Estimating intergenerational income mobility on two samples: sensitivity to model selection

Francesco Bloise (Uni. Rome), Paolo Brunori (Uni. Florence), Patrizio Piraino (UCT)

Paris – 6th December 2018

Intergenerational mobility

- Relationship between the adult economic success of children and their family background
- Do poor children become poor adults? Do rich children become rich adults?
- How can we obtain a synthetic measure of intergenerational mobility/persistence in economic status?

“Economic” mobility

- Earnings or Income as typical dimensions along which to characterize one’s status.
- Fathers and sons as representative of the succeeding generations
- Now more common to look at mothers and daughters

Linear Regression Approach

$$Y_{ci} = \alpha + \beta\, Y_{pi} + u_{ci}$$

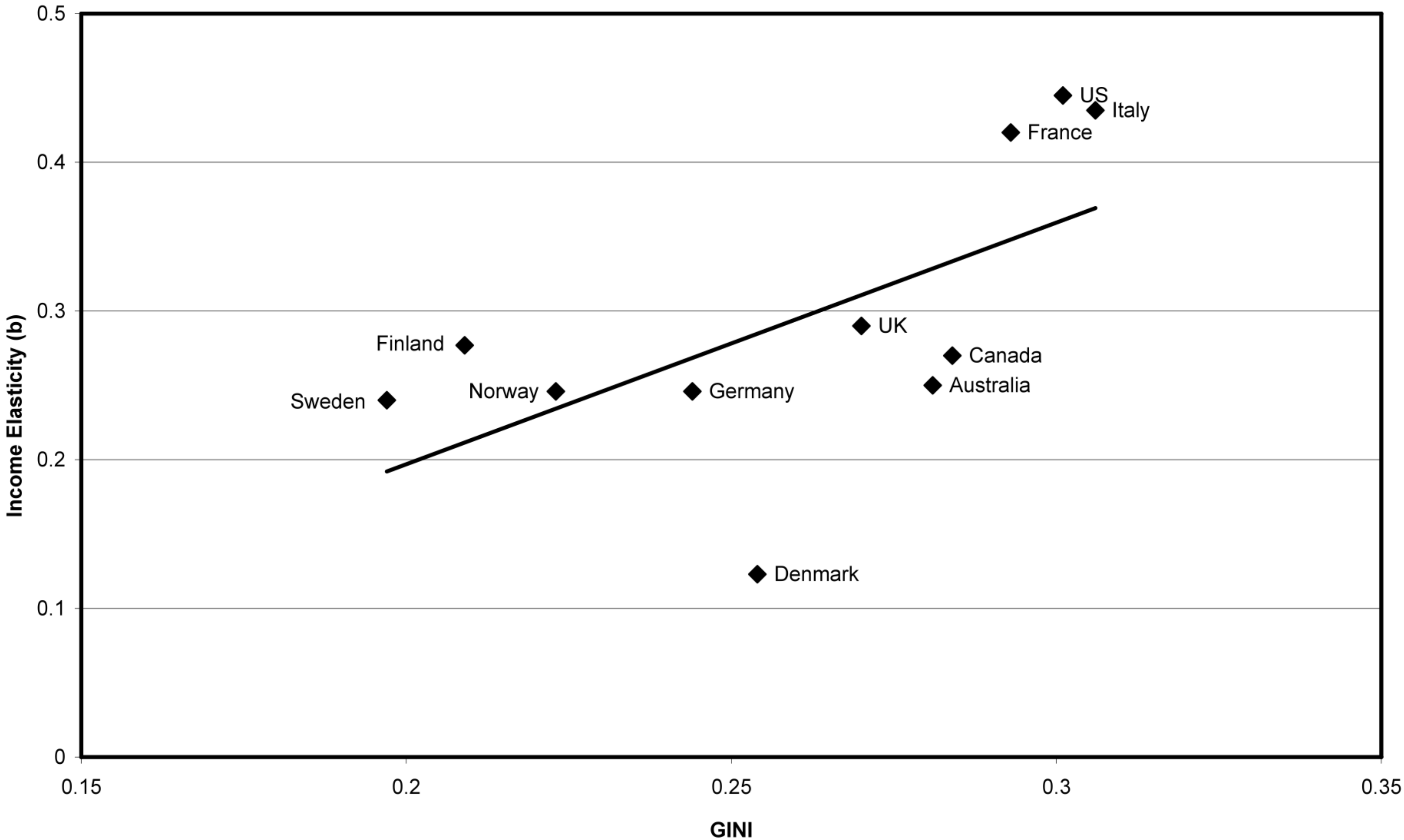
Y is income/earnings and u an error term.

β is intergenerational elasticity: measure the change in the child income associated with changes in parent's income

Summary statistic, "**overall**" measure

- $\beta = 0 \Rightarrow$ complete mobility: child status is independent of parental background
- $\beta = 1 \Rightarrow$ complete immobility: child status is fully explained by parents

Figure 1. Estimates of Intergenerational Income Elasticities for Fathers and Sons Plotted with GINI Coefficients for Eleven Developed Countries during the Early 1980s.



Data provided by M. Jantti from Figure 20.1, Bjorklund, A., and M. Jantti. 2009. Intergenerational Income Mobility and the Role of Family Background. In W. Salverda, et al. (eds.), The Oxford Handbook of Economic Inequality. Oxford: OUP.

Data requirements

- Protracted longitudinal survey
- Nationally representative
- Accurate measures of <u>permanent</u> income for both generations

Estimation Issues

- Early literature (for the US): intergenerational elasticity of earnings at about 0.2
- Much of this work was based on single-year measures of earnings → downward bias
  - Mazumder (2005) → 0.6
- A considerable number of studies have followed raising other relevant issues
  - Lifecycle bias
  - IV

Repeated cross-section

- Many countries do not have longitudinal datasets.

- If we want to say something about mobility around the world we need an alternative strategy.

- Retrospective information in repeated cross-sections may be exploited.

Two Sample Two Stages Least Squares

- Two-stage regression based on two samples:
  - children who have reported their parents' socio-economic characteristics
  - sample of adults (pseudo-parents) from a dataset representative of the same population.

- Then:
  1. estimate an income equation from the older sample;
  2. use the estimated coefficients to predict parents' income on the base of recall variables;
  3. regress child income on the predicted parental income.

Two Sample Two Stages Least Squares

First stage $y_{it}^{ps} = \gamma z_i^{ps} + \vartheta_{it}$

Second stage $y_{it}^{c} = a + \beta \hat{y}_i^{p} + \omega_{it}$

where $\hat{y}_i^{p} = \hat{\gamma} z_i^{p}$

Bias

- TSTSLS may produce upwardly biased estimates of the ‘true elasticity”.

- Variables used in the first stage are typically positively related to child income, independently of parental income.

$$y_i^c = \gamma_1 y_i^p + \gamma_2 \hat{y}_i^p + u_i$$

- probability limit of the TSTSLS estimator is (Nicoletti and Ermisch, 2008; Jerrim et al, 2016):

$$plim\ \beta_{TSTSLS} = \beta + \gamma_2 (1 - R^2)$$

- where β is the unbiased IGE and R^2 is from the first-stage model

Bias

- Two sources of bias in TSTSLS :
  - direct effect on child income of the first-stage predictors: γ_2
  - incorrect prediction of parental income $1 - R^2$.

- Most studies using the TSTSLS are not explicit about these two sources of bias
  - choice of the model is largely determined by data availability
  - several IGE estimates based on different combinations of first-stage predictors are then presented as robustness checks

Our method

- For a given sample, $(1-R^2)$ can always be minimised using a sufficient number of parameters

- However, an overfitted model will very imprecisely predict out-of-sample

- Our aim is to maximize the ability of the model to predict earnings of ‘unseen’ fathers

Method

- Elastic-net shrinkage operator (Zou and Hastie, 2005)

$$\sum_{i=1}^{n}\left(y_i - \beta_o - \beta_1 X_{1,i} - \beta_2 X_{2,i} \ldots - \beta_k X_{k,i}\right) + \lambda\left(\alpha \sum_{j=1}^{k} |\,\beta_j| + (1-\alpha) \sum_{j=1}^{k} {\beta_j}^2\right)$$

- This operator is a weighted average of two standard operators in machine learning: Lasso (least absolute shrinkage and selection operator) and ridge regression.

- Using elastic net we obtain different set of βs depending on λ and α.
  - algorithm needs to "tune" λ and α so that equation above is minimized.
  - a standard method to tune elastic nets is k-folds cross-validation

- We use cross-validation to iteratively search α and λ that minimize MSE
  - this specification will be based on a subset of regressors and shrunk coefficients
  - by selecting parameters that produce the smallest MSE, we aim to minimize (1 – R^2)

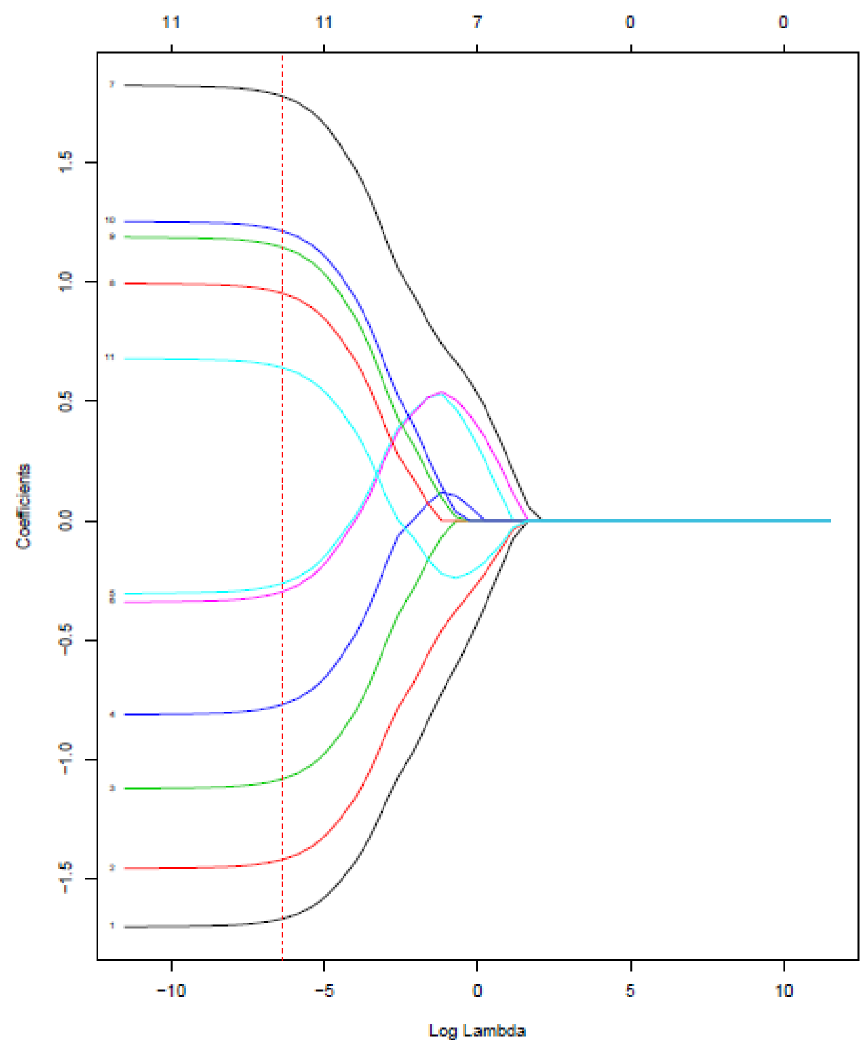
Empirical application

- We use survey data from South Africa.
  - same data and sample selection rules as in Piraino (2015)
  - analysis is restricted to males only

- focus on two parsimonious specifications of the 1st-stage equation:
  - parental education
  - parental education + occupation.

- Due to data availability, these are the two most common specifications used in the literature, particularly in low- and middle-income countries

Results



Results

Two-sample intergenerational income elasticity

	First-stage variables	
	(1)	(2)
	Education	Education & Occupation
OLS	$\beta = 0.621$ *s.e.* = 0.059 *MSE* = 1.128	$\beta = 0.626$ *s.e.* = 0.079 *MSE* = 1.018
Elastic net (base model)	$\beta = 0.622$ *s.e.* = 0.059 *MSE* = 1.128	$\beta = 0.635$ *s.e.* = 0.075 *MSE* = 1.018
Elastic Net (Full model)		***β = 0.578*** *s.e. = 0.076* ***MSE = 0.980***
N	2,587	1,442

Remarks

- For a fixed set of variables included in the prediction equation, the method will reduce the bias in the IGE
  - unless the model concurrently increases γ_2 so much to offset the improvement in $(1-R^2)$
  - useful result for researchers who are constrained to work with only one or two recall variables
  - since our method does not simply add parameters, γ_2 is less likely to increase

- Event when one can choose among multiple variables, this approach provides less arbitrary robustness tests for alternative sets of predictors