An Agent-Based Model for tertiary educational choices in Italy

Silvia Leoni

University of Leicester

AISSEC Conference
October 9, 2020

Tertiary education in Italy

- Low attainement:
  - in 2017 the 26.5% of aged 30-34 has obtained a higher education degree; a lower share registered only in Romania (26.3%) (Eurostat);
  - second lowest attainment among adults after Mexico; 18% of 25-64 years old graduated (OECD, *Education at a Glance 2017*).
- Not attractive: scarse job perspectives and low economic return to tertiary education (OECD).
- Evidence by Naticchioni et al., 2016: the generation from 1975-1979 suffered a remarkable earning loss at first job market entry, with respect to previous generations; larger effect for high-skilled rather than low-educated workers.

Research question and literature

Research question

- Analyse the determinants behind university enrolment decisions and explore whether these determinants could explain the low educational attainment characterising Italy.
- This work introduces an ABM of how individual educational preferences form with the aim to verify if, simulated over t periods of time, the model is able to provide a realistic representation of the socio-economic phenomenon investigated.

A close example to this work is Manzo, 2013: among the determinants of the distribution of educational choices across social groups, social influence cannot be ignored.

The Model

Agents

- *Juniors* vs *Seniors*
- *Juniors* are provided with a monetary endowment, $X_{i,t}$, deriving from their parental income (bequest), which can be spent in education or in the transition from school to the job market
- Agents also own a certain *ability* capturing innate talent and personal skills
- *Juniors* will ponder the choice to enrol at university if the following budget constraint is satisfied:

$$X_{i,t} - CostEdu > 0$$

Environment

Social map with N neighbourhoods representing agent's social relations: family, friends, acquaintances, structured in social circles (Hamill & Gilbert, 2009)

- The circumference of a circle will contain all those points within a distance set by a radius and creates a cut-off, limiting the size of personal networks.

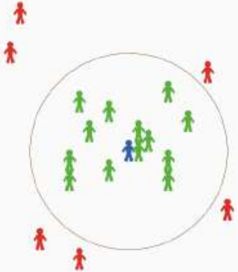


Figure 1: An example of social circle.

Environment

- One reach: agents are only permitted to link with agents who can reciprocate, i.e. others whose reach includes ego.

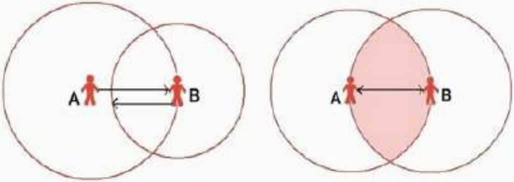


Figure 2: Reciprocity in social circles

Preference for enrolling

Building on Manzo (2013), agents enrol at university with a probability increasing in the level of preference P,

$$Pr_{it}(enroll) = \frac{exp(P_{it})}{1 + exp(P_{it})}$$

where

Preference for enrolling

$$P_{it} = \underbrace{ln\left(\frac{C^{e}_{is,t+1}}{C^{e}_{iun,t+1}}\right)}_{economic\ motivation} + \underbrace{SI_{it}}_{Social\ influence} - \underbrace{EF}_{effort}$$

Preference for enrolling

Expectations on future consumption

Real expected consumption for skilled workers at time $t+1$ will be:

$$C^e_{is,t+1} = X_{i,t} - CostEdu + Y^e_{is,t+1}$$

while real consumption for unskilled workers will be:

$$C^e_{iu,t+1} = X_{i,t} + Y^e_{iu,t+1}$$

Expectations on future income are modelled as naive expectations based on the information set of *senior* neighbours:

$$Y^e_{i,t+1} = E_t(Y_{t+1}|\Omega_{n,t}) = \frac{\sum_{i=1}^{n} w_i Y_{nsen,t}}{\sum_{i=1}^{n} w_i}$$

$w \rightarrow$weight assigned to agents' parents

The agent will compare the expected consumption in the two cases by taking the natural logarithm of their ratio:

$$ln\left(\frac{C^e_{is,t+1}}{C^e_{iu,t+1}}\right) = ln\left(\frac{X_{i,t} - CostEdu + Y^e_{is,t+1}}{X_{i,t} + Y^e_{iu,t+1}}\right)$$

Preference for enrolling

Social Influence

It reflects a merely imitative behaviour; SI can be considered as a measure of "educational conformism"

$$SI_{it} = \frac{Npeersenr}{N_{npeers}}$$

Effort of Education

Effort necessary to obtain a university degree, assumed to depend only on individual ability

$$EF = (1 - a_{it})^{\gamma}$$

- a_{it} measures individual ability
- $\gamma > 0$ measures the concavity of the function$\rightarrow$*returns to scale*

The Model

Simulating the model

Initialisation and calibration

At the set-up, $t = 0$

- 250 agents
- random age starting from 21
- agents aged between 30 and 40 *hatch* a child (*junior*) according to *birth rate*$= 0.4$
- Greeen/blue→Skilled/Unskilled
- Big/small→Senior/Junior
- Juniors are allowed to move 7 steps away to differentiate their neighbourhood from the one of their parent

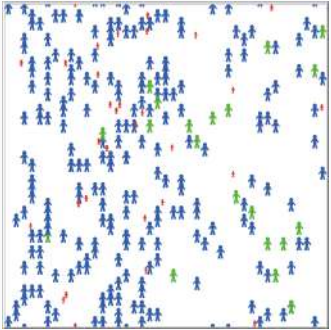


Figure 3: The NetLogo world of the model at its set-up.

Initialisation and calibration

Variable	Inputs and calibration
N. senior agents	250
N. steps	7
Social reach	10
Probability of segregation p_s	0.5
Proportion skilled/unskilled	9% (SHIW Bank of Italy 2002-2016)
Endowment	9% (Istat, last two years)
Cost of education	5000€ per year (Federconsumatori 2017)
Average working life	32 years (Eurostat 2016)
Ability	$\sim N(0.5, 0.1)$ (Breen & Goldthorpe, 1997)
γ	1.2 (Staffolani & Valentini, 2007)
Income distribution	Skilled $\sim \mathit{Lognormal}(9.97, 0.85)$ ▸ Estimates
	Unskilled $\sim \mathit{Lognormal}(9.46, 0.92)$

Table 1: Variables' values used in the model set-up

The Model

Model dynamic

Model dynamic

Iterate the model 100 t. For each t:

- agents' age is updated by adding 5 years
- juniors aged 20 and satisfying the budget constraint:
  - observe income of neighbours and compute their preference for enrolling
  - make a decision about education
  - change consequently their educational and generational status (agents "grow up")
  - obtain a skilled/unskilled income
- seniors *hatch* and die according to birth rate and death rate (0.9) to smooth population dynamic

Run 100 Monte Carlo experiments

The Model

Computational results

Summing up

Overview of the model

Input	Income Cost of education	Proportion S/U Personal Network	Segregation Family effect
Model	Economic motivation	Social Influence	Effort
Output	% enrolling	% Enrol. from S % Enrol. from U	% Immobile % Upward % Downward

Results

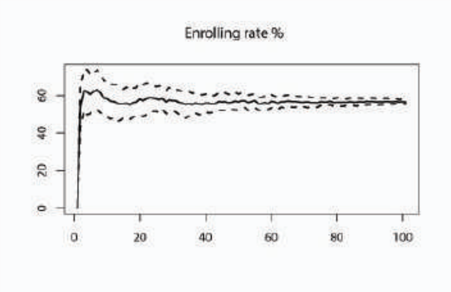
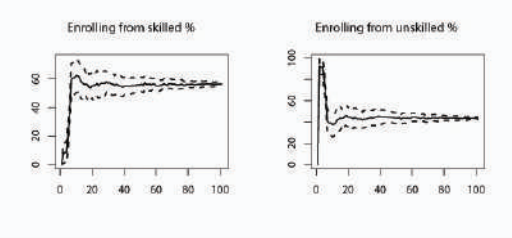


Figure 5: Average rate (%) of agents enrolling to university, % of enrolling from skilled/unskilled family (*continuous line*) and corresponding standard deviation (*dashed line*) over the simulations time span



Results

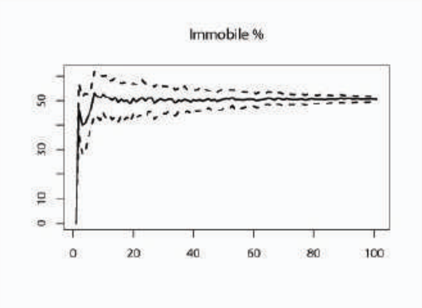
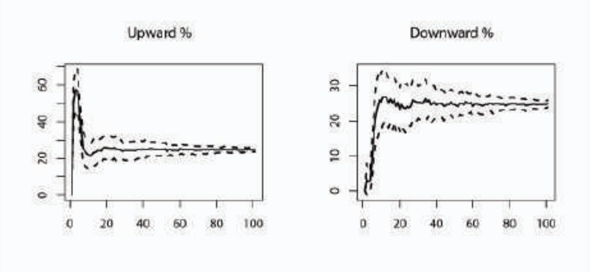


Figure 6: Average rate (%) of agents remaining immobile, movign upward or downward across the educational levels (*continuous line*) and corresponding standard deviation (*dashed line*)



Results

- The average enrolling rate is about 56%. ▸ Results
- Among those deciding to continue their studies, the majority, although slight (56.19%) comes from parents owing a university degree.
- About half of those deciding will maintain the same educational level than their parents, while the rest will evenly advance or remain behind the education level of their family

▸ back

Year	Rate of transfer from secondary education %
2008	65.8
2009	63
2010	63.3
2011	61.3
2012	58.2
2013	55.7

Table 2: Rate of transfer of Italian students from secondary school to university, in percentage. Source: Istat

The Model

Sensitivity Analysis

Sensitivity Analysis

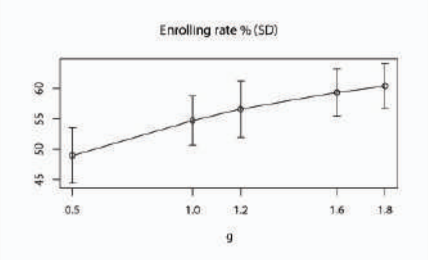
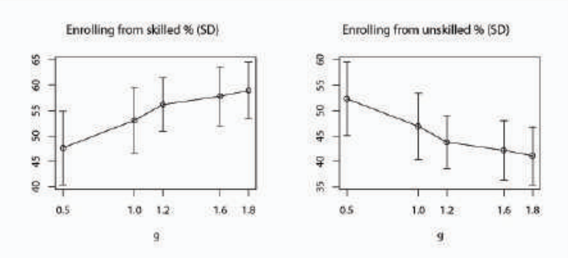
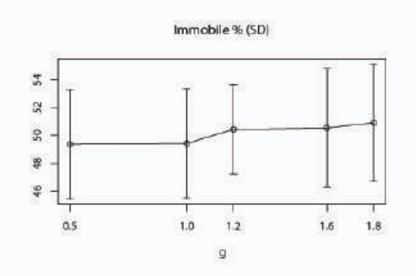


Figure 7: Synthetic statistics for the sensitivity experiment performed on γ



Sensitivity Analysis



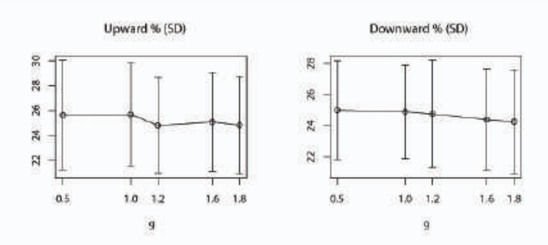


Figure 8: Synthetic statistics for the sensitivity experiment performed on γ

The Model

A calibration experiment

A calibration experiment

- ABMs have proven to be able to reproduce economic and social features, observable in the real world (Cont, 2001; Manzo, 2013). Given this property, calibration and validation can play a major role;
- Empirical validation as the process of ensuring that an ABM is consistent with empirical data (Tesfatsion, 2006)
- This analysis deals input validation$\rightarrow$ requires that the exogenous inputs of the model are empirically meaningful and appropriate

Purpose

Provide a more accurate representation of Italian personal income in the calibration of the ABM developed, by fitting to the SHIW data employed three Beta-type distributional forms.

Beta-type distributions

- The generalised beta of the second kind (GB2) is a four-parameter distribution defined over the support $(0, \infty)$. Its probability density function (pdf) is given by:

$$f(x) = \frac{aqx^{a-1}}{b^a[1 + (x/b)^a]^{1+q}}, x > 0 \tag{1}$$

  with $a > 0$, $b > 0$, $p > 0$, $q > 0$ are the four parameters identifying the distribution, where
  - b is a *scale* parameter, which stretches and squeezes the distribution
  - a, p, q are the *shape* parameters, affecting the shape of the distribution.
- The Singh–Maddala distribution corresponds to the case of the GB2 distribution, with $p = 1$.
- The Dagum is a GB2 distribution with the shape parameter $q = 1$.

Beta-type distributions

Since their discovery (Burr, 1942; McDonald, 1984; Singh & Maddala, 1976), Beta-type distributions have been widely employed as income distributions (Brachmann et al., 1996; McDonald & Xu, 1995).

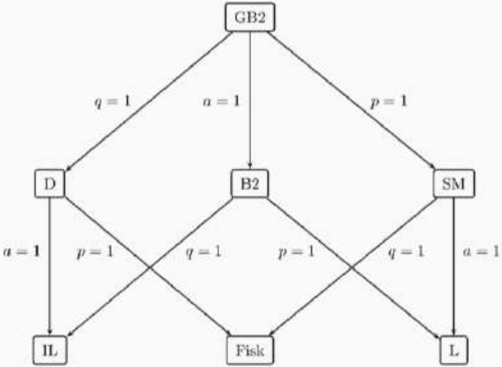
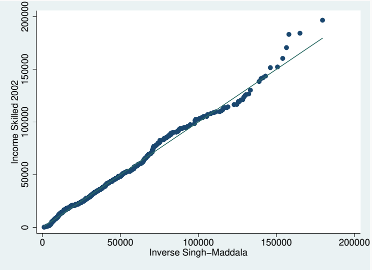


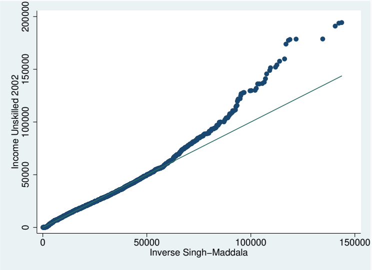
Figure 9: Source: Kleiber and Kotz (2003, p.188)

Fitting Italian income distribution

- SHIW Bank of Italy 2002-2016, 8 waves
- Each dataset has been split in two datasets: skilled vs unskilled
- Values have been converted to 2010 prices using the Istat household consumption deflator
- STATA packages `smfit`, `dmfit`, `gb2lfit`



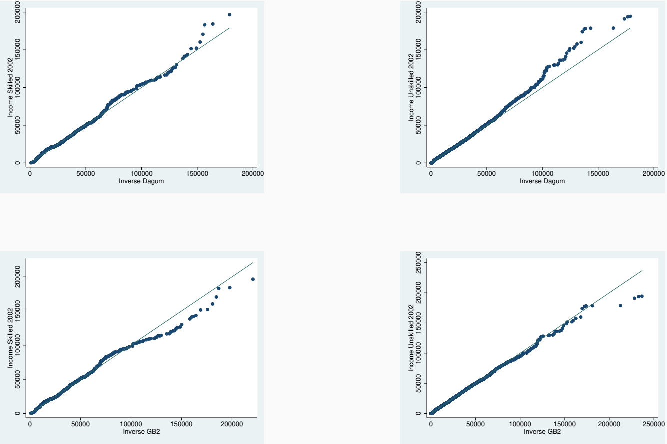
(a) Skilled



(b) Unskilled

Figure 10: Fit for the Singh-Maddala distribution, year 2002

Fitting the Italian income distribution



(a) Skilled **(b)** Unskilled

Figure 12: Fit for the Dagum (top) and GB2 (bottom) distributions, year 2002

Fitting the Italian income distribution

▸ Fit estimates ▸ GB2

	Skilled			Unskilled		
	AIC	BIC	LR	AIC	BIC	LR
Dagum	27432.09	27447.37	8.57^{**}	275969.1	275991.5	50.05^{***}
GB2	27425.51	27445.89		275921.0	275950.9	
S-M	27442.92	27458.2	19.41^{***}	276288.6	276311	369.58^{***}
GB2	27425.51	27445.89		275921.0	275950.9	
N		1205			12821	

Table 3: AIC, BIC and LR test for the year 2002

The Model

Model simulations

Model simulations

▸ Rate of transfer

The R extension in Netlogo is employed to run 30 Monte Carlo simulations. Outcomes are in line with the decreasing tendency observed in the rate of transfer from secondary to tertiary education.

Variable	Mean	SD
Enrolling rate %	53.61	0.60
Enrolling from skilled %	54.51	0.94
Enrolling from unskilled %	45.49	0.94
Immobile %	51.27	0.39
Upward %	24.38	0.31
Downward %	24.35	0.25

Table 4: Mean value and SD of the outcome variables computed across the simulations run.

Final remarks

- The modelling framework proposed showed that by taking into account the ability of agents to interact and adapt their choices to their neighbours, it is possible to reproduce features mirroring reality
- The experiment highlights the potential of both distribution (lognormal and GB2) to be a good candidate to describe the income distribution in the model, however the calibration experiment allows for a more rigorous imputation of parameter values
- Future development could introduce heterogeneity in the level of ability or add modules, such as Universities and geographical environment (*spatial analysis*)

Thank you

sl707@leicester.ac.uk

References i

Brachmann, K., Stich, A., & Trede, A. (1996). Evaluating parametric income distribution models. *Allgemeines Statistisches Archiv*, *80*, 285–298.

Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differential: Towards a Formal Rational Action Theory. *Rationality and Society*, *9*(3), 275–305.

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of mathematical statistics*, *13*(2), 215–232.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223–236.

Hamill, L., & Gilbert, N. (2009). Social Circles: A Simple Structure for Agent-Based Social Network Models. *12*(2).

References ii

- Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences* (Vol. 470). John Wiley & Sons.
- Manzo, G. (2013). Educational Choices and Social Interactions: A Formal Model and a Computational Test.
- McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 647–663.
- McDonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, *66*(1-2), 133–152.
- Naticchioni, P., Raitano, M., & Vittori, C. (2016). La Meglio Gioventù: earnings gaps across generations and skills in Italy. *Economia Politica*, *33*(2), 233–264.

Singh, S. K., & Maddala, G. S. (1976). A function for the size distribution of incomes.. *Econometrica*, *44*(5), 963–970.

Staffolani, S., & Valentini, E. (2007). Bequest taxation and efficient allocation of talents. *Economic Modelling*, *24*(4), 648–672.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. *Handbook of computational economics*, *2*, 831–880.

back

	Skilled			Unskilled		
	μ	σ	N	μ	σ	N
2002	10.27*** (0.0228)	0.791*** (0.0161)	1205	9.671*** (0.00739)	0.837*** (0.00523)	12281
2004	10.20*** (0.0248)	0.889*** (0.0176)	1283	9.625*** (0.00802)	0.901*** (0.00567)	12634
2006	10.13*** (0.0243)	0.894*** (0.0172)	1356	9.626*** (0.00752)	0.826*** (0.00531)	12063
2008	10.04*** (0.0231)	0.881*** (0.0163)	1457	9.533*** (0.00789)	0.872*** (0.00558)	12220
2010	10.09*** (0.0193)	0.788*** (0.0137)	1664	9.479*** (0.00907)	0.995*** (0.00641)	12040
2012	9.513*** (0.0188)	0.787*** (0.0133)	1750	9.184*** (0.00873)	0.951*** (0.00617)	11880
2014	9.591*** (0.0230)	0.964*** (0.0163)	1752	9.245*** (0.00921)	0.998*** (0.00651)	11759
2016	9.896*** (0.0217)	0.843*** (0.0154)	1506	9.344*** (0.00998)	1.015*** (0.00706)	10340

Table 5: Estimates of the parameters μ and σ of a lognormal distribution to Italian personal income. Significance levels are indicated as follows: $^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$; standard errors are reported in parentheses.

back

Variable	Mean	SD
enrolling rate %	56.61	0.51
enrolling from skilled %	56.19	0.74
enrolling from unskilled %	43.81	0.74
Immobile %	50.44	0.42
Upward %	24.80	0.28
Downward %	24.76	0.30

Table 6: Average values and corresponding standard deviations across 100 simulations, calculated over the time span 20-100

back

	Skilled			Unskilled		
	S-M	Dagum	GB2	S-M	Dagum	GB2
a	2.247	2.832***	4.515***	1.967***	3.616***	5.415***
	(0.0936)	(0.117)	(0.7905)	(0.0217)	(0.0542)	(0.3629)
b	35528.5***	36019.8***	31949.39***	31541.1***	25195.3***	22825.45***
	(2721.2)	(1673.0)	(1581.98)	(1028.0)	(283.6)	(333.04)
p		0.702***	0.417***		0.446***	0.2844***
		(0.0572)	(0.083)		(0.0109)	(.0213)
q	1.337***		0.496***	2.604***		0.551***
	(0.151)		(0.1164)	(0.117)		(0.048)
N	1205			12821		

Table 7: Parameters estimates for the fit on the year 2002

back

	$\hat{a}$		$\hat{b}$		$\hat{p}$		$\hat{q}$	
	S	U	S	U	S	U	S	U
2002	4.52	5.42	27156.89	19401.60	0.42	0.28	0.50	0.55
2004	5.04	6.81	30816.64	22427.11	0.32	0.21	0.40	0.44
2006	6.82	5.44	28294.48	22608.41	0.24	0.28	0.29	0.60
2008	4.03	6.95	27393.31	20348.92	0.40	0.21	0.55	0.44
2010	4.91	8.57	26742.75	21210	0.37	0.15	0.45	0.36
2012	5.38	9.86	17183.32	16280.64	0.32	0.13	0.45	0.32
2014	5.65	12.57	23415.56	19892.36	0.26	0.10	0.44	0.27
Mean	5.19	7.94	25857.56	20159.84	0.33	0.20	0.44	0.43

Table 8: Average parameters estimates for the GB2 distribution of Italian income for skilled and unskilled individuals over the years 2002-2014; data provided by the Bank of Italy (SHIW).