogenous regressor y_{2i} in the model.
  1. **First stage regression:** regress the endogenous regressor x_i on $\boldsymbol{z}_i$

     $$y_{2i} = \boldsymbol{z}_i' \pi + \nu_i \tag{12}$$

     and calculate the fitted value $\hat{x}_i$
  2. **Second stage regression:** regress the dependent variable y_i on constant and $\hat{y}_{2i}$ and the predetermined regressors.
- One can check the validity of the instrument relevance assumption by performing the first stage regression and perform a F-test on the **excluded instruments** (i.e. the variables in $\boldsymbol{z}_i$ which are not in $\boldsymbol{x}_i$).

Bias of 2SLS (GMM)

- Under GM assumptions, OLS estimator is consistent AND unbiased.
- The 2SLS estimator is consistent, but biased. In small samples, the bias can be considerable.
- The 2SLS estimator is most biased when
  - the instruments are "weak", meaning that the partial correlation between excluded instruments and the endogenous regressor is low.
  - there are many over-identifying restrictions.
- It can be shown show that in case of a single endogenous regressor

$$E(\hat{\beta}_{2SLS} - \beta) \approx \frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}\left[\frac{E(\boldsymbol{\pi}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\pi})/K}{\sigma_\nu^2} + 1\right]^{-1} = \frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}(F+1)^{-1} \tag{13}$$

- If $F \approx 0$ then the bias in the 2SLS estimate is almost equal to the bias in the OLS estimate ($\frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}$)
- The bias of 2SLS vanishes when F gets large.
- Rule of thumb (see Bound et al, 1992): $F > 10$.

The C-test for endogeneity

- The C statistic (also known as the "difference-in-Sargan" statistic) allows a test of the exogeneity of one or more instruments.
- It is defined as the difference of the Sargan-Hansen statistic of
  1. the equation with the full set of instruments (i.e., including the instruments whose validity is suspect) and
  2. the equation with the smaller set of instruments (valid under both the null and alternative hypotheses)
- Under the null hypothesis that both the smaller set of instruments and the additional, suspect instruments are valid, the C statistic is distributed as chi-squared in the number of instruments tested.

Fort, Manaresi and Trucchi (2016)

- Fort, Margherita, Francesco Manaresi, and Serena Trucchi. "Adult financial literacy and households' financial assets: the role of bank information policies." *Economic Policy* 31.88 (2016): 743-782.
- Italian Banks that entered the **PattiChiari** Consortium, had to provide to their customers information on basic economic concepts (e.g., about compound interests, fixed/flexible interest rates, and the calculation of debt repayment instalments) and more transparent information on current account expenditures.
- This info provided on regular basis in a easy accessible way (Internet and/or at local bank branch).
- Any bank can apply to be a member of the Consortium, provided that it undertakes the "Quality Commitments".
- Dataset: 2010 wave of SHIW
- FL questions on portfolio diversification, the risk associated to fixed or adjustable interest rate and the effect of inflation.
- Instrument: dummy indicating whether the 'main' bank of the household is part of the PattiChiari Consortium.

Empirical strategy

$$wealth_{ip} = \beta_0 + \beta_1 finlit_{ip} + \boldsymbol{x}_{ip}\boldsymbol{\beta}_2 + \gamma_p + \varepsilon_{ip} \tag{14}$$

$$finlit_{ip} = \alpha_0 + \alpha_1 Patti_{ip} + \boldsymbol{x}_{ip}\boldsymbol{\alpha}_2 + \delta_p + \nu_{ip} \tag{15}$$

$$wealth_{ip} = \theta_0 + \theta_1 Patti_{ip} + \boldsymbol{x}_{ip}\boldsymbol{\theta}_2 + \mu_p + \xi_{ip} \tag{16}$$

- The authors expect $\alpha_1 > 0$ and $\theta_1 > 0$ ('intention to treat' effect) and argue that the variable $Patti_{ip}$ is exogenous in both equations (15) and (16) (no self-self-selection into a *Pattichiari* bank).
- β_1 can be estimated consistently by IV.
- **LATE:** The IV estimate of β_1 applies to the subpopulation of compliers, i.e., individuals whose financial knowledge increases because of information policies carried out by banks.
- When both the instrumental variable and the treatment variable are binary, the first-stage estimate of the effect of the instrument on the endogenous regressor (α_1) (cf. table 3)

Summary statistics

Table 1. Descriptive statistics

Variable	Mean	Standard deviation	*N*
Financial literacy (number of correct answers out of three)	1.93	0.98	4,865
Financial literacy, binary indicator 1(three out of three answers correct)	0.34	0.47	4,865
Financial literacy, binary indicator 1(question on inflation correct)	0.74	0.44	4,865
Financial literacy, binary indicator 1(question on loan correct)	0.64	0.48	4,865
Financial literacy, binary indicator 1(question on portfolio diversification correct)	0.55	0.5	4,865
Financial literacy (number of correct answers out of two present in 2006 and 2008)	1.38	0.71	4,865
Instrumental variable, binary indicator 1(Household is *PattiChiari* client) (*Patti*)	0.73	0.45	4,865
Instrumental variable, binary indicator 1(Household is client of at least one *PattiChiari* bank)	0.75	0.43	4,865
Financial assets (thousands)	19.4	22.4	4,865

Baseline specification

Table 2. Baseline regressions

	OLS	FS	ITT	IV
Dependent variable	Fin. assets	Fin. lit.	Fin. assets	Fin. assets
Financial literacy	**2.231*****			**8.273*****
	(0.630)			**(2.979)**
Client of *PattiChiari*		**0.235*****	**1.941*****	
		(0.035)	**(0.723)**	
Age	0.832***	0.054***	0.939***	0.494**
	(0.136)	(0.006)	(0.134)	(0.197)
Age squared	−0.005***	−0.000***	−0.006***	−0.001
	(0.001)	(0.000)	(0.001)	(0.002)
Male	1.017	0.170***	1.347**	−0.063
	(0.642)	(0.025)	(0.652)	(0.705)
Number of observations	4865	4865	4865	4865
F-test on the excluded instruments				43.849

Notes: The table reports OLS, first-stage, intention-to-treat, and instrumental variable results of the model $w_{ip} = \tilde{\alpha} + \tilde{\beta}\mathrm{fl}_{ip} + X_{ip}\tilde{\gamma} + \lambda_p + \epsilon_{ip}$, where w_{ip} is the financial assets owned by household i, living in province p, in year 2010; fl_{ip} is the financial literacy of the household head, X_{ip} is a vector of household controls (age, age squared, gender, marital status, years of education of the household head and number of household components, household labour income in 2010), and λ_p is a province fixed effect. Financial literacy is instrumented with a dummy that takes the value 1 if the household's main bank belongs to the *PattiChiari* Consortium and 0 otherwise. Financial literacy is the number of correct answers out of the three questions asked in 2010. 2010 SHIW Data. Heteroskedasticity-robust standard errors clustered at the province level in parentheses.
$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Binary measure for financial literacy

Table 3. Robustness: different measures for financial literacy

	OLS	FS	ITT	IV
Dependent variable	Fin. assets	Fin. literacy	Fin. assets	Fin. assets
Number of answers correct out of 3				
Financial literacy (nb/3)	2.231***			8.273***
	(0.630)			(2.979)
Client of *PattiChiari*		0.235***	1.941***	
		(0.035)	(0.723)	
Number of observations	4,865	4,865	4,865	4,865
F-test on the excluded instr.				43.849
Binary indicator for 3 answers correct out of 3				
Financial literacy (3 correct)	4.512***			43.378**
	(1.086)			(20.156)
Client of *PattiChiari*		0.045***	1.941***	
		(0.016)	(0.723)	
Number of observations	4,865	4,865	4,865	4,865
F-test on the excluded instr.				8.152
Binary indicator for inflation question correct				
Inflation correct	1.921			19.096***
	(1.204)			(7.315)
Client of *PattiChiari*		0.102***	1.941***	
		(0.019)	(0.723)	
Number of observations	4,865	4,865	4,865	4,865
F-test on the excluded instr.				29.128
Binary indicator for loan question correct				
Loan correct	1.604**			27.207**
	(0.798)			(10.831)
Client of *PattiChiari*		0.071***	1.941***	
		(0.013)	(0.723)	
Number of observations	4,865	4,865	4,865	4,865
F-test on the excluded instr.				28.297
Binary indicator for portfolio question correct				
Portfolio correct	5.287***			31.491**
	(1.136)			(12.929)
Client of *PattiChiari*		0.062***	1.941***	
		(0.019)	(0.723)	

Different subgroups

Table 5. Heterogeneity in the effect of financial literacy on financial assets

	OLS	FS	ITT	IV
Dependent variable:	Fin. assets	Fin. literacy	Fin. assets	Fin. assets
Education ≤ median[a]				
Financial literacy	1.985***			9.426**
	(0.628)			(3.747)
Client of *PattiChiari*		0.256***	2.412**	
		(0.044)	(0.998)	
Number of observations	2,563	2,563	2,563	2,563
F-test on the excluded instr.				33.677
Education > median				
Financial literacy	3.034***			7.333
Client of *PattiChiari*		0.127***	0.928	
		(0.046)	(1.110)	
Number of observations	2,122	2,122	2,122	2,122
F-test on the excluded instr.				7.542
Age ≤ 60[b]				
Financial literacy	1.544**			5.154
	(0.769)			(6.261)
Client of *PattiChiari*		0.167***	0.860	
		(0.044)	(1.096)	
Number of observations	2,419	2,419	2,419	2,419
F-test on the excluded instr.				14.507
Age 60+ years				
Financial literacy	2.665***			10.309**
	(0.709)			(4.032)
Client of *PattiChiari*		0.256***	2.635**	
		(0.048)	(1.121)	
Number of observations	2,367	2,367	2,367	2,367
F-test on the excluded instr.				28.385
Low education, age 60+ years				
Financial literacy	1.722**			12.595**
	(0.808)			(5.153)
Client of *PattiChiari*		0.232***	2.918**	
		(0.052)	(1.204)	
Number of observations	1,647	1,647	1,647	1,647

Lewbel (2012)

- Lewbel, A. (2012), "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models", Journal of Business & Economics Statistics, 30(1), pp. 67-80.
- This article provides a new method of identifying structural parameters in models with endogenous or mismeasured regres- sors.
- The method may be used in applications where other sources of identification, such as excluded instrumental variables, and repeated measurements are not available.

Lewbel (2012)

- Consider the following **triangular** system of equations

$$y_1 = \boldsymbol{x}\boldsymbol{\beta}_1 + y_2\gamma_1 + \varepsilon_1 \tag{17a}$$

$$y_2 = \boldsymbol{x}\boldsymbol{\beta}_2 + \varepsilon_2 \tag{17b}$$

- Traditionally, this model would be identified by imposing equality constraints on $\boldsymbol{\beta}_1$, e.g. $\beta_{1j} = 0$. Then x_j can be viewed as an excluded instrument
- The model is also identified if we assume $\mathrm{cov}(\varepsilon_1, \varepsilon_2) = 0$ (recursive system).
- Assumptions
  1. $E(\boldsymbol{x}'\varepsilon_1) = \mathbf{0}$ (orthogonality condition), $E(\boldsymbol{x}'\varepsilon_2) = \mathbf{0}$ and for some random vector $\boldsymbol{z}$, $\mathrm{cov}(\boldsymbol{z}, \varepsilon_1\varepsilon_2) = \mathbf{0}$
  2. $\mathrm{cov}(\boldsymbol{z}, \varepsilon_2^2) \neq \mathbf{0}$
- The elements of $\boldsymbol{z}$ can be discrete or continuous, and $\boldsymbol{z}$ can be a vector or a scalar.
- Some or all of the elements of $\boldsymbol{z}$ can also be elements of $\boldsymbol{x}$.

Lewbel (2012)

- The parameters $\boldsymbol{\beta}_1$, γ_1 and $\boldsymbol{\beta}_2$ can be estimated by means of GMM based on the following moment conditions

$$E(\boldsymbol{x}'\varepsilon_1) = \mathbf{0} \tag{18a}$$

$$E(\boldsymbol{z} - E(\boldsymbol{z}))'\varepsilon_2\varepsilon_1 = \mathbf{0} \tag{18b}$$

$$E(\boldsymbol{x}'\varepsilon_2) = \mathbf{0} \tag{18c}$$

- Estimation strategy:
  - Estimate β_2 by running the first stage regression (17b) (cf. moment condition 18c)
  - Compute the residuals of the first-stage regression $\hat{\varepsilon}_{2i} = y_{2i} - \boldsymbol{x}_i\hat{\boldsymbol{\beta}}_2$
  - Generate the 'excluded' instruments $(\boldsymbol{z}_i - \bar{\boldsymbol{z}})'\hat{\varepsilon}_{2i}$ (cf. moment condition 18b).
  - The parameters $\boldsymbol{\beta}_1$ and γ_1 can be estimated by using a standard IV regression (e.g the Stata routine ivreg2.
  - See example in file smpjfe2012.log. In the example the method does not work well (rejection overidentifying restrictions)