Unemployment expectations in an agent-based model with education*

Luca Gerotto[†] Paolo Pellizzari[‡]

Abstract

Why are unemployment expectations of the "man in the street" markedly different from professional forecasts? We present an agent-based model to explain this deep disconnection using boundedly rational agents with different levels of education. A good fit of empirical data is obtained under the assumptions that there is staggered update of information, agents update episodically their estimate and there is a fraction of households who always and stubbornly forecast that the unemployment is going to raise.

The model also sheds light on the role of education and suggests that more educated agents update their information more often and less obstinately fixate on the worst possible forecast.

Keywords: Agent-based modeling; Bounded rationality; Unemployment expectations.

1 Introduction

Why are unemployment expectations of the "man of the street" so strikingly different from the ones produced by professional economists or highly regarded institutions? The question is extremely important as the correct unemployment forecast is related to the quantification of the zero-income probability that was proven to be one of the most important drivers in saving decisions (Carroll [6]).

We examine in this work survey data of Italian households and show that their forecasts are only loosely related to or seemingly disconnected from professional estimates.[1]

*The final authenticated version [10] is available online at `https://doi.org/10.1007/978-3-319-94580-4_14`.

[†]Corresponding author, Department of Economics, Ca' Foscari University, Cannaregio 873, Fondamenta San Giobbe, 30121, Venice, Italy luca.gerotto@unive.it

[‡]Department of Economics, Ca' Foscari University, Cannaregio 873, Fondamenta San Giobbe, 30121, Venice, Italy paolop@unive.it

[1]By professional forecasts in this paper we refer to the figures made public by OECD every quarter. The use of a specific set of "professional" forecasts is not affecting the results sensibly as other forecasts, say produced by different research centers or governmental offices, are typically very similar and strongly correlated with the OECD data.

The fact that standard agents produce a vastly different set of forecasts when compared with professional data demonstrates once more how the assumption of perfect rationality of economic agents is flawed even if it is still present in the majority of scholarly work. Indeed, there is overwhelming evidence in experimental economics and psychology that shows the limits of human rationality (Rabin [12]) and how poorly humans behave in problems that involve complex situations (Camerer [4]). Realistic agents were shown to have troubles to gather and correctly internalize a lot of information in a probabilistic setup that are likely to be needed in the proper economic analysis of unemployment.

Clearly, it's also hard to believe that agents entirely fail to acquire information or are unable to grasp even the most basic elements driving the risk of unemployment. In what follows we assume that households are boundedly rational in the sense that their expectations are rational but based on wrong grounds due to their limited ability or willingness to observe (and absorb!) relevant information. The idea is similar, for instance, to Axtell and Epstein ([1]) where a handful of workers are rational in their retirement decisions whereas most of the population just adopts herding strategies imitating the behavior of the peers in the social network. Moro and Pellizzari [**?**] study the labour market participation where agents mimic (only) the ones with approximately the same age and health who appear to have higher satisfaction (or "utility") from work and leisure. Our framework was also inspired by the epidemiological metaphor used in Carroll [5], where the diffusion of information is depicted in pretty much the same way in which epidemics spread: only some agents are "infected" at time t by a common source (say, mass media coverage) and react accordingly, whereas the ones that were not infected keep using the obsolete information or act on the basis of spurious facts.

We assume that only the most informed fraction of the population acquires the relevant knowledge and uses a professional forecast. The remaining portion behaves as (very) naive econometricians (Branch [3]), projecting old trends into the future, or stubbornly and pessimistically keeps saying that unemployment is always going to increase in the future.

Departing notably from previous work, we also include in the model the dependence on the education of the agents, as it's likely that more (less) educated households are more (less) likely to get informed and use professional forecasts. We calibrate the model to empirical data fitting, for each education level, the probability to obtain the professional forecast as well as the probability to stubbornly declare that unemployment is going to increase.

Interestingly, our results show that observed forecasts are quite similar to the ones produced by the model. In addition, more educated agents more often get and use professional forecasts and less frequently exhibit stubborn pessimism. In this sense, our work shows a plausible way to reconcile data with a realistic behavior of the agents (who are not perfectly informed and fail to be fully rational at all times).

Our model is "explicit", as suggested in Epstein [9], it is sufficient to replicate the most salient features of the data and, even if we cannot claim it necessarily says the last word, it sheds light on reasons accounting for different forecasts. Our work is quite evidently related to the generative' approach to modelling and offers a tentative explanation as we can "grow" the phenomenon of interest, see Epstein [8].

The paper is organized as follows. Section 2 describes the data we have used and shows how *prima facie* survey observations appear to be very weakly aligned with professional forecasts. The agent-based model using education-dependent types is described in Section 3. We then present and discuss the results of a series of NetLogo simulations. Section 5 contains the conclusion and some final remarks.

2 The data

In this paper we make use of data gathered by the Italian National Institute of Statistics (ISTAT) on unemployment expectations. Specifically, ISTAT each month asks to a randomly selected sample of 2000 households the following question:

> How do you expect the number of people unemployed in Italy to change over the next 12 months?

Users have the option to provide an answer in the set {increase sharply, increase slightly, remain the same, decrease slightly, decrease sharply, don't know}. Answers are encoded using the values $\{+2, +1, 0, -1, -2, 0\}$, respectively.

Data are aggregated forming the average of the answers and multiplying by 100, effectively weighting twice as much the "sharp" replies with respect to the "slight" ones. Therefore, the balance[2] index has a theoretical range going from -200, if 100% of interviewees expect unemployment to *decrease* sharply in one year, to $+200$, if 100% of interviewees expect unemployment to *increase* sharply in one year. Moreover, sub-indexes that provide expectations for particular demographic categories are available. The comparison of the sub-indexes is not a negligible topic, since several empirical works suggest that the degree of financial literacy (Lusardi and Mitchell [11]) and expectations (Souleles [13] and Easaw et al. [?]) are related to the demographic characteristics of the individual, with a relevant role of the education level.

Professional forecasts are contained in the OECD Economic Outlook for Italy, which is released on a biannual basis.[3] In our analysis we use the forecasted change in the unemployment rate, measured as the difference between the forecasted unemployment rate in four quarters and the unemployment rate of the current quarter.

The top panel of Figure 1 represents the quarterly indexes for the three different education levels, namely *Less than High School, High School, College*. The plot shows that the three series are correlated but have different levels and that the range of values actually attained by the indexes in the last 22 years ($[-25, 115]$) is quite narrow with respect to the theoretical one ($[-200, 200]$). For example, the index for less educated households has never attained negative values meaning that, in this 22 years window, they never expected

[2] A "balance index" is constructed as the difference between the percentages of respondents giving positive and negative replies.

[3] From 2003, OECD releases forecasts also for each quarter, while for the period from 1995 to 2002 OECD releases forecasts only for each semester. Therefore until 2002 we estimate the missing quarters forecasts through interpolation of the biannual forecasts.

the unemployment rate to decrease. This is clearly not justified by neither OECD forecasts (Figure 1, in the bottom panel) nor ex-post realizations, since the 90 quarters we consider are split almost equally between periods of increasing and of decreasing unemployment, and the average value of the change in the unemployment rate is close to zero. Hence, the series of professionals' and households' forecasts have a strikingly different interpretation.

Figure 1: Above: Households unemployment expectations balance indexes by education level (1995Q1-2017Q2). Below: Forecasts of the one-year change in the unemployment rate, as released by the OECD (1995Q1-2017Q2)

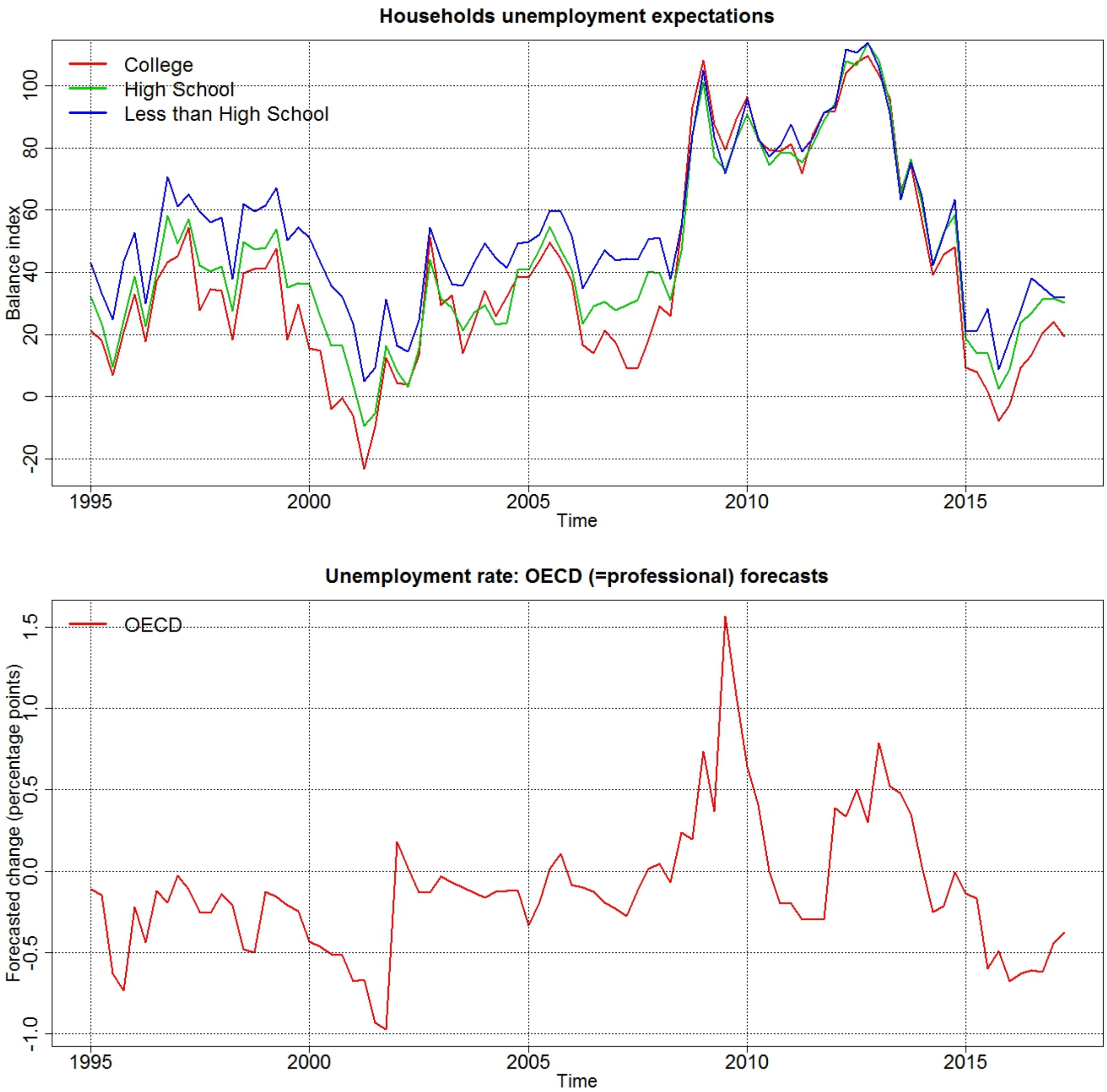


Table 1: Summary statistics (1995Q1-2017Q2)

	Min	1st Qu	Median	Mean	3rd Qu	Max	St.Dev
Less than High School	4.97	35.84	50.4	53.34	64.9	114	24.79
High School	-9.6	24.89	39.73	44.27	58.32	113.7	28.35
College	-23.47	15.77	32.7	39.09	54.32	109.5	32.39

Table 1 shows some descriptive statistics of ISTAT data. It is apparent that less educated individuals are more pessimistic (higher mean and median) and change less frequently their opinion (lower standard deviation). This stylized evidence, implying that the knowledge and use of relevant information (i.e., professional forecasts) differs as a function of the education level, will be incorporated in the model described in the next section.

3 The model

Assume there are N agents, whose education level is $edu_i \in E = \{lths, hs, col\}$, with the strings denoting *Less than High School, High School* and *College*, respectively.[4]

At (the beginning of) time t all agents observe a signal s_{it} defined as:

$$s_{it} = \begin{cases} o_t & \text{with probability } \lambda_i \\ s_{i(t-1)} & \text{with probability } \beta(1-\lambda_i) \\ 0 & \text{with probability } (1-\beta)(1-\lambda_i) \end{cases}, \tag{1}$$

where o_t is the latest professional forecast released by OECD, $\lambda_i = \lambda(edu_i)$ is the probability to obtain it and β is the probability to remember the past signal if one does not obtain the latest professional forecast. In other words, at each time step agents can either get an informative signal o_t, with some probability that depends on their education, or remember the old signal $s_{i(t-1)}$ if they get no fresh news, or forget even the old signal. Observe that, with probability $\beta(1-\lambda_i)$, the agent is not informed but keeps using the only item of obsolete information he owned in the past. If, say, $\beta = 0$ then any uninformed agent would pretend that unemployment will not change; at the other extreme, if $\beta = 1$, any uninformed agent would remember and make use of the old signal.

Once the (informative or uninformative) signal is available, agents convert it into an answer ans_{it} to be reported in the survey. Recall from Section 2 that answers are encoded with integers in $\{-2, -1, 0, +1, +2\}$ to mean a range going from "unemployment will decrease sharply" (-2) through "unemployment will remain the same" (0) to "unemployment will increase sharply" $(+2)$. We assume that agents stubbornly report $ans_{it} = 2$ with probability $\mu_i = \mu(edu_i)$; otherwise they "translate" their signal based on its perceived

[4] 2000 individuals are interviewed by ISTAT per month (6000 individuals per quarter). About $\frac{1}{2}$ of the Italian population belongs to the *lths* group, about $\frac{1}{3}$ to the *hs* group and about $\frac{1}{6}$ to the *Col* group. Hence, the baseline simulation has 3000 agents per quarter with *lths*, 2000 with *hs* and 1000 with *col*.

magnitude in the following way:

$$ans_{it} = \begin{cases} +2 & \text{with probability } \mu_i \\ f(s_{it}|\gamma) & \text{with probability } 1 - \mu_i \end{cases} . \tag{2}$$

The translation function f depends on a threshold $\gamma > 0$ that shapes 5 ranges of values driving the interpretation of the signal:

$$f(s_{it}|\gamma) = \begin{cases} +2 & \text{if } s_{it} \geq \gamma; \\ +1 & \text{if } \gamma/2 < s_{it} < \gamma; \\ 0 & \text{if } -\gamma/2 \leq s_{it} \leq \gamma/2; \\ -1 & \text{if } -\gamma < s_{it} < -\gamma/2; \\ -2 & \text{if } s_{it} \leq -\gamma. \end{cases} \tag{3}$$

In words, while the signal is very large, i.e., exceeding γ, the agent will claim that unemployment will increase sharply. If, instead, the signal is larger than $\gamma/2$ but smaller than γ, the milder conclusion is that "unemployment is going to increase slightly". Finally, if the signal is close to zero, in the sense that $-\gamma/2 \leq s_{it} \leq \gamma/2$, the answer will state that the rate is not changing (and so on, with obvious modifications for negative values of the signal s_{it}).

At the end of period t, when all $ans_{it}, i = 1, ..., N$, are available, it is straightforward to compute the aggregate index according to the way used by ISTAT:

$$x_t = \frac{100}{N} \sum_{i=1}^{N} ans_{it}.$$

Let y_t denote the balance index and let $y_t^{lths}, y_t^{hs}, y_t^{col}$ be the three subindexes relative to the different levels of education in E. Each run of the model produces a sequence of computed expectations $\{x_t, t = 1, ..., n\}$ and indeed the code can be thought of as an artificial data-generating process of surrogate data that can be compared with the empirically observed expectations $\{y_t, t = 1, ..., n\}$. The differences between the values of x_t and y_t clearly depend on the values of the parameters β and γ, which are constant for all agents, and of parameters λ and μ, which are heterogeneous and affected by the education level (*lths, hs, col*). If $\Theta^{edu} \in [0,1]^2 = (\lambda^{edu}, \mu^{edu})$ denotes the vector of the parameters specific to education level $edu \in E$, we can calibrate the model minimizing the fitting error on the related subindex y_t^{edu}.

There is a wide variety of possible and meaningful fitting criteria and, in order to obtain $\hat{\Theta}^{edu}$, we have used a very simple approach and have computationally minimized the sum of the relative deviations of the first two moments of the time-series x_t (generated by the model) and y_t^{edu} (by ISTAT):

$$\hat{\Theta}^{edu} = \arg\min_{\Theta^{edu}} \left(\left(\frac{E[x_t] - E[y_t^{edu}]}{E[y_t^{edu}]} \right)^2 + \left(\frac{V[x_t] - V[y_t^{edu}]}{V[y_t^{edu}]} \right)^2 \right), \tag{4}$$

where $E[\cdot]$ and $V[\cdot]$ are the usual mean and variance operators. A list and a brief description of the symbols used in the model are included in Table 2.

It is worth stressing that the model describes a tale in which boundedly rational agents, who seldom acquire and process the relevant data, generate a series of expectations that can be reconnected to some extent with those very same data. Indeed the "man in the street", despite its failure to use professional analysis immediately, makes episodic use of the data and resort to pessimistic forecasts from time to time, based on his level of education.

Table 2: List of symbols and acronyms

Symbol	Description
edu	Education level in the set $E = \{lths, hs, col\}$
o_t	Official professional forecast by OECD
y_t, y_t^{edu}	ISTAT balance index and subindex relative to $edu \in E$
s_{it}	Signal received by agent i at time t
ans_{it}	Survey answer provided by agent i at t
$f(s_{it})$	Translation function
β	Probability of remembering the past signal
λ^{edu}	Probability of absorbing the current official forecast
μ^{edu}	Fraction of stubbornly pessimistic households
γ	Parameter determining the qualitative translation of the signal

Acronym	Meaning
lths	*Less than High School*
hs	*High School*
col	*College*

In a more abstract interpretation, the model is just a filter taking as an input the OECD data and generating the "filtered data" published by ISTAT. Even though purely orthodox economists often expect the two sequences to be the same, this is clearly not the case. The filter, embedding scattered update of information, effects of education and use of bleak forecasts, shows that the observed data can be realistically linked with the professional estimates.

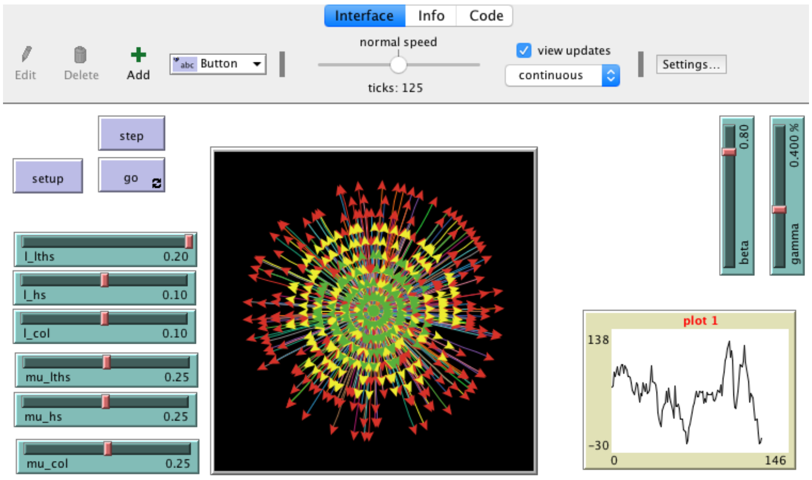
Next section will present the results obtained using the model with calibrated parameters.

4 Results

The model was implemented in NetLogo [14] a powerful and popular computational platform to develop agent-base models in an object-oriented setup with plenty of graphical facilities (download is free at `http://ccl.northwestern.edu/netlogo/`).

Essentially, the user can change the parameters of the model and obtain a simulated ISTAT time-series starting with the unique inputs provided by OECD every quarter. A screenshot of the model is visible in Figure 2.

Figure 2: A screenshot of the NetLogo model. Sliders can change the values of the parameters and the time series of the simulated data appear in the Plot window.



As a preliminary step, we calibrate the two parameters β and γ, which are common among the three education groups, over a grid of possible values. The calibration was obtained using NetLogo's `BehaviorSpace`, a tool allowing the efficient generation of data through the systematic exploration of the parameter space of the model and we picked the values $\beta = 0.9$ and $\gamma = 0.4$. Suppose, for instance, that an agent equipped with his individual μ receives the signal $s_t = o_t = 0.21$ at time t: as the signal is bigger than $\gamma/2$ (but smaller than γ), the answer will be $ans_t = 1$ with probability $1 - \mu$ or $ans_t = 2$ if he acts stubbornly with probability μ. Assume next that the agent fails to get the OECD update o_{t+1}. Then, with probability $\beta = 0.9$ he remembers the past signal: $s_{i(t+1)} = s_{it} = 0.21$, so the answer will again be $ans_{t+1} = 1$, or $ans_{t+1} = 2$ if he acts stubbornly. Conversely, with with probability $1 - \beta = 0.1$ he does not remember the past signal: in this case $s_{i(t+1)} = 0$ and the answer would be $ans_{t+1} = 0$ (or $ans_{t+1} = 2$ if he acts stubbornly).

Subsequently, we run separate calibrations for each education level, with the aim of obtaining the values of λ^{edu} and μ^{edu} as described by (4). Table 3 displays the calibrated parameters and illustrates how education influences the behavior of the agents. A look at the first column, reporting μ as a function of the education level, shows that the probability to act stubbornly visibly declines for more educated households (for example agents with a college degree have a μ about 20% smaller than their peers who did not complete the high school).

As far as the probability to acquire information is concerned, the more (less) educated agents more (less) often are infected by OECD data and the λ's are in the range 0.085 to 0.110 per quarter.[5] Even though the figures may look quite close, the effect is clearly detectable over a yearly period: less than one third among the agents whose education is

[5]In the interpretation of the results, it has to be remembered that some of the individuals (on average, a fraction $\lambda^{edu} \cdot \mu^{edu}$) receive the signal o_t but still provide a stubbornly pessimistic answer. This implies

Table 3: Calibrated values of μ, the probability to answer stubbornly, and λ, the probability to acquire the current OECD forecast.

	μ	λ
LTHS	0.309	0.085
HS	0.270	0.095
College	0.251	0.110

lower than high school degree come to know professional forecasts over one year whereas nearly 40% of agents equipped with a college degree acquire the information. Equivalently, agents in the *lths* group are acquainted on average with the professional forecasts once in three years, whereas the ones with a degree typically need approximately nine months less.

Table 4: Summary statistics - Simulated and Original (1995Q1-2017Q2)

	Min	1st Qu	Median	Mean	3rd Qu	Max	St.Dev	Cor
lths - Simulated	1.70	37.98	52.13	53.33	69.93	106.80	25.12	0.67
lths - Original	4.97	35.84	50.40	53.34	64.90	114.00	24.79	
hs - Simulated	-16.00	27.11	43.93	44.20	62.68	106.10	28.17	0.70
hs - Original	-9.60	24.89	39.73	44.27	58.32	113.70	28.35	
col - Simulated	-24.60	19.94	39.90	39.49	59.46	108.00	31.90	0.73
col - Original	-23.47	15.77	32.70	39.09	54.32	109.50	32.39	

Table 4 contrasts the distributions of the simulated and original data. The mean and the standard deviations are closely matched for any education level. Even more interestingly, the simulated data closely reproduce the minima, maxima and quartiles of the observed data. Differences are small and very rarely exceed 5 points (over a possible range of 400 points). The table also shows that the correlation between the real and the simulated indexes x_t and y_t^{edu} is about 70% and indeed turns out to be higher than the correlation between the index and the raw series of professional forecasts y_t^{edu} and o_t.

Figure 3 provides a visual comparison of the original and simulated index. Confirming the interpretation of Table 4, the fit is quite good especially in the first decade of the XXI century and, despite some misalignment around 2012, artificial data are reasonably close to their empirical target.

5 Conclusion

Our research question revolved around the observation that unemployment forecasts produced by surveyed "men in the street" are vastly different and, typically, stickier and gloomier, than the ones produced and disseminated by professionals. This is clearly evident

that the fraction of individuals who receive the signal o_t *and really* incorporate it into the answer ans_{it} is approximately $\lambda^{edu} \cdot (1 - \mu^{edu})$, that is 0.059 for *lths*, 0.070 for *hs* and 0.083 for *col*.

Figure 3: Households unemployment expectations balance (1995Q1-2017Q2)

contrasting the unemployment balance index provided by ISTAT with the data extracted from the Economic Outlook by OECD, that was used in this work to proxy the professional outcomes of rational experts and respected institutions.

We developed an agent-based model where the respondents to the survey are typified by different education, affecting their willingness or ability to use the professional forecast as well as the frequency with which they stubbornly claim that "unemployment is likely to raise". The acquisition of the professional signal and the way it is processed (or "translated", to use the terminology of Section 2) are both stochastic events driven by different individual parameters. ABMs are well known to allow for the possibility to exploit the interactions of many boundedly rational agents who behave in a realistic way and are not required to be perfectly informed and able to analyze complex economics problems. As Conlisk [7] pointed out, rational choices can only be justified in some circumstances, say in the presence of a simple context, when good feedback can be obtained and there are plenty of opportunity to validate alternative decisions/estimates. The task of forecasting future unemployment rate is indeed demanding, nuanced and plagued with structural uncertainty and noise and, we believe, an ABM is a promising tool to reconcile the data with some plausible description of agents' behavior.

The model shows that scattered update of information together with the frequent recourse to over-pessimistic forecasts and oblivion are sufficient to accurately replicate the empirical data.

The results tells a story in which only a minority of agents acquire (i.e., is infected with) the public professional forecast and answer the survey accordingly. A much larger proportion of households, for reasons that are not investigated in this work but may be related, say, to extreme risk aversion or lack of trust in the official statistics, embrace the most pessimistic view. Most of the agents, yet, give an intermediate and lazy answer based on outdated information, whose effect can last with dubious effectiveness for several years. Observe that such agents may emotionally feel in a comfortable herd, as they belong to the majority of the population, see Baddeley [2].

One of the main novelty of our model is the introduction of education as a powerful driver of heterogeneity in behavior and the intuition that the schooling level is extremely relevant is a valuable insight of the model: more educated agents more often obtain and use the professional forecasts and less often resort to the cheap answer that unemployment will increase with no regard to the current economic situation. The relationship between education and "more rational" behavior turns out to be monotone and extremely significant in our agent-based setup.

It would be interesting to explore in future research the alternative assumption that there is peer-to-peer (local) information exchange (by contrast, in the present framework there is a centralized and unique source for the rational forecast). The NetLogo code would allow for similar computational generalizations that implement other strains of agents' bounded rationality.

References

[1] R. L. Axtell and J. M. Epstein. *Coordination in Transient Social Networks: An Agent-Based Computational Model of the Timing of Retirement*, pages 161–183. Brookings Institution Press, Washington DC, 1999.

[2] M. Baddeley. Herding, social influence and economic decision-making: socio-psychological and neuroscientific analyses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1538):281–290, 2010.

[3] W. A. Branch. The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *The Economic Journal*, 114(497):592–621, 2004.

[4] C. F. Camerer. Progress in behavioral game theory. *The Journal of Economic Perspectives*, 11(4):167–188, 1997.

[5] C. D. Carroll. The Epidemiology of Macroeconomic Expectations. In L. E. Blume and S. N. Durlauf, editors, *The Economy as an Evolving Complex System, III: Current Perspectives and Future Directions*, pages 5–29. Oxford University Press, Oxford, New York, 2006.

[6] C. D. Carroll, R. E. Hall, and S. P. Zeldes. The buffer-stock theory of saving: Some macroeconomic evidence. *Brookings papers on economic activity*, 1992(2):61–156, 1992.

[7] J. Conlisk. Why Bounded Rationality. *Journal of Economic Literature*, 34:669–700, 1996.

[8] J. M. Epstein. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, 2006.

[9] J. M. Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12, 2008.

[10] L. Gerotto and P. Pellizzari. Unemployment expectations in an agent-based model with education. In Y. Demazeau, B. An, J. Bajo, and A. Fernández-Caballero, editors, *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection*, pages 175–186, Cham, 2018. Springer International Publishing.

[11] A. Lusardi and O. S. Mitchell. Financial literacy around the world: an overview. *Journal of pension economics & finance*, 10(4):497–508, 2011.

[12] M. Rabin. Psychology and economics. *Journal of economic literature*, 36(1):11–46, 1998.

[13] N. S. Souleles. Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. *Journal of Money, Credit, and Banking*, 36(1):39–72, 2004.

[14] U. Wilensky. *NetLogo*. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999.