Università
Ca'Foscari
Venezia

Corso di Laurea magistrale (*ordinamento ex D.M. 270/2004*)
in Economia - Economics

Tesi di Laurea

Ca' Foscari
Dorsoduro 3246
30123 Venezia

Precautionary savings and expectations of Italian households during the crisis: an agent-based model

Relatore
Ch. Prof. Paolo Pellizzari
Ch. Prof.ssa Cinzia Di Novi

Correlatore
Ch. Prof.ssa Noemi Pace

Laureando
Luca Gerotto
Matricola 827794

Anno Accademico
2013 / 2014

INDEX

INTRODUCTION

In the light of the current crisis and of the relationship between expectations, rational or irrational as they might be, and individuals behavior, I believe that a very interesting topic is to understand the way in which individuals develop their own expectations and take the subsequent decisions. The reason is straightforward: the more we are able to explain individual behavior, the more we are necessarily able to capture the aggregate effect of policy interventions, that is one the main goals in economics.

This is the case because expectations towards the future of a given individual have a widely documented influence, both at theoretical and empirical level, on decisions (e.g. consumption, saving, investment). Moreover, sometimes initially unjustified expectations turn out to be correct because they turn out to be "self-fulfilling prophecies": a relatively recent stream of research has started focusing on the role of psychological factors and "less-than-fully-rational" shifts in expectations on the business cycles (Milani 2011).

The Rational Expectations Model is an useful starting point, but at the empirical level it has been proved being far from precise (Roberts 1998). One of the main problems is the assumption that all individuals have access to the same, complete set of information, exploited to develop the rational expectation. On the other hand, both information collection (Branch 2004) and information processing are costly. As a further consideration, the stream of research on financial literacy (see for example Lusardi (2013)) has shown that the ability[i] of the median individual to understand economic and financial phenomena is not enough to develop expectations correctly, even if fed with complete statistics and information.

Therefore, I decided to study the behavior, and in particular the saving behavior, of Italian households at the very beginning of the current crisis, when "unemployment", "spread" and "recession" were not yet so debated topics but, at the same time, the seeds of the current problems were sprouting. Do their pessimism anticipate the actual crisis, turning out to be a self-fulfilling prophecy? Or, at the other extreme of the possible scenarios range, have they been surprised by the impending crisis they did not expect at all?

With this aim, I first present the theory of precautionary savings, which describes what are the drivers of the saving behavior, like individual preferences characteristics (e.g. patience, risk aversion) and expectations towards the future (e.g. zero-income probability, income variance). In light of this theory, I describe what kind of behavior I expect, then I look at microdata from the Bank of Italy survey *Indagine sui bilanci delle famiglie Italiane*[1] in order to analyze the actual households saving behavior, using a difference–in–differences model. I will find results that are different from what a fully rational expectation would imply. Then, I will look at an index for the households unemployment expectations developed by ISTAT

[1] Data available at the Bank of Italy website http://www.bancaditalia.it/statistiche/indcamp/bilfait/dismicro.

(*Indagine sulla fiducia dei consumatori*)[2] and, under the assumption that individuals make a particular kind of choice due to incomplete or imperfect information (so, not irrational decisions, but rational decisions based on the wrong ground), I will develop an agent-based model which tries to describe the evolution of expectations, without renouncing to the rationality assumption. I will find that educational level is an important driver, with the more educated agents having the expectations closer to what the Rational Expectations Model would predict.

This thesis is organized as follows: in section I, I present a review of the literature on precautionary savings; in section II, I describe what kind of behavior I expect and I run a diff-in-diff model on Italian microdata; in section III, I develop the agent-based model; in section IV I conclude.[ii]

[ii] I am grateful to my thesis advisors Prof. Paolo Pellizzari and Prof. Cinzia Di Novi. I also thank Prof. Enrica Croda, Prof. Antonio Paradiso and Prof. Dino Rizzi for helpful discussion.

[2] Data available at ISTAT online datawarehouse (I.stat) http://dati.istat.it.

I. PRECAUTIONARY SAVINGS: REVIEW OF PREVIOUS LITERATURE

The begin of the stream of research on precautionary savings is dated in the Sixties; one of the first works is the pioneering paper by Hayne L. Leland (1968), who analyzed from a purely mathematical and theoretical point of view the problem, defining precautionary savings as follows:

"The "precautionary" demand for saving usually is described as the extra saving caused by future income being random rather than determinate"

In this paper, Leland analyzes a classical two period model where the consumer has to choose the saving level in period 1 in order to maximize the expected utility obtained by consumption in periods 1 and 2:

$$\max_k E(U(C_1, C_2))$$

$$\text{s.t.} \begin{cases} C_1 = (1-k)I_1 \\ C_2 = I_2 + (1+r)kI_1 \\ E(I_2) = I_2^* \\ VAR(I_2) = \sigma_2 \end{cases}$$

Where U is the utility function, characterized by decreasing marginal utility, C_i the consumption in period *i*, I_i income in period *i*, *r* the real riskless interest rate at which the agent can lend (or borrow) and *k* the saving rate in the first period, which is the control variable. Denoting by k_0 the optimal saving rate in case of certain income in period ($I_2 = I_2^*$), a risk-averse individual would, under reasonable conditions[3], choose an higher optimal *k* in case of uncertain income with same expected value, and this value of *k* would be increasing in the degree of uncertainty. That is, part of the savings, $k_0 I_1$, would be due to life-cycle consumption smoothing reasons, and part, $(k - k_0)I_1$, to precautionary saving motives, proving the existence of positive precautionary demand for saving.

A more recent paper by Stephen P. Zeldes (1989) couples a theoretical part with the calibration of the inherent model. He compares the behavior of an economic agent under certainty and uncertainty, trying to solve some empirical puzzles like the excess sensitivity of consumption to transitory income, the high growth of consumption in periods of low real interest rates, and the under spending of the elderly. The framework of the problem is similar to that of Leland, but in a multi-period instead of a two-period framework. The only source of uncertainty is assumed to be exogenous labor income, which is composed by a "permanent" component and a "transitory" component.

[3] These "reasonable assumptions" are those of additive utility from consumption in different periods and of a decreasing absolute risk aversion coefficient for the utility function, an assumption corroborated both by common sense and empirical data: ceteris paribus, and individual is more willing to face a lottery (Leland uses as an example a lottery L=(10000, 0.55; -10000, 0.45)) the higher his level of wealth.
If additive utility is not accepted as an assumption, the result holds if the utility function shows decreasing risk aversion to concentration, that is if the agent "become[s] less averse to risk in a variable as that variable becomes increasingly predominant in a constant utility function".

Under certainty, that is under the Modigliani life-cycle hypothesis, the growth rate of consumption is equal to the difference between the real interest rate (*r*) and the rate of time preference (δ); in the particular case in which the two are equal, one of the elements which could influence saving decisions[4] is eliminated and there is complete consumption smoothing: expected consumption is flat over the agent lifetime and, consequently, there is zero consumption growth. Consumption is proportional to the value of human (the present value of expected future earnings) plus non-human wealth (wealth actually accumulated through saving), saving is positive when income is higher than the average and negative the other way round.

An important point found in this article is that, when labor income is a random variable, the propensity to consume out of transitory income[5] is higher when current assets are low with respect to (expected) future income: the intuition is that, if future is expected to be rosy, there is little reason to save now as future income is expected to be even higher. Another policy consideration is that if present higher income is perceived to be a signal of expected even higher future income, the most of it would be spent; if it is perceived to have happened just by chance[6], most of it would be saved.

Moreover, a risk-averse individual tends to respond differently to changes in human and non-human wealth with respect to the life-cycle hypothesis: it will "over-respond" to changes in non-human wealth or current income, which are certain, and it will "under-respond" to changes in human wealth, which is uncertain. This leads to a different optimal pattern of consumption with respect to the certainty case, with a positive growth rate even if the equality condition between the real interest rate and the rate of time preference is respected, and the slope of the consumption path rising as a function of the level of income uncertainty (Caballero 1991). Another interesting finding is that, even if risk averse individuals have a higher amount of precautionary savings ceteris paribus, this does not necessarily mean that a more risk-averse individual would have a higher propensity to consume out of transitory income: this would depend also on the initial wealth[7].

Another difference with the life-cycle hypothesis concerns borrowing constraints, which are not considered in the model by Modigliani[8] but, if added, strongly influence behavior, in particular at relatively low wealth levels: even if constraints do not bind in the present, they could bind in the future, creating a further reason for precautionary savings and for consumption to track income more closely. Of course, the stronger the binding constraints, the stronger such a motive.

[4] That is preference of present over future consumption.
[5] That is, out of the random component of labour income.
[6] For example, if the individual assumes that yearly labour income is i.i.d. an income higher than the average in a given year is not assumed to have any relationship with future incomes.
[7] Since, ceteris paribus, an higher wealth decreases the need for additional precautionary savings.
[8] Which assumes that it is possible to borrow and lend at the same riskless interest rate.

This branch of research was studied in depth by Angus Deaton (1991), who studied the relationship between savings and liquidity constraints. The framework of his study is similar to the models analyzed before: the peculiarity is that he considers the behavior of a risk-averse but relatively impatient consumer ($\delta > r$), who in case of income certainty would borrow against future income to consume more in the present. Moreover, in order to make it consistent with empirical findings, the model is then calibrated in order to be characterized by negatively autocorrelated individual income growth, positively autocorrelated aggregate income growth, and procyclical aggregate saving.

The other core assumption concerns a strict borrowing constraint ($W_t \geq 0$), which requires a nonnegative non-human wealth level in any period. This condition implies that, in any given period, the agent can consume at most his assets plus current income (which Deaton calls "cash at hand", x_t). The inability to borrow reinforces the precautionary motive for saving.

The assumptions of the model lead to the result that, given nonnegativity constraints on the level of assets, the level of consumption must be lower or equal to the level of cash in hand; consequently, defining as c_t^* the optimal level of consumption in period t in the unconstrained problem, the agent consumes the optimal level c_t^* if the constraint is not binding ($c_t^* \leq x_t$), or consumes all the available resources x_t if the constraint is binding ($c^* > x_t$). This means that, if the level of assets plus current income is sufficiently low, it would be optimal to consume all, since, despite prudency considerations, the marginal utility of money in the current period is higher than what it is expected to be in the future.

One of the most basic simulations of the paper, with i.i.d. income N(100,10) and 200 periods, whose graph is reported in Figure 1, shows that it is possible to smooth consumption even holding a relatively low level of assets (never higher than 20% of expected income). Consumption is smoother than income and savings are pro-cyclical.

On the other hand, if income is stationary positively serially correlated, that is, if good draws of income are expected to be followed by other good draws of income and bad draws by other bad draws, more can be spent out of given cash in hand, given that the precautionary motive has gone down[9] (even if it is still present). Anyway, consumption is again smoother than income, savings are procyclical and, in particular occasions, relatively large asset stocks are accumulated: this happens when there are good draws in many subsequent periods (Deaton defines such cluster of periods as "booms" or, when there are instead bad draws, "slumps").

Simulations with different levels for the autocorrelation parameter, ranging from -0,4 to 0.9, show that the higher the degree of autocorrelation, the more consumption is as noisy as income. This because, given that

[9] Remember that the agent is relatively impatient, so it is not worthwhile to save in order to consume many periods from the current one in the future: "money is worth more now than it is expected to be in the future" (Deaton 1991).

consumers are impatient, saving a lot in order to postpone consumption when bad draws will eventually come is not worthwhile; so the agent would consume most of the income in case of good draws and, necessarily, consumption would track income also when it is low because there are too few assets to adequately smooth consumption.

Finally, probably the most interesting simulation concerns an economy characterized by an income process described as a random walk with drift and by periods of booms and slumps. Formally, there are two possible states (s=1,2), with state 1 being the "boom" and state 2 being the "slump". When s = 1, $\Delta lny_t = g_1 + \varepsilon_t$, while when s = 2 $\Delta lny_t = g_2 + \varepsilon_t$, where $g_2<0<g_1$, and ε_t is white noise drawn from the same distribution $N(0,\sigma^2)$ whatever the state. This means that income is expected to increase during "booms" and decrease during "slumps".

States show persistence, meaning that the conditional probability of having the same state the following period is higher than 0.5. What are the findings of the simulation and their interpretation? First of all, when there is a state change, results are maybe counterintuitive: on the one hand, when a slump begins, the agent starts saving because he expects income to fall in the following periods; then, in the subsequent periods the saving rate approaches zero and, if the slump lasts enough, he starts dissaving, drawing resources from the accumulated wealth in order to sustain consumption. On the other hand, if a boom begins, the agent uses all its assets accumulated during bad times in order to finance the "spending boom", remaining with zero assets. During booms, then, given that income is expected to rise, consumers have no motive to save: they would even like to dissave, but they are prevented from doing it by the zero-level assets and the borrowing constraints. Results are reported in Figure 3, even if it must be highlighted that Deaton has deliberately "exaggerated" the parameters with respect to the original model to which he has referred (Hamilton 1989) in order to amplify the effect.

Finally, calibrating the model with estimates for the US market taken from the literature, he confirms that, with the level of earnings uncertainty recorded in the US, "precautionary motives would tend to increase the desired growth rate of consumption and generate a great deal of saving early in the life cycle".

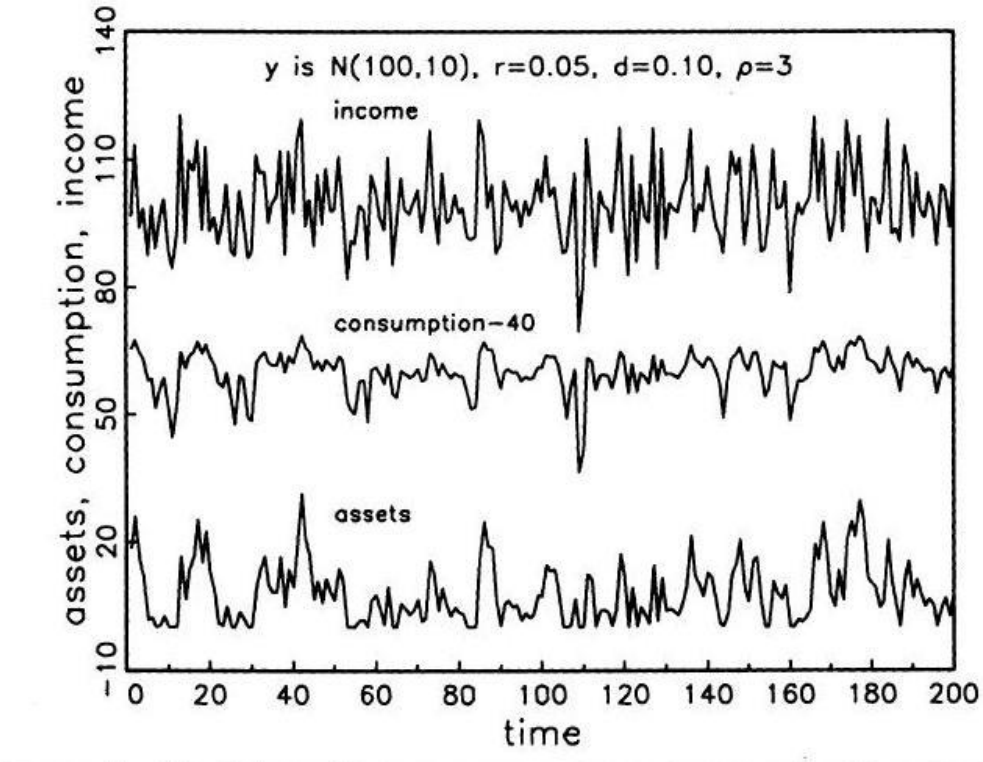


Figure 1: Simulations of income, consumption, and assets, with white noise income. Taken from Deaton (1991) pag. 1230.

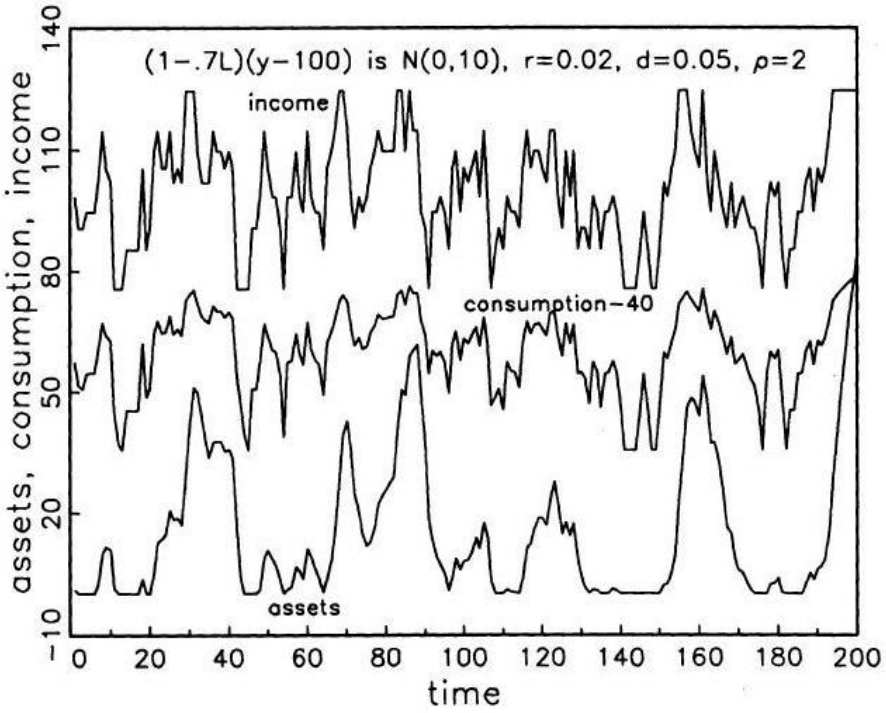


Figure 2: Simulations of income, consumption, and assets, with positively autocorrelated income. Taken from Deaton (1991) pag. 1234.

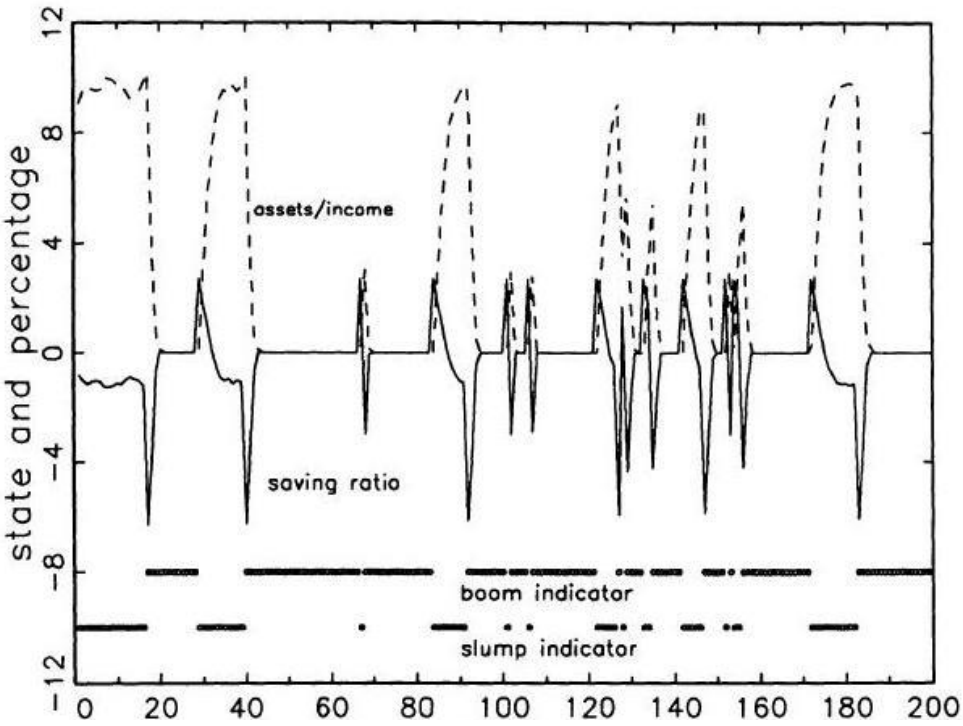


Figure 3 Simulations of saving and assets with autocorrelated growth and two possible states, "boom" and "slump". Taken from Deaton (1991) pag. 1241.

The work of Ricardo Caballero (1991) starts from the results of the research agenda which goes from Leland (1968) to Zeldes (1989) in order to provide a measure of the relative importance of precautionary savings in the United States; his analysis reinforces the previous results and gives some interesting insights, like the pattern of wealth accumulation (see Figure 4), with wealth accumulated for precautionary reasons increasing until middle-age (positive savings) and decreasing afterwards (negative savings), and that total wealth level in any time period increases as a function of risk aversion; moreover, he finds that precautionary saving motives can explain about 60% of observed US net wealth, a result consistent with a different study by Skinner (1988) who found that almost 56% of the accumulated wealth could be explained by prudent behavior of young consumers, who save for precautionary reason. These numbers provide an idea of the fundamental role of precautionary savings, and consequently of their main driver (uncertainty), on the economy and of the importance of studying such a topic.

Carroll (1992), analyzing macroeconomic evidence, tries to link the literature on life-cycle savings with the one on precautionary savings. The assumption is that, like in Deaton (Deaton 1991), the consumer is impatient, in the sense that in case of certain income would borrow against future income to consume more in the current period, and prudent, in the sense that in case of uncertain income would have precautionary saving motives. As a consequence, the model is constructed such as to be similar to the one of Zeldes (1989) with additional impatience ($\delta>r$, while in Zeldes $\delta=r$). Income is noisy, with expected growth rate equal to g. The utility function considered is an isoelastic one.

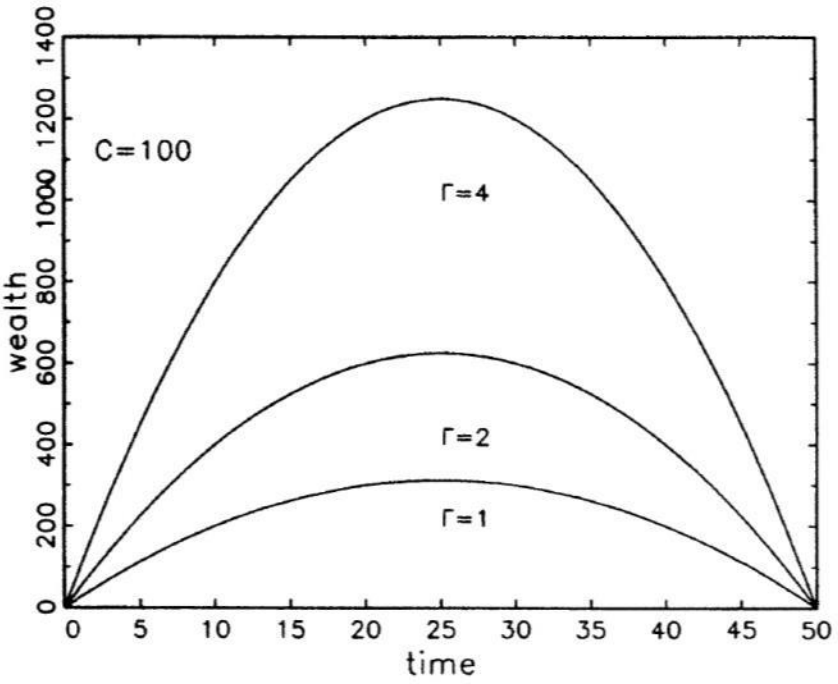


Figure 4: the path of wealth for different values of Γ (where Γ is measure of risk aversion, linearly increasing in the risk aversion coefficient). Taken from Caballero (1991), pag. 861.

These two forces, impatience and prudence, move the saving rate in any given period in the opposite directions[10], and they are in equilibrium if the wealth is equal to what Carroll defines as "target wealth": if wealth is above the target, impatience would dominate prudence leading to dissaving, while if wealth is below the target, prudence would dominate impatience leading to saving[11].

As a consequence, when the wealth level is low[12] prudence dominates, the level of consumption is depressed by precautionary savings and the growth rate of consumption is high; as the wealth level increases, the precautionary saving motive is reduced, freeing resources for consumption. The opposite works if the wealth level is high: see Figure 5. There is an equilibrium when these two forces neutralize each other, and this happens when the growth rate of consumption is equal to the growth rate of income, *g*.

In particular, if shocks to consumption are lognormally distributed, it can be shown that consumption will grow according to the following:

[10] Impatience tending to decrease it, prudence to increase it.
[11] In Carroll's words: "If wealth is below the target, fear (prudence) will dominate impatience and the consumer will try to save, while if wealth is above the target, impatience will be stronger than fear and consumers will plan to dissave. Unemployment expectations are important in this model because when consumers become more pessimistic about unemployment, their uncertainty about future income increases, so their target buffer-stock increases, and they increase their saving to build up wealth toward the new target".
[12] Relative to the current income of expected future income level.

(1) $$\Delta \ln C_{t+1} \approx \rho^{-1}(r-\delta) + \frac{1}{2}\rho E_t var(\Delta \ln C_{t+1}) + e_{t+1}$$

Where $E_t var(\Delta \ln C_{t+1})$ is positively correlated with income variance and negatively correlated with wealth: the lower the wealth, the lower the ability of consumers to draw from a buffer and, consequently, the higher the variance in the growth rate of consumption (and in the consumption level itself).

From the analysis of the formula and the related graph, some insights can be obtained. On the one hand, if the growth rate of income *g* increases, the target wealth/income ratio moves to the left: the higher the expected future income, the higher consumption in the current period and the lower the savings, ceteris paribus. On the other hand, an increase in the degree of uncertainty in income will increase the variance of consumption growth, ceteris paribus, shifting the $E_t \Delta \ln C_{t+1}$ curve upwards and increasing the target level.

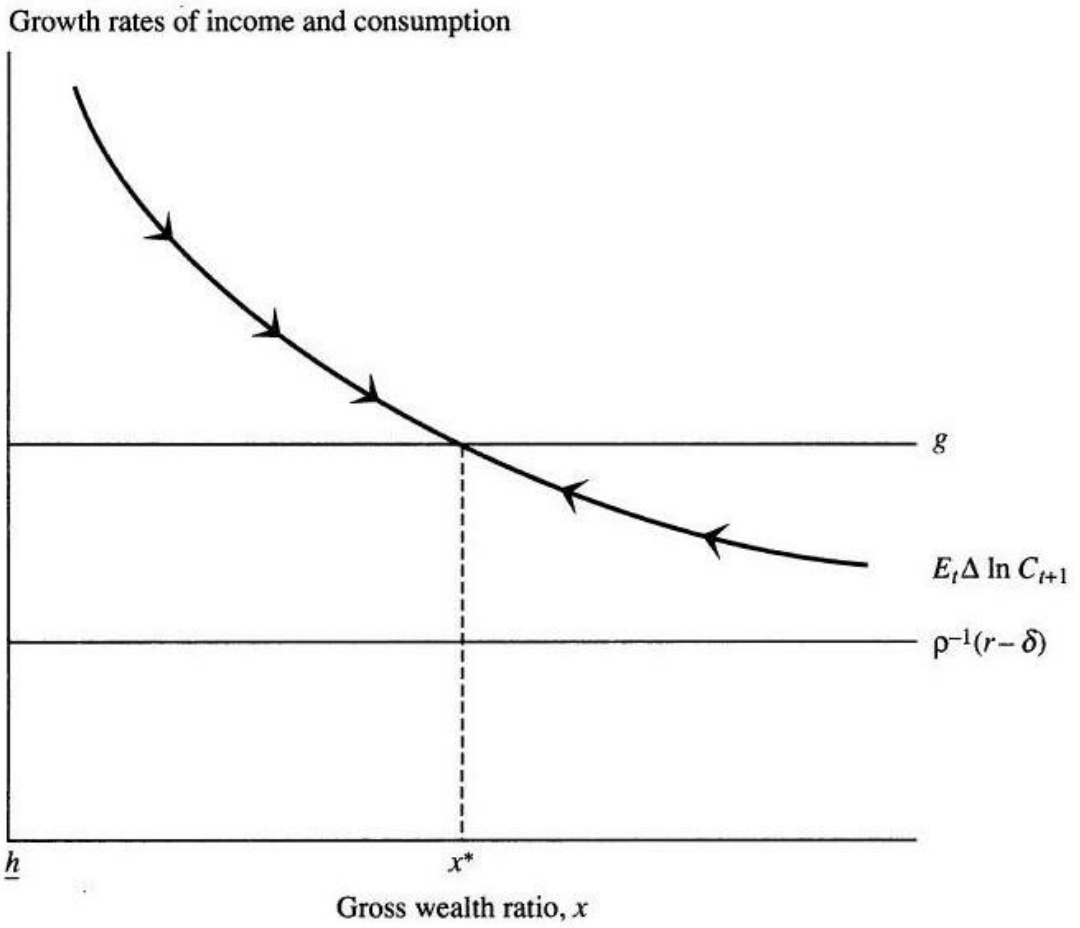
The effect of an increase in risk aversion, represented by the risk aversion coefficient ρ, is ambiguous: as far as the uncertainty part $E_t var(\Delta \ln C_{t+1})$ is considered, ceteris paribus it would increase the expected growth rate of consumption and the target wealth/income ratio, but it would also decrease the perfect certainty growth rate $\rho^{-1}(r-\delta)$. The first is the direct effect of increased risk aversion, the second is the effect of a lower intertemporal elasticity of substitution.

The result of the simulation performed by Carroll according to that formula is that the target wealth-to-income ratio is very sensitive to the degree of risk aversion, implying that the first effect outweighs the second, and that also the amount of precautionary savings is very sensitive to the degree of consumers' risk aversion. Other interesting results concern the importance of the probability of zero income[13] *p*, the standard deviation of permanent shocks, $\sigma_{\ln N}$, and the growth rate of income, *g,* that is the role of expectation towards the future values of those variables. As intuitive, the higher the level of *p* and $\sigma_{\ln N}$, the higher the target wealth, while the higher *g* the lower the target wealth.

Considering the option to put explicit liquidity constraints, like in Deaton (1991), Carroll concludes that "in a buffer-stock model with relevant borrowing constraints, loosening the constraints should result in behavior qualitatively identical to the effect of increasing the minimum guaranteed future income in a model without liquidity constraints[14]". Reasoning the other way round, given that an increase in the minimum guaranteed future income implies a decrease in the precautionary saving, putting into the model explicit liquidity constraints would increase precautionary savings, ceteris paribus.

[13] We can surely consider the unemployment rate as a proxy for that probability.
[14] Intuition: both loosening of borrowing constraints and increase in minimum guaranteed future income lead to less potential problems in the future.

Figure 5: The Buffer-Stock Relationship between Growth Rates and the Gross Wealth Ratio. Taken from Carroll (1992), pag. 78.



Source: Author's calculations as described in the text. Arrows represent convergence toward the "target" wealth ratio. Gross wealth ratio is the ratio of gross wealth to permanent income. The minimum possible amount of gross wealth is h. The expected consumption growth next period is $E_t \Delta \ln C_{t+1}$. The target wealth on hand is x^*. The perfect certainty growth rate of consumption is $\rho^{-1}(r - \delta)$. The growth rate of income is g.

Linking the theoretical framework of the model exposed and empirical findings, according to the buffer-stock model consumption and income growth should converge in the long-run (otherwise the wealth-to-income ratio would explode up or down); another finding is that, in case of changed expectations that influence the target wealth-to-income ratio, the adjustment is gradual and not instantaneous[15]. Moreover, the variance of consumption growth decreases as the wealth level increases, since consumers have more resources from which it is possible to draw in order to smooth consumption. Households typically reduce debt burden during recessions, and similarly express greater reluctance to dip into savings in periods when they are pessimistic about income growth.

[15] Consumers do not reach the new desider target in just one period.

Also, concerning unemployment, the author conducted tests comparing the buffer-stock model with the classical CEQ-PIH[16] model, using data taken from the University of Michigan *Surveys of consumers*, coupled with aggregate data on personal saving and unemployment. Unemployment data are shown to have an important influence on saving, ceteris paribus[17]: "Consumers everywhere were inclined towards precautionary savings to provide a cushion against the threat of unemployment"[18]. Consequently, given a higher propensity to save, the consumption is depressed in periods of heightened unemployment fears.

Concerning fear of unemployment, the opinion of Carroll is that there is more fear when, for example, the unemployment rate is 10% than when it is equal to 5% but rising: levels matter more than changes.

Another finding of Carroll is that "A negative consumption shock means high saving now, but the drop in C reduces aggregate demand. As demand falls, firms fire workers and pay the remaining workers less. So an increase in saving is followed by drops in employment and income" and that an extreme, but possible, interpretation is that "consumers expectations are largely determined by random shocks, but that expectations turn out to be correct because they constitute self-fulfilling prophecies".

Moreover, there is an hypothesis, that Carroll will study in depth in his following works, that it is possible to link buffer-stock and life-cycle models: "if expected income growth is high early in life but lower (or negative) as retirement approaches, it is entirely possible that consumers will engage in buffer-stock saving when young but, after a certain age, will switch to more traditional life-cycle saving behavior as their expected income growth falls".

Five years later (1997), Carroll issued another paper on this stream of research; the first theoretical findings are related to the work of 1992 (Carroll 1992): the assumptions on prudence and impatience[19] are exactly the same, leading to the finding that consumption does not respond one-for-one for transitory shocks in income, since the buffer is used or created, and there is an additional consumption growth induced by permanent income shocks[20].

He then moves a critique to the previous empirical work on consumption, since it has ignored the expected variance of consumption term in the Euler equation, assuming it to be equal to zero or anyway constant, whereas we have seen in equation 1 that this term, interacting with the risk-aversion coefficient, makes the

[16] Certainty equivalence-Permanent Income Hypothesis.
[17] That is, even controlling for income expectations.
[18] Quotation, made by Carroll, of a Gallup Poll director after a public opinion poll in Britain in August 1991 (taken from a newspaper article: "Consumer Caution Constrains Recovery," The Times of London, August 7, 1991, business section, p. 20). A "Gallup Poll" is a public opinion poll run by the performance-management consulting company Gallup, Inc.
[19] Concerning impatience, he moves a critique towards Deaton's insights (Deaton 1991) arguing that his results are due mainly to the impatience assumption rather than to the liquidity constraints.
[20] The agent saves more, since future income has higher uncertainty, and consume more in the future (using the buffer).

curve shift, having consequently an impact on the wealth-to-ratio target level and the corresponding precautionary savings. Consequently, he considers the most important implication of this misspecification that typical methods of Euler equation estimation are meaningless if the consumers involved are buffer-stock savers, given that these methods do not consider the expected consumption variance term[21].

One important conclusion that can be inferred from the buffer-stock model is that there is a one-to-one relationship between an increase in the growth rate in permanent income and the growth rate of consumption[22]. Moreover, the Life Cycle/Permanent Income Hypothesis (LC/PIH) and buffer-stock models provide very different results about marginal propensity to consume (MPC) out of transitory income: the life-cycle model implies an estimated MPC of only 2%, while for buffer-stock consumers ranges from 15 to 50%[23], with those who face a greater variance in transitory shocks that, because of a lower MPC, should, on average, show a greater divergence between consumption and income. Given this small MPC, the LC/PIH is unable to explain the consumption income parallel, while the buffer-stock is.

Concerning the relationship between the growth rate of income and the saving rate in steady-state, given that to maintain the same wealth-to-income ratio the saving rate must be proportional to both the growth rate of income and the target ratio ($s = gw^*$), one could be tempted to conclude that there is a one-to-one relationship between *s* and *g*. Anyway, w^* is inversely proportional to *g*[24], so this effect is mitigated and the relationship is not one-for-one: so, the human wealth effect is smaller than in the standard model, in the sense that expectation towards the future, which is uncertain, are weighted less in the current consumption-saving decision.

Carroll highlights also what are some important drawbacks of the buffer-stock model: first of all, it is unable to accurately explain the accumulation of housing wealth, which is an illiquid asset (and not a liquid assets to be used if necessary, like assumed by the model); the second drawback is that not all consumers, and in particular not the very rich, act like buffer-stock consumer (nor according to the life-cycle model) and this precludes the possibility of capturing the whole aggregate effect, given that the very rich account for a substantial component of the aggregate[25]. Also, it is not proper for the elderly, because it does not account for sources of uncertainty other than labor income, like emergencies due to health problems or the alike.

[21] Actually, the only consumers for which those Euler equation estimation methods give reliable results are consumers with infinite wealth, since they can use it as a buffer and the growth rate of consumption, being independent from the (stochastic) income, has zero variance.
[22] Although the adjustment is, as highlighted talking about (Carroll 1992), not instantaneous.
[23] According to different parameters for taste preferences, wealth level, and so on.
[24] For a refresh, review discussion about Figure 5.
[25] Carroll reports that the top 1% wealthier US households possess the 64% of the financial wealth.

Finally, making reference also to other works published between his 1992 work and this 1997 one, he reinforces his intuition that consumers behavior can be described as buffer-stock when young, switching to life-cycle saving around 45.

Moving to less theoretical and more empirical papers, Engen and Gruber (2001), starting from the theoretical finding that one of the major potential determinants for precautionary saving is the lost wage due to unemployment, and so that the effect should be greater for those facing higher unemployment risk or for the younger[26], use the different Unemployment Insurance (UI) systems in the different US states as a natural experiment to test the theory. They found that higher social security generosity reduces saving by decreasing the precautionary motive and taxing away individual savings, since consumers (who are also taxpayers) have to pay the "premium" through taxation.

The simulation results show that the "crowding out" effect of even low unemployment insurance is substantial and, coherently with the starting theoretical findings, this "crowding out" effect decreases with age and increases with unemployment risk. These results are tested and confirmed by empirical data, with estimates which are also of an higher magnitude than the ones implied by the simulation (possibly due to a much smaller starting level of savings in the data than in the model). Moreover, Engen and Gruber (2001) too found drawbacks similar to the ones of Carroll, in the sense that they consider only liquid assets (and not other kinds of wealth accumulation) and that wealthier households behavior is unlikely to be influenced by the generosity of the UI system, so the model is not able to capture the behavior of this kind of consumers.

Gourinchas and Parker (2002) estimate a model of life-time consumption in the presence of labor income uncertainty, with uncertainty estimated using data from the PSID[27] and CEX (Consumer Expenditure Survey) datasets and simulation techniques. Differently from previous work, they also include a bequest function, in order to incorporate bequest motives for saving.

They found that the expenditure pattern depends crucially on the expected growth rate of income: if it and the discount rate are low relative to real interest rate, the behavior is better approximated by the CEQ-LCH (Certainty-equivalent life-cycle hypothesis model), otherwise consumers behave according to the buffer-stock model. Gourinchas and Parker (2002) confirms the intuition of Carroll (1997) that consumers behave as buffer-stock consumers when young, since they face a greater labor income uncertainty in the long time span before retirement, while retirement and bequests motive are more important after age 40, when

[26] Who have a longer uncertain time span before retirement.
[27] Panel Study of Income Dynamics.

retirement age approaches. According to their estimates, household discount the future at low rates[28] and are not particularly risk averse.

The precise moment of the "shift" depends on the time preferences of the individual: if the discount factor is high, the future is important and the individual starts saving for retirement (or bequest) purposes earlier in life; on the other hand, if the consumer is impatient consumption will parallel income for a longer time span, to decrease sharply approaching retirement in order to build up the necessary level of assets.

The paper defines as life-cycle savings the difference between income and life-cycle consumption and precautionary savings as the complement of life-cycle savings to obtain total savings. Dissecting the two effects, according to the calibration of Gourinchas and Parker (2002) model, early in life impatient young consumers would like to borrow against future income if it were certain (negative life-cycle savings) while they have a positive demand for precautionary savings, with the second effect outweighing the first and leading to overall positive savings and positive wealth level; around age 40, life-cycle savings becomes larger than precautionary savings and the consumer begins building wealth for retirement. As wealth increases, the variance of consumption growth decreases, leading to a lowering of precautionary savings both in absolute value and in relative importance with respect to life-cycle savings. Eventually, at the very end of the working life precautionary savings component is negative, since the agent consumes the buffer he has accumulated given he will face no income uncertainty from that point onwards. Moreover, as it is possible to see also from Figure 6, given (uninsurable) income risk, the wealth level is higher at any age level than the one implied by the life-cycle model only.

The work of Marco Cagetti (2003) further extends the model, by considering not only retirement, bequests and precautionary motives, but also the possibility that wealth is accumulated in order to finance significant expenditures in given moments of life (e.g. buying an house, financing children education). Also an uncertain lifetime (between 65 and 91 years old) is introduced.

He found differences in time preference and risk aversion among educational groups: considering "no high school", "high school" and "college", he highlights that more educated people (college graduates) are less impatient and more risk-averse. Then the rate of time preference is mainly the same for the lowest two groups, while risk aversion is increasing in the educational level. The educational group at the opposite side of the educational range (no high school) maybe is not so risk-averse and patient because of the impact of welfare programs, that reduces the necessity of taking care of their future for these individuals.

[28] Consequently, the discount factor is high.

Precautionary savings is still found to be an important component of total savings, making up the almost totality of wealth for young consumers, while for older individuals (those near to retirement) the wealth level with the uncertainty framework is twice the one that the life-cycle model would imply. So, even with this extension, precautionary savings still explains a large fraction of the wealth of the median household, and the estimates confirm that the assumption of an impatient and prudent consumer is realistic. Also, it

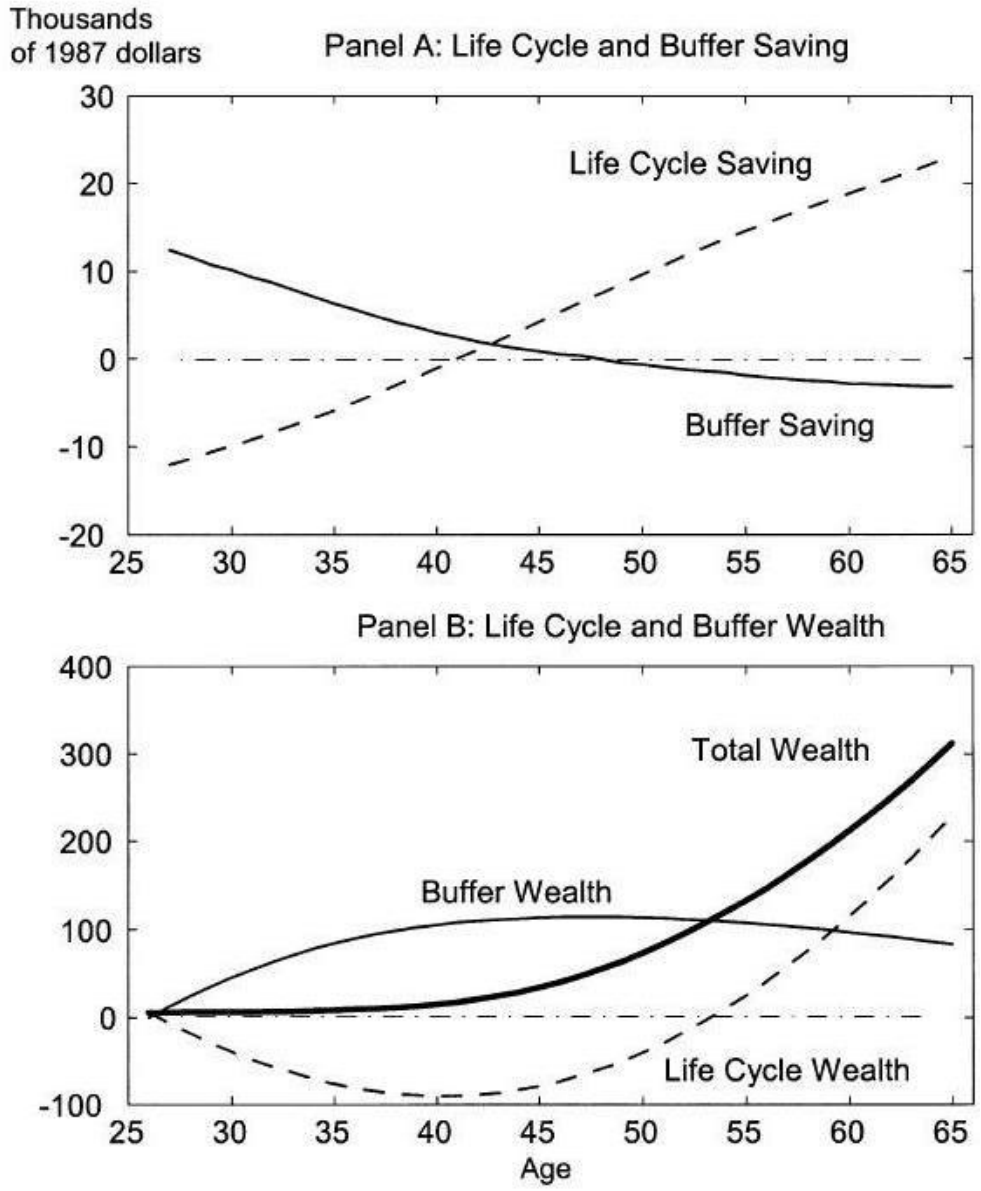


Figure 6: the role of risk in saving and wealth accumulation. Taken from Gourinchas and Parker (2002), page 75 Fig. 7.

has been shown that there are some differences in the degree of patience among different demographic groups.

Finally, in a very recent paper of May 2012, Giavazzi and MacMahon (2012) conducted a quasi-natural experiment in order to understand the impact of a change in uncertainty on the behaviour of consumers. They used German microdata, considering the period around 1998 election (1995 to 2000), one of the

closest elections of post-war Germany. The interest towards this election comes from the fact that the outgoing government (Kohl) had issued a pension reform to move the social security system from a path defined as "unsustainable". This reform would have had an influence on workers of the private sector, leading, among the other changes, to a decrease of the replacement rate (from 70% to 67%) and of payroll contribution rates (that were projected to reach 25% in 2050, from the current 18%). Schröder, the "challenger", promised that he would withdraw the reform in case of success at the elections, that is what he actually did when he became the Chancellor. Anyway, Germans were conscious that the previous path was unsustainable, and that in case of withdraw another reform necessarily would come in a few years: actually, not to intervene when a reform is necessary does not perpetuate status quo, it raises uncertainty.

In order to study the change in behavior, Giavazzi and MacMahon run a diff-in-diff regression on the data of German Socioeconomic Panel (GSOEP), using private sector workers, affected by the reform, as treatment group characterized by a theoretical increase in uncertainty, and civil servants as control group. The results are striking: the treatment group saved markedly more than the control group in the period of heightened uncertainty, with the coefficient on the interaction term[29] group being both statistically and economically significant. An example reported in the paper shows that an household that previously was saving 8.9%[30] of disposable income in 1998, ceteris paribus, saved 15.9% in year 2000: a 7 percentage points increase, mostly twice as much! The increase in uncertainty induced households to save more than they would otherwise, lowering consumption and contributing to the slowdown of the German economy at the beginning of the millennium.

The interpretation of the results is that individuals affected by the reform have to increase their wealth in order to offset the loss in pension rights wealth. For this reason, despite the result found in the literature that precautionary savings is more important the younger the worker, in this situation the ones with the sharper change in saving pattern were the ones closest to retirement, since they had a shorter time span to recover the pension rights wealth loss. Moreover, since individuals had to reach an higher target-to-income ratio, to use Carroll's original language, the increase in saving level is not supposed to last forever: they are supposed to save markedly more only for the period necessary to reach the new steady-state.

The conclusion of this study is really important: in the authors' words: "consumption may fall and the economy might slow down for no other reason than political uncertainty".

[29] Between the two dummies: one for the period (heightened uncertainty=1, otherwise 0) and treatment (treated=1, control=0).
[30] The average saving rate for the balanced panel in 1998.

WHY IS IT IMPORTANT TO STUDY EXPECTATIONS AND PRECAUTIONARY SAVINGS?

The role of expectations in the economy is crucial, as we have repeatedly seen in the literature review; in particular, to recall Carroll's words (Carroll 1992), one extreme interpretation is that "consumers expectations are largely determined by random shocks, but that expectations turn out to be correct because they constitute self-fulfilling prophecies". And we have seen that there is a strong relationship between future income expectations, of which one of the main indicators is constituted by unemployment, and saving behavior; savings themselves have a strong influence on the real economy, since, always according to Carroll (1992) "A negative consumption shock means high saving now, but the drop in C reduces aggregate demand. As demand falls, firms fire workers and pay the remaining workers less. So an increase in saving is followed by drops in employment and income".

The literature on precautionary savings suggests that a significant amount of wealth accumulation is imputable to precautionary motives, that the main driver of precautionary savings is income uncertainty, and one of the main drivers of income uncertainty is unemployment. Moreover, many more people cite having a buffer from which to draw in case of emergencies rather than saving for retirement as the most important reason for saving (Carroll 1997). Also in Cagetti (2003) it is reported that, according to the Summary of Consumer Finance (SCF) 1992 data, in which consumers were asked what was their main reason for saving, precautionary saving appears to be the most important reason for saving for all age groups lower than 55 years old. Also for the age group that is closer to retirement (56-65), precautionary saving is reported as the main reason for saving by almost one out of three (32.72%) of the interviewees, not far from retirement purposes (38.65%). These findings suggest that, in people's mind, precautionary savings could be even more important than the substantial amount estimated with theoretical models and related simulations. As a further clue, in Carroll (1997) it is reported also that financial planning guides widely used in the US suggest that the consumer should keep as liquid assets between four and six times his average monthly wage, and if the employment situation is unstable this quantity should be increased; a point sustained by the finding that, using PSID data, saving rates and wealth are substantially higher for people with more variable income.

Anyway, the literature is mainly based on US macro and microdata; it is fundamental to try to understand if it can be applied also to countries like Italy, where the drivers of saving do not need necessarily to be the same. A very interesting point concerning the difference between US and European countries like Italy can be found in Deaton (Deaton 1991), where it is reported that the median household wealth, excluding

housing and pension rights, is equal to 1000$[31], an amount much lower than Italy, which has one of the highest private wealth in the world[32] (Bartiloro and Rampazzi 2013). Moreover, the median propensity to save of Italian households in the last 20 years has been 3-4 times the one of US households (Bartiloro and Rampazzi 2013).

So, unfortunately, it appears not to be possible to make a direct inference on the role of precautionary savings and the other main drivers of saving for other countries like Italy. On the other hand, from a theoretical point of view, the main drivers of saving are:

1) Saving for retirement (life-cycle)
2) Saving for precautionary motives

in countries like Italy we have a relatively generous (at least with respect to the US) Social Security system, which provides a lower motive for saving for retirement, considering an argument similar to that provided by Engen and Gruber (2001) for precautionary savings and unemployment insurance.
On the other hand, this provides a further motive to study precautionary savings in Italy: if saving for retirement is less important, this should increase the relative role of the other kind of savings, like the precautionary one: for this reason, it is possible to consider 60%, the estimate of Caballero (1991) for the US, as a lower bound for the relative importance of precautionary savings in Italy and to guess that even more than 60% of wealth accumulation could be due to precautionary motives.

Assuming that the theoretical framework of precautionary savings literature can be applied also to Italy, it is interesting to test the findings and to interpret them. I decided to use a difference-in-difference estimation procedure, similar to that adopted by Giavazzi and MacMahon (2012), comparing the consumption-saving decisions of groups which should react differently, according the theoretical literature, to a given shock to uncertainty (or to the perceived uncertainty). Giavazzi and MacMahon (2012) use private sector workers as treatment group and civil servants as controls; another option is to compare the self-employed against employees, whose income is much less volatile.
The beginning of the crisis, despite being sometimes conventionally identified with the Lehman Brothers crash, is not so clear-cut; moreover, the advent of the crisis (and, more important, the perception of the advent of the crisis) is different for different countries. On the one hand, between the very end of 2006 and the beginning of 2007 the subprime crisis arises; on the other hand, the first year of negative real GDP growth in Italy has been 2008, but has this been a cause or a consequence of consumers behavior? Remember Carroll: if as a consequence of negative expectations consumption is depressed, this reduces

[31] Of course, we are considering year 1991 USD, equivalent to 1543.27 year 2013 USD (as calculated using the Price Indexes provided by Eurostat database http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do;jsessionid=9ea7d07d30d622b0b55523604d3487ebfa9725c6ef44.e34OaN8PchaTby0Lc3aNchuMchiPe0.
[32] Even if, unfortunately, coupled with one of the biggest public debt in the world.

aggregate demand, leading to a decrease of GDP and maybe an increase of unemployment, turning out to be a self-fulfilling prophecy.

Assuming that people are rational, we expect that, when the perception of future uncertainty changes, the consumer defines a new level of target wealth, higher in case of increased uncertainty or worse future expectations and lower in case of decreased uncertainty or better future expectations, adapting consequently the consumption-saving decisions in order to reach the new target wealth level. For example, if an higher wealth-to-income ratio is desired, we expect a sharp increase in the saving rate followed by a return to a rate similar to the initial one[33] in the following periods. This behavior should be more pronounced for the groups which face an higher relative uncertainty: for example, for a given shock a civil servant should face a lower (if any) increase in uncertainty than a worker of the private sector, and similarly a self-employed higher than an employee. Assuming, again, that people are rational, when the expected reaction is observed, it could be considered as a clue of the fact that a change in the economic environment has been perceived. This could help studying the development of expectations, representing a starting point to build the agent based-model.

For this reason, I used Italian households microdata taken from the *Indagine sui bilanci delle famiglie*[34], run by the Bank of Italy (*Banca d'Italia*) every two years. The survey is conducted in every odd year, collecting information on each household income, saving and wealth levels, consumption patterns and habits of the previous year: for example, the 2013 survey is named *Indagine sui bilanci delle famiglie 2012*, since data collected concern year 2012. Every additional detail about data elaboration is presented in the appendix.

[33] It depends on the new steady state saving rate.

[34] Data available at the Bank of Italy website http://www.bancaditalia.it/statistiche/indcamp/bilfait/dismicro.

II. DIFFERENCE-IN-DIFFERENCES MODEL

I have chosen to run a difference-in-differences (diff-in-diff) comparing years 2006 versus 2008 and 2008 versus 2010[35], to study saving patterns.

A difference-in-differences estimation is aimed to highlight the effect of a given exogenous shock (in this case, the advent of the crisis) between a group that should be hit by the shock (the "treated") and another group that should not be hit, or hit to a lower extent, (the "controls"). This effect is estimated calculating the difference between the two groups in the differences in the value assumed by the dependent variable before and after the shock, ceteris paribus.

More formally, I estimate the following model:

$$(2) \qquad Y_{it} = \alpha + \beta_1 Treat + \beta_2 Year_{2008} + \beta_3 Treat * Year_{2008} + \gamma X_{it} + \varepsilon_{it} \; t = 2006{,}2008$$

Where Y_{it} is the saving rate of household *i* in period *t*, *Treat* is a dummy equal to one for treated households, $Year_{2008}$ is a dummy equal to one for the second year of the comparison (in this case, 2008)[36], X_{it} is a vector of demographic and economic controls and ε_{it} the usual error term. α, β_1, β_2, β_3 and γ are the parameters to be estimated.

Here I present the results of the regression having as treated group the self-employed and as control the employees (with the exclusion of temporary workers, or "temp", which are present in the regression as an alternative treatment group: this because their employment situation is at least as unstable as those of self-employed); in the Appendix II I present the results for the analysis with civil servants as treated and other workers as controls, which provides a similar intuition but the division of treatment and control groups is less precise, and this could have influenced results.

The results for the regression for years 2006-2008 is reported in Table 1. The dummy Self-employed assumes value 1 for the treated and 0 for controls, while the dummy YEAR assumes value 1 in year 2008 and 0 in 2006.

Among the control variables are included the main economic drivers of saving: income (Y), wealth (W), the number of components of the household (NCOMP) and that of income perceptors (NPERC), a dummy for civil servants (PUBLIC) and some demographic controls: age group (where lower than 35 is the reference group), location (where North-West is the reference group), gender (where male is the reference group) and educational level (where less than high school is the reference group)

[35] Data for year 2012 are also available, but I prefer not to use them because of the many events ("Fornero reform, increased taxation and so on) that could have influenced the change in behavior: interpretation would be very difficult.

[36] For the 2008-2010 comparison, t=2008, 2010 and $Year_{2010}$ (no more $Year_{2008}$) is a dummy equal to one for the second year of the comparison: that is 2010.

The results are mainly consistent with the theoretical background: the propensity to save, ceteris paribus, increases with the level of income and decreases with the level of wealth, decreases with the number of components of the household (intuition: it is necessary to spend more, ceteris paribus). The coefficient on PUBLIC is negative, but not statistically significant. One point is interesting: the coefficient on NPERC is positive and statistically significant, while applying the concept of precautionary saving motives I expected the coefficient to be negative: if the same household income derives from more sources, this income is less variable and uncertain, the precautionary motive is lower and this should lead to lower, not higher, savings. It would be interesting to study in depth the problem (maybe there are nontrivial interactions with other variables), but this topic is outside the scope of this thesis and is left for future research.

The age controls are not statistically significant, probably due also to a problem of numerosity, while the controls for location are. For example, the coefficients for the dummies that identify workers leaving in Southern Italy or in the two big islands (Sicily and Sardinia) are positive. In these regions fears of unemployment are higher than in the rest of Italy and the saving rate, consistently with the theory, is higher. For the dummy North-East, this effect is probably due to a much higher median income for workers living in this area, and to cultural reasons. There is no economic nor statistically significant difference among genders, but controls for educational level are both economically and statistically significant at 1% level.

Of the three regressors that characterize a diff-in-diff model, the most interesting is the interaction between Self-employed and YEAR, that represents the difference in the change of saving rate from the first (2006) to the second period (2008) between the treated (self-employed) and the untreated (employees). This coefficient (3.11%) is positive but not statistically significant.

The coefficient on YEAR is not statistically significant, but the coefficient on Self-employed is (at the 5% level). Anyway, the sign of the coefficient on Self-employed is the opposite of what one would eventually expect: empirical studies found that the saving rate is higher for those facing a more uncertain income, while this coefficient would suggest that the saving rate of the self-employed, ceteris paribus, is lower than the one of the employees. One possible explanation could be a self-selection of more risk-averse workers into less uncertain job position, meaning that self-employed are less risk-averse, but this effect would bias the estimate towards zero: it is improbable that it is strong enough to invert the sign. This result is anyway interesting, from a theoretical point of view.

Table 1: Saving regression, baseline results. Years 2006-2008. Dependent variable: reported saving rate

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.59E+00	1.10E-01	-23.522	2.00E-16	***
D(temporary worker)	-1.75E-02	3.44E-02	-0.509	0.61079	
D(Self-employed)	-3.25E-02	1.48E-02	-2.201	0.02779	*
D(PUBLIC)	-8.01E-03	1.07E-02	-0.751	0.45247	
YEAR	-9.93E-03	9.43E-03	-1.053	0.2923	
log(Y)	2.73E-01	1.13E-02	24.059	2.00E-16	***
NCOMP	-3.86E-02	4.08E-03	-9.441	2.00E-16	***
W	-1.16E-08	5.71E-09	-2.034	0.04208	*
NPERC	4.11E-02	7.27E-03	5.649	1.77E-08	***
D(AGE 35-44)	2.21E-02	2.57E-02	0.861	0.38932	
D(AGE 45-54)	1.79E-02	2.53E-02	0.707	0.47951	
D(AGE 55-64)	-5.92E-04	2.55E-02	-0.023	0.9815	
D(AGE >65)	5.77E-02	5.82E-02	0.991	0.32159	
D(North-East)	3.83E-02	1.12E-02	3.405	0.00067	***
D(Center)	1.74E-02	1.25E-02	1.389	0.16493	
D(South)	7.92E-02	1.33E-02	5.965	2.74E-09	***
D(Islands)	6.06E-02	1.55E-02	3.915	9.24E-05	***
D(Female)	5.13E-03	1.00E-02	0.511	0.60941	
D(High School)	-5.23E-02	9.45E-03	-5.534	3.41E-08	***
D(Bachelor, Master or PhD)	-7.06E-02	1.49E-02	-4.753	2.10E-06	***
YEAR:D(temporary worker)	-3.28E-02	4.78E-02	-0.686	0.49274	
YEAR:D(Self-employed)	3.11E-02	2.01E-02	1.552	0.1208	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2197 on 2882 degrees of freedom
Multiple R-squared: 0.2911, Adjusted R-squared: 0.286
F-statistic: 56.36 on 21 and 2882 DF, p-value: < 2.2e-16

A clue for studying the problem comes from the work of Bartiloro and Rampazzi (2013), that is based on the same datasets (*Indagine sui bilanci della famiglie italiane*). In Figure 7 is reported Fig. 2 taken from that work; graph (a) represents the median of the saving rate by quartile of equivalent income[37] for different surveys ranging from 1991 and 2010. Graph (b) represents the same by quartile of net wealth.

[37] The equivalent household income is the income level any individual should have, if he was to leave alone, in order to maintain the same standard of living. It is calculated through the modified OECD equivalence scale: the householder has a coefficient equal to 1, any other individual older than 14 0.5 and younger than 14 0.3. For each household the number of "equivalent adults" is calculated, summing up these coefficients, and household income is finally divided by this sum. The rationale is that two persons living together incur less expenses than two persons living separately, ceteris paribus. For further details, see Bartiloro and Rampazzi (2013).

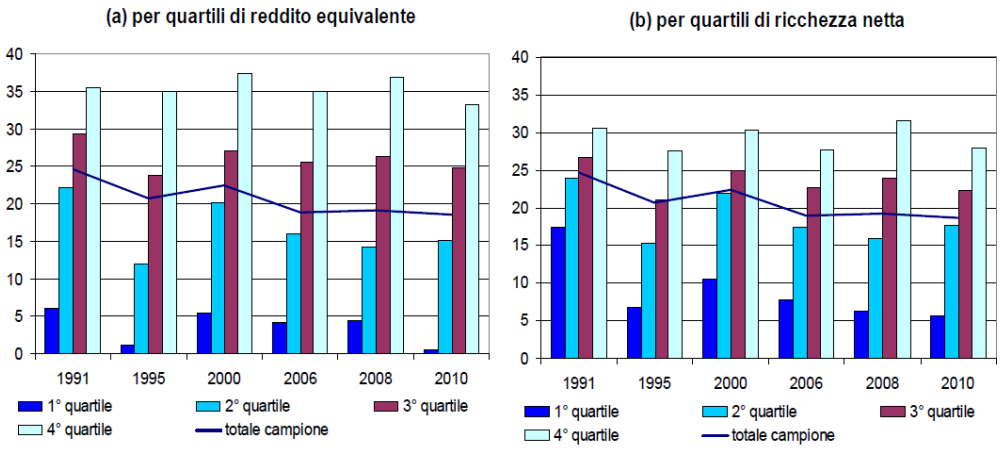


Figure 7: median of propensity to save, by quartiles of equivalent household income and net wealth. Taken from Bartiloro and Rampazzi (2013), Fig. 2.

Looking at graph(a), in particular the medians of years 2006, 2008 and 2010, it is possible to note that the behavior of the 3rd and 4th quartiles is different from the one of the 1st and the 2nd ones. On the one hand, for the highest quartiles there is an increase in the median of the saving rate from 2006 to 2008, followed by a decrease in 2010 that brings the saving rate under the starting one of 2006. For the 2nd quartile, on the other hand, there is a decrease in 2008 followed by an increase in 2010. For the 1st quartile, there is an initial increase followed by a sharp decrease from 2008 to 2010, probably due to the saving behavior of the growing number of unemployed, which are anyway excluded from the analysis (see appendix for details). These very different behaviors induced me to try to run the regression on two different subsets, those having equivalent household income higher or lower than the median[38], in order to compare the results of the different diff-in-diff coefficients.

In Table 2 we report the results of the regression for the subsample having household equivalent income higher than the median (18440€). The coefficient on the interaction term is 1.85%, but it is not statistically significant (probably also because the sample size has been halved). The coefficients on YEAR and Self-employed are not statistically significant: the expected positive and significant coefficient on Self-employed (2.42%) has not been found yet, but this finding, due to the possible bias toward zero, is a less important puzzle than the negative coefficient of Table 1. The economic and control variables (Y, W, NCOMP, NPERC) maintain their sign, significance and interpretation, with the exception of W which is not statistically significant in this subsample, while the demographic ones maintain their sign and interpretation but are now less statistically significant, with the exception of the educational ones that are still significant at the 0.1% level.

[38] The median was not exactly the same across the different datasets used. For the 06-08 comparison for self-employed and employees is was equal to 18440€, for the 08-10 comparison for self-employed and employees to 19740€. Intuitively, these value can be interpreted as an individual earning about 1500€ per month.

The interpretation of the results of Table 2 is that, ceteris paribus, self-employed and employees had the same saving rate in 2006 and maintained it in 2008.

Table 2: Saving regression, only households reporting equivalent household income higher than the median. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.92E+00	1.82E-01	-10.537	2.00E-16	***
D(temporary worker)	-5.02E-03	7.66E-02	-0.066	0.94772	
D(Self-employed)	2.42E-02	1.89E-02	1.278	0.20132	
D(PUBLIC)	1.43E-03	1.33E-02	0.108	0.91431	
YEAR	5.10E-03	1.22E-02	0.417	0.67683	
log(Y)	2.10E-01	1.82E-02	11.545	2.00E-16	***
NCOMP	-4.46E-02	6.49E-03	-6.873	9.35E-12	***
W	-3.83E-09	5.41E-09	-0.708	0.47891	
NPERC	5.99E-02	9.83E-03	6.092	1.43E-09	***
D(AGE 35-44)	1.32E-02	3.47E-02	0.381	0.70339	
D(AGE 45-54)	-1.25E-02	3.40E-02	-0.367	0.71395	
D(AGE 55-64)	-5.09E-02	3.43E-02	-1.484	0.13804	
D(AGE >65)	1.02E-02	6.84E-02	0.15	0.88107	
D(North-East)	3.69E-02	1.34E-02	2.764	0.00579	**
D(Center)	1.56E-02	1.50E-02	1.041	0.29815	
D(South)	3.45E-02	1.98E-02	1.74	0.08202	.
D(Islands)	5.42E-02	2.38E-02	2.283	0.02259	*
D(Female)	5.64E-04	1.24E-02	0.046	0.96359	
D(High School)	-5.20E-02	1.37E-02	-3.801	0.00015	***
D(Bachelor, Master or PhD)	-5.89E-02	1.74E-02	-3.379	0.00075	***
YEAR:D(temporary worker)	-4.10E-02	1.01E-01	-0.404	0.68601	
YEAR:D(Self-employed)	1.85E-02	2.45E-02	0.755	0.45014	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1991 on 1430 degrees of freedom
Multiple R-squared: 0.1868, Adjusted R-squared: 0.1749
F-statistic: 15.65 on 21 and 1430 DF, p-value: < 2.2e-16

In Table 3 we report the results of the regression for the subsample composed by households with equivalent income lower than the median. In this case the estimates of the diff-in-diff regression are much different: the only statistically significant regressor (at the 1% level) is Self-employed, with a strongly negative coefficient: -6.72%. This is a puzzle much more problematic than the one present in the regression with the whole sample, because the risk-aversion argument can justify the bias toward zero, but cannot justify such strongly negative coefficient. The other two coefficient, on YEAR and on the interaction term, are not statistically significant.

Table 3: Saving regression, only households reporting equivalent household income lower than the median. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.24E+00	2.28E-01	-14.229	2.00E-16	***
D(temporary worker)	2.94E-03	4.13E-02	0.071	0.94333	
D(Self-employed)	-6.72E-02	2.33E-02	-2.877	0.00408	**
D(PUBLIC)	-1.92E-02	1.69E-02	-1.138	0.25533	
YEAR	-2.11E-02	1.42E-02	-1.49	0.13652	
log(Y)	3.38E-01	2.43E-02	13.896	2.00E-16	***
NCOMP	-3.48E-02	6.83E-03	-5.093	4.00E-07	***
W	-5.92E-08	2.64E-08	-2.243	0.02503	*
NPERC	2.75E-02	1.14E-02	2.414	0.01591	*
D(AGE 35-44)	1.66E-02	3.75E-02	0.444	0.65715	
D(AGE 45-54)	2.83E-02	3.72E-02	0.759	0.44778	
D(AGE 55-64)	3.58E-02	3.76E-02	0.951	0.34187	
D(AGE >65)	1.21E-01	1.03E-01	1.171	0.24163	
D(North-East)	4.28E-02	1.88E-02	2.273	0.02316	*
D(Center)	2.02E-02	2.09E-02	0.968	0.33337	
D(South)	1.03E-01	1.90E-02	5.402	7.73E-08	***
D(Islands)	6.99E-02	2.15E-02	3.247	0.00119	**
D(Female)	9.54E-03	1.62E-02	0.59	0.55538	
D(High School)	-5.17E-02	1.34E-02	-3.855	0.00012	***
D(Bachelor, Master or PhD)	-6.61E-02	3.01E-02	-2.197	0.02817	*
YEAR:D(temporary worker)	-3.34E-02	5.74E-02	-0.582	0.56064	
YEAR:D(Self-employed)	2.82E-02	3.25E-02	0.867	0.38613	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2352 on 1430 degrees of freedom
Multiple R-squared: 0.2204, Adjusted R-squared: 0.209
F-statistic: 19.26 on 21 and 1430 DF, p-value: < 2.2e-16

So, the interpretation of the results of Table 3 is that in 2006, ceteris paribus, self-employed workers had a saving rate lower by more than 6 percentage points with respect to employees; in 2008, both categories (employees and self-employed) have kept constant their saving rate.

A possible critique is that the estimates on the interaction of Self-employed times YEAR reported in Table 2 (1.85%) and Table 3 (2.82%) are similar to that of Table 1 (3.11%), they are not statistically significant only because of a problem of "micronumerosity" (sample size halved), but they are economically significant and their economic meaning is that the reaction of the "poorer" self-employed was higher in magnitude than the "richer" self-employed. Even if no statistician would ever accept such an argument, even interpreting the coefficients like if they were statistically significant would lead to a different conclusion: "poorer" employees saved less (-2.11 percentage points) while self-employed saved more (2.82-2.11=0.71 percentage points) but their increase is lower than self-employed workers with equivalent income higher

than the median, which saving rate increases by about 2 percentage points (1.85+0.51=2.36 perentage points). Assuming that they have perceived the same shock, the behavior of the highest two quartiles seems "more rational", at least as far as the theory of precautionary savings is concerned.

In Table 4, Table 5 and Table 6 are reported the results of the regressions for the years 2008-2010. The regression with the whole sample, looking at the insignificant coefficient on the interaction term, seems to suggest that the saving rate of treated and untreated has moved in a parallel way: given the negative and significant coefficient on YEAR, both self-employed and employees have, ceteris paribus, decreased their saving rate by a bit less than 2 percentage points (-1.77%). Looking at the controls, it is interesting to note that now gender is a regressor statistically significant at the 5% level. Anyway, splitting the sample in two as a function of equivalent household income[39], the results are really different.

The "richer" subsample has a strongly positive and significant coefficient on Self-employed (5.34%), a negative and significant coefficient on YEAR (-3.11%) and an insignificant coefficient on the interaction term. This findings are consistent with the graph (a) of Figure 7, with a decrease of the saving rate from the first to the second period for the top two quartiles. Treated and controls saving rate moved in a parallel way, decreasing.

The result for the "poorer" subsample is totally different: the coefficient on Self-employed is negative (-1.60%) but not significant, posing a less problematic puzzle than the strongly negative coefficient found in the 2006-2008 dataset. The ones on YEAR and the interaction terms are not significant.

[39] Where the threshold is the median (19740€).

Table 4 Saving regression, baseline results. Years 2008-2010. Dependent variable: reported saving rate

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.60E+00	1.08E-01	-24.066	2.00E-16	***
D(temporary worker)	-4.55E-02	3.20E-02	-1.42	0.15569	
D(Self-employed)	1.30E-02	1.47E-02	0.885	0.3761	
D(PUBLIC)	1.10E-03	1.05E-02	0.106	0.91587	
YEAR	-1.77E-02	9.54E-03	-1.856	0.0635	.
log(Y)	2.73E-01	1.11E-02	24.588	2.00E-16	***
NCOMP	-3.62E-02	4.13E-03	-8.769	2.00E-16	***
W	-2.22E-08	6.25E-09	-3.552	0.00039	***
NPERC	3.20E-02	7.50E-03	4.268	2.04E-05	***
D(AGE 35-44)	-8.18E-03	2.50E-02	-0.327	7.44E-01	
D(AGE 45-54)	1.74E-03	2.46E-02	0.071	0.94372	
D(AGE 55-64)	-3.78E-03	2.47E-02	-0.153	0.87869	
D(AGE >65)	4.74E-02	4.15E-02	1.142	0.25365	
D(North-East)	4.93E-02	1.13E-02	4.377	1.25E-05	***
D(Center)	2.43E-03	1.29E-02	0.189	8.50E-01	
D(South)	8.09E-02	1.36E-02	5.955	2.94E-09	***
D(Islands)	4.51E-02	1.55E-02	2.903	3.73E-03	**
D(Female)	2.15E-02	9.73E-03	2.214	0.02691	*
D(High School)	-3.77E-02	9.63E-03	-3.914	9.32E-05	***
D(Bachelor, Master or PhD)	-6.34E-02	1.41E-02	-4.503	6.99E-06	***
YEAR:D(temporary worker)	4.85E-03	4.45E-02	0.109	9.13E-01	
YEAR:D(Self-employed)	-4.70E-03	1.97E-02	-0.239	0.81124	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2129 on 2704 degrees of freedom

Multiple R-squared: 0.3102, Adjusted R-squared: 0.3048

F-statistic: 57.9 on 21 and 2704 DF, p-value: < 2.2e-16

Table 5 Saving regression, only households reporting equivalent household income higher than the median. Years 2008-2010. Dependent variable: reported saving rate

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.92E+00	1.84E-01	-10.466	2.00E-16	***
D(temporary worker)	5.42E-02	6.68E-02	0.812	0.4172	
D(Self-employed)	5.34E-02	1.86E-02	2.879	0.00405	**
D(PUBLIC)	2.10E-02	1.36E-02	1.544	0.1229	
YEAR	-3.11E-02	1.26E-02	-2.465	0.01382	*
log(Y)	2.12E-01	1.83E-02	11.602	2.00E-16	***
NCOMP	-4.24E-02	6.92E-03	-6.12	1.23E-09	***
W	-1.38E-08	5.95E-09	-2.317	0.02063	*
NPERC	4.59E-02	1.06E-02	4.341	1.52E-05	***
D(AGE 35-44)	-1.69E-02	3.64E-02	-0.464	0.64301	
D(AGE 45-54)	-2.58E-02	3.55E-02	-0.728	0.46685	
D(AGE 55-64)	-3.24E-02	3.56E-02	-0.91	0.36273	
D(AGE >65)	1.46E-02	5.16E-02	0.283	0.77702	
D(North-East)	4.81E-02	1.35E-02	3.572	0.00037	***
D(Center)	-7.30E-03	1.55E-02	-0.47	0.63823	
D(South)	3.82E-02	2.00E-02	1.911	0.05615	.
D(Islands)	3.18E-03	2.44E-02	0.13	0.89651	
D(Female)	1.82E-02	1.22E-02	1.484	0.13793	
D(High School)	-3.98E-02	1.42E-02	-2.805	0.0051	**
D(Bachelor, Master or PhD)	-5.89E-02	1.80E-02	-3.277	0.00108	**
YEAR:D(temporary worker)	-3.09E-02	8.96E-02	-0.345	0.73005	
YEAR:D(Self-employed)	-1.03E-02	2.45E-02	-0.42	0.67445	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1968 on 1341 degrees of freedom
Multiple R-squared: 0.1879, Adjusted R-squared: 0.1752
F-statistic: 14.78 on 21 and 1341 DF, p-value: < 2.2e-16

Table 6 Saving regression, only households reporting equivalent household income lower than the median. Years 2008-2010. Dependent variable: reported saving rate

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.44E+00	2.32E-01	-14.827	2.00E-16	***
D(temporary worker)	-3.77E-02	3.87E-02	-0.975	0.32985	
D(Self-employed)	-1.60E-02	2.39E-02	-0.668	0.50455	
D(PUBLIC)	-3.13E-02	1.62E-02	-1.937	0.05292	.
YEAR	-5.96E-03	1.41E-02	-0.423	0.67268	
log(Y)	3.56E-01	2.46E-02	14.512	2.00E-16	***
NCOMP	-3.50E-02	6.92E-03	-5.054	4.93E-07	***
W	-1.33E-07	4.57E-08	-2.897	0.00382	**
NPERC	2.00E-02	1.15E-02	1.738	0.08242	.
D(AGE 35-44)	-3.36E-03	3.44E-02	-0.098	0.92221	
D(AGE 45-54)	1.91E-02	3.40E-02	0.562	0.57414	
D(AGE 55-64)	2.29E-02	3.43E-02	0.667	0.50519	
D(AGE >65)	1.17E-01	7.36E-02	1.585	0.11324	
D(North-East)	5.42E-02	1.88E-02	2.881	0.00403	**
D(Center)	1.69E-02	2.15E-02	0.787	0.43121	
D(South)	1.14E-01	1.98E-02	5.735	1.21E-08	***
D(Islands)	7.95E-02	2.16E-02	3.678	0.00024	***
D(Female)	2.27E-02	1.53E-02	1.488	0.13699	
D(High School)	-3.89E-02	1.34E-02	-2.911	0.00367	**
D(Bachelor, Master or PhD)	-3.90E-02	2.51E-02	-1.558	0.11951	
YEAR:D(temporary worker)	-1.16E-02	5.34E-02	-0.217	0.82807	
YEAR:D(Self-employed)	4.74E-03	3.13E-02	0.151	0.8797	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2243 on 1341 degrees of freedom
Multiple R-squared: 0.2448, Adjusted R-squared: 0.233
F-statistic: 20.7 on 21 and 1341 DF, p-value: < 2.2e-16

How is it possible to interpret these results? Why do I not find any significant evidence of a change in behavior, like in Giavazzi and MacMahon (2012)? One possible explanation is that Giavazzi and MacMahon take into account more than two years, getting a bigger sample size and having consequently an higher significance level. The (unsignificant) estimates are anyway far from the striking 8 percentage points increase in the saving rate found in Giavazzi and MacMahon (2012).

According to the theoretical findings, one expects first of all to highlight that, before the crisis, self-employed workers had a saving rate higher than employees, ceteris paribus, since a greater income variability should imply an higher propensity to save. Secondly, one expects to register some change in behavior in one of these years (2006 vs 2008, or 2008 vs 2010): for example, a positive and significant coefficient on YEAR, denoting that the households have reacted to the shock increasing their saving rate, and a positive and significant coefficient on the interaction term, because the self-employed workers, due to the higher volatility intrinsic in their job, have increased their saving rate more than employees. These could have been registered between 2006 and 2008, if the households had reacted to the subprime crisis shock, or between 2008 and 2010, if a reaction to the first year of recession in Italy (2008) had occurred.

On the other hand, what we found is that low-income self-employed workers save less than low-income employees, something very different from the theoretical prediction (or from what was found in other empirical studies), that the diff-in-diff coefficient is not statistically significant for any subsample and that between 2008 and 2010 high-income households decreased their saving rate, instead of increasing it. Should we assume that householders are irrational in their decision? Maybe, but economists do not like very much to give up the rationality assumption. Another hypothesis to try to explain the phenomenon is not that householders are irrational, but that they take rational decisions based on the incomplete or imperfect information that they have in that given moment.

Why different reactions for different subsamples? What is the driver? The sample has been split as a function of equivalent household income, but this variable is definitely correlated with some demographic variables. Many works, like Cagetti (2003) which uses different risk-aversion and time preferences parameters for different educational groups, or Souleles (2001), suggest that education could have a strong influence on behavior and attitudes, so I decided to try to understand if education level could be a driver. There are three educational groups: “Less than high school”, “High school” and “Bachelor, Master or PhD”.

First of all, looking at regressions of Table 1 to Table 6, there is a categorical variable concerning educational level inserted among the control variables. The reference group is “Less than High School”, while in the regressions are presented two dummies, “High School” and “Bachelor, Master or PhD”, which indicate the percentage points difference in the saving rate between that group and the reference one, ceteris paribus. It is possible to appreciate that these control variables are statistically significant in most

regressions, but “Bachelor, Master or PhD” for the subsamples composed by households with equivalent income lower than the median[40].

Moreover, the distributions of education level on the top two and the lower two quartiles of household equivalent income is vastly different.

Table 7: percentage of householders in each educational group, for different subsamples

	Less than high school	High School	Bachelor, Master or PhD
2006-08, YEQ<median	0.51	0.44	0.05
2006-08, YEQ>median	0.23	0.55	0.22
2008-10, YEQ<median	0.48	0.45	0.08
2008-10, YEQ>median	0.22	0.53	0.25

Note: percentages may not sum to one due to rounding

The distributions are different, in particular as far as the extreme groups are considered: the proportion of individuals with less than high school among the “richest” is half of the proportion among the “poorest”, while the number of individuals with at least a bachelor degree among the “richest” is three to four times the one among the “poorest”.

The difference in the educational level distribution between the two subsamples is so clear-cut that it appears to be a good candidate to help explain the different behavior.

I created three subsamples according to the educational level; anyway, householders having a degree higher than High School are not enough to have significant results for the diff-in-diff (e.g. just about 50 householders having a Bachelor, Master or PhD and working as self-employed), so I put together “High School” and “Bachelor, Master or PhD” versus “Less than high school”. The results of the four regressions are reported in Table 8 to Table 11. Looking only at the coefficients of interest for a diff-in-diff model, neither the interaction terms nor the dummies that denote Self-employed workers are statistically significant; on the other hand, as far as the educational level “Less than High School” is concerned, both in the regression for 2006-08 and in the one for 2008-10, we find a negative coefficient on the variable YEAR that is statistically significant at the 10% level. This is a further clue that educational level could be an important driver of behavior, and that this hypothesis deserves a deeper study.

In order to try to explain this “irrational” behavior and the link with the educational level, we now move to the third section of this dissertation. In this section I make reference to an interesting work of Christopher Carroll and I develop an agent-based model to track the evolution of unemployment expectations[41].

[40] Which, looking at Table 7, could be due to a problem of “micronumerosity” of that educational group in those subsamples.

[41] One of the main drivers of precautionary savings.

Table 8: Saving regression, only households having "High School" or "Bachelor, Master or PhD" as educational level. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.45E+00	1.33E-01	-18.487	2.00E-16	***
D(temporary worker)	-9.06E-02	5.10E-02	-1.776	0.07595	.
D(Self-employed)	-2.49E-02	1.88E-02	-1.323	0.18616	
D(PUBLIC)	-9.68E-03	1.26E-02	-0.769	0.44176	
YEAR	5.01E-04	1.19E-02	0.042	0.96653	
log(Y)	2.56E-01	1.33E-02	19.252	2.00E-16	***
NCOMP	-3.72E-02	5.18E-03	-7.182	1.00E-12	***
W	-8.20E-09	5.88E-09	-1.395	0.16312	
NPERC	4.34E-02	9.42E-03	4.607	4.38E-06	***
D(AGE 35-44)	3.03E-03	3.19E-02	0.095	0.92423	
D(AGE 45-54)	2.25E-03	3.16E-02	0.071	0.94323	
D(AGE 55-64)	-3.86E-02	3.20E-02	-1.206	0.22803	
D(AGE >65)	1.64E-02	7.20E-02	0.228	0.81976	
D(North-East)	3.93E-02	1.40E-02	2.805	0.00509	**
D(Center)	1.81E-02	1.57E-02	1.15	0.25033	
D(South)	6.64E-02	1.67E-02	3.967	7.56E-05	***
D(Islands)	4.09E-02	2.02E-02	2.024	0.04309	*
D(Female)	4.68E-03	1.25E-02	0.374	0.70856	
YEAR:D(temporary worker)	1.16E-02	7.02E-02	0.166	0.8683	
YEAR:D(Self-employed)	2.72E-02	2.53E-02	1.076	0.28207	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.221 on 1805 degrees of freedom

Multiple R-squared: 0.2675, Adjusted R-squared: 0.2598 F-statistic: 34.7 on 19 and 1805 DF, p-value: < 2.2e-16

Table 9: Saving regression, only households having "Less than High school" as educational level. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.06E+00	2.01E-01	-15.272	2.00E-16	***
D(temporary worker)	5.10E-02	4.66E-02	1.094	0.274	
D(Self-employed)	-4.00E-02	2.44E-02	-1.639	0.10154	
D(PUBLIC)	-3.74E-03	2.05E-02	-0.183	0.85521	
YEAR	-2.78E-02	1.53E-02	-1.815	0.06986	.
log(Y)	3.17E-01	2.06E-02	15.389	2.00E-16	***
NCOMP	-3.81E-02	6.74E-03	-5.648	2.08E-08	***
W	-9.12E-08	3.64E-08	-2.505	0.0124	*
NPERC	3.31E-02	1.16E-02	2.861	0.0043	**
D(AGE 35-44)	5.76E-02	4.34E-02	1.327	0.18467	
D(AGE 45-54)	4.96E-02	4.22E-02	1.174	0.24047	
D(AGE 55-64)	6.74E-02	4.24E-02	1.589	0.11238	
D(AGE >65)	1.39E-01	9.86E-02	1.407	0.15985	
D(North-East)	3.57E-02	1.88E-02	1.899	0.05789	.
D(Center)	1.86E-02	2.07E-02	0.897	0.36988	
D(South)	1.04E-01	2.18E-02	4.772	2.08E-06	***
D(Islands)	9.23E-02	2.43E-02	3.808	0.00015	***
D(Female)	5.57E-03	1.72E-02	0.324	0.74585	
YEAR:D(temporary worker)	-6.52E-02	6.50E-02	-1.003	0.31586	
YEAR:D(Self-employed)	4.00E-02	3.29E-02	1.216	0.22443	

Residual standard error: 0.2163 on 1059 degrees of freedom

Multiple R-squared: 0.3431, Adjusted R-squared: 0.3313 F-statistic: 29.11 on 19 and 1059 DF, p-value: < 2.2e-16

Table 10: Saving regression, only households having "High School" or "Bachelor, Master or PhD" as educational level. Years 2008-2010. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.54E+00	1.29E-01	-19.688	2.00E-16	***
D(temporary worker)	-4.76E-02	4.73E-02	-1.007	0.3139	
D(Self-employed)	1.49E-02	1.83E-02	0.813	0.4166	
D(PUBLIC)	-2.20E-03	1.24E-02	-0.178	0.85851	
YEAR	-1.33E-02	1.19E-02	-1.122	0.2622	
log(Y)	2.68E-01	1.29E-02	20.693	2.00E-16	***
NCOMP	-3.44E-02	5.12E-03	-6.718	2.48E-11	***
W	-1.95E-08	6.52E-09	-2.996	2.77E-03	**
NPERC	3.30E-02	9.61E-03	3.429	0.00062	***
D(AGE 35-44)	-4.51E-02	2.96E-02	-1.522	0.12814	
D(AGE 45-54)	-2.92E-02	2.93E-02	-0.997	0.31867	
D(AGE 55-64)	-5.28E-02	2.97E-02	-1.779	0.07535	.
D(AGE >65)	-1.22E-02	5.58E-02	-0.219	0.8265	
D(North-East)	4.34E-02	1.40E-02	3.105	0.00193	**
D(Center)	4.34E-03	1.60E-02	0.271	0.7862	
D(South)	7.63E-02	1.70E-02	4.489	7.62E-06	***
D(Islands)	8.14E-03	1.98E-02	0.411	6.81E-01	
D(Female)	1.54E-02	1.20E-02	1.283	0.19978	
YEAR:D(temporary worker)	5.99E-02	6.69E-02	0.896	0.37062	
YEAR:D(Self-employed)	-2.64E-02	2.44E-02	-1.085	0.27813	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2154 on 1759 degrees of freedom

Multiple R-squared: 0.2892, Adjusted R-squared: 0.2815 F-statistic: 37.66 on 19 and 1759 DF, p-value: < 2.2e-16

Table 11: Saving regression, only households having "Less than High school" as educational level. Years 2008-2010. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.79E+00	2.02E-01	-13.8	2.00E-16	***
D(temporary worker)	-4.73E-02	4.32E-02	-1.095	0.27382	
D(Self-employed)	5.82E-03	2.50E-02	0.233	0.81591	
D(PUBLIC)	-4.62E-03	1.98E-02	-0.233	0.81577	
YEAR	-2.63E-02	1.58E-02	-1.66	0.09726	.
log(Y)	2.83E-01	2.08E-02	13.609	2.00E-16	***
NCOMP	-3.66E-02	7.09E-03	-5.163	2.98E-07	***
W	-5.49E-08	2.76E-08	-1.991	0.04683	*
NPERC	3.18E-02	1.21E-02	2.621	0.00892	**
D(AGE 35-44)	8.46E-02	4.72E-02	1.792	0.07338	.
D(AGE 45-54)	9.12E-02	4.58E-02	1.992	0.04665	*
D(AGE 55-64)	1.16E-01	4.58E-02	2.529	0.01159	*
D(AGE >65)	1.54E-01	6.59E-02	2.332	0.01992	*
D(North-East)	5.58E-02	1.89E-02	2.944	0.00332	**
D(Center)	-9.76E-03	2.19E-02	-0.446	0.65567	
D(South)	8.65E-02	2.26E-02	3.819	0.00014	***
D(Islands)	9.92E-02	2.51E-02	3.961	8.04E-05	***
D(Female)	2.78E-02	1.71E-02	1.623	0.10484	
YEAR:D(temporary worker)	-3.40E-02	5.91E-02	-0.576	0.56456	
YEAR:D(Self-employed)	3.99E-02	3.32E-02	1.201	0.2299	

Residual standard error: 0.2068 on 927 degrees of freedom

Multiple R-squared: 0.3589, Adjusted R-squared: 0.3458 F-statistic: 27.32 on 19 and 927 DF, p-value: < 2.2e-16

III. EXPECTATIONS FORMATION PROCESS: AN AGENT-BASED MODEL

CARROLL'S MODEL

The model developed by Christopher Carroll is relatively simple and very intuitive, but at the same time really interesting. Inspired by the medical literature on epidemiological models, he tries to export them to social sciences and economics treating expectations diffusion like a virus: each individual can be "infected" by a new, professional forecast on a given macroeconomic indicator (e.g. inflation rate, unemployment rate, GDP growth…) over the near future, and consequently update his own expectation, or maintain his own idea in case he is not "infected".

The idea is that individuals do not have their own macroeconomic model to develop their own expectations, but that they absorb the economic content and related forecasts of news reports diffused by the media (television, newspapers and so on). In any given period[42], each individual has a given probability of reading (or hearing) these news reports, updating his expectation or, if this is not the case, maintaining his previous expectation.

Of course, the "forecast virus" can be passed also between one individual and another; anyway, in the simpler model present in Carroll a "common-source infection" version is presented.

In the "common-source" model, individuals do not exchange information among them but can acquire news reports from an unique source, that is the media. All media are assumed to publish on a regular basis the same "professional forecast" of a given macroeconomic indicator, so that there is no heterogeneity in the news reports at the same time. In any given period, any individual has a probability *p* of paying attention to the news reports and obtaining the latest professional forecasts and *1-p* of not obtaining it.

Other than the model, the work of Carroll (2003) has found a stronger relationship between households forecast and professional forecasts in periods in which there was a greater coverage of economic indicators in newscasts. This provides an intuition for the reason that could have led to a different economic significance for the results of Giavazzi and MacMahon and of the present work: being the pension reform one of the most debated topics in the political campaign, the subject was well-known by the public opinion. Inflation and unemployment rates, in relatively normal times, are a much less debated topic.

[42] For Carroll, a quarter.

DATA SOURCE: PROFESSIONAL FORECASTS

First of all, we need the "professional forecasts" present in the model of Carroll. I decided to use forecasts taken from the OECD[43] Economic Outlook[44] in order to build a time series.

OECD Economic Outlook is released by OECD twice a year, in June and December[45]. In June editions are presented forecasts for the reference year and the next one, while in December there is the forecast also for another year. For example, in 2008, in June 2008 OECD Economic Outlook unemployment rate forecasts for 2008 and 2009 are present, while in December 2008 edition forecasts for 2008[46], 2009 and also 2010 can be found.

OECD Economic Outlook Datasets can be found in the OECD Statistics Database stats.oecd.org, under Economic Projections. The oldest Economic Outlook for which unemployment forecasts at the country level are available is N.38 of December 1985, so I built a time series of professional forecasts by picking from each edition the unemployment rate forecast for the following year. The series is composed by 56 semesters between December 1985 and June 2013[47].

DATA SOURCE: A REAL MEASURE OF HOUSEHOLD UNEMPLOYMENT EXPECTATIONS

Once obtained professional forecast, some measure of households unemployment expectations is needed in order to calibrate the model and verify its ability to track households expectations.

ISTAT (*Italian National Institute of Statistics*) households confidence indexes (*Indagine sulla fiducia dei consumatori*)[48] have been developed monthly since 1982. Until December 2010 the survey has been conducted by ISAE (*Istituto di studi e analisi economica*). From January 2011 it is managed by ISTAT, applying the same methodology adopted previously. Data collection is run in the first 10 working days of the reference month, through telephone interviews carried out with CATI (Computer assisted telephone interviewing) technique. The sample, which size is of about 2000 units, is random, formed in two stages (telephone directory, consumers), proportional to the Italian adult population and stratified for region of address and municipality dimension.

The list used is the telephone directory, the sample unit is the subscriber, randomly chosen in the layer, and the statistical unit is the consumer, where for consumer is meant an adult person who is part of the

[43] Organisation for Economic Co-operation and Development.
[44] Data can be found at OECD online database http://stats.oecd.org.
[45] But in 2013, when the second edition has been released in November instead of December.
[46] It is difficult to define as "forecast" an estimate for 2008 released in December 2008; anyway, it can be considered as an end-year projection based on data collected in the first part of the year.
[47] I excluded the latest edition, November 2013, because I do not have enough data for the corresponding household expectations measure.
[48] Data available at ISTAT online datawarehouse (I.stat) http://dati.istat.it.

household which corresponds to the selected subscriber and contributes, also in non-monetary terms[49], to build the income of his household.

The questionnaire for the survey comprehends, together with some questions on the structure and the income of the household, some qualitative questions, characterized by three to five ordinal modalities[50], about the situation of the household and the economy in the last period (last 12 months) and expectations towards the near future (next 12 months).

One of the qualitative questions concerns expectations towards unemployment in the near future. There are six possible answers: "it will go up a lot"[51], "it will go up a little"[52], "it will remain stable"[53], "it will go down a little"[54], "it will go down a lot"[55], "don't know/no answer"[56]. The index that will be used for the agent-based model, is obtained using the following procedure:

- First, the percentage of sample units that has chosen each of the six answers is calculated
- The percentage who answered "it will go up a lot" is multiplied by 2; "it will go up a little" by 1; "it will remain stable" or "don't know/no answer" by 0; "it will go down a little" by -1; "it will go down a lot" by -2[57].
- The numbers so obtained are summed up, multiplied by 100 and finally rounded to the nearest integer number

The index so obtained has a range of -200 to 200[58], and extrema are reached if the whole sample has answered "it will go down a lot" or "it will go up a lot", respectively. The interpretation of a positive value index is that the general opinion can be summarized as "the unemployment rate is going to rise in the next 12 months", while a negative value would mean that "the unemployment rate is going to decrease in the next 12 months"; then, the higher the absolute value, the stronger such an opinion.

Now, the problem is to reconcile a biannual measure of unemployment rate forecasts with a monthly index of households expectations, expressed on a scale of -200 to 200 points.

[49] For example, an housewife contributes to the formation of the household income.
[50] Plus the possibility not to answer or to answer "I don't know".
[51] *Aumenterà molto*.
[52] *Aumenterà poco*.
[53] *Rimarrà stabile*.
[54] *Diminuirà poco*.
[55] *Diminuirà molto*.
[56] *Non sa/non risponde*.
[57] Until December 1994, those numbers were the opposite: -2 for "will go up a lot", then -1, 0, 1, 2.
[58] Until December 1994, those numbers were the opposite: 200 if the whole sample has answered "it will go down a lot", -200 if the whole sample has answered "it will go up a lot".

One option could be to use the value of the monthly index of the month next to the release of each OECD Economic Outlook edition (e.g: January 1986 household index value against December 1985 OECD Economic Outlook unemployment rate forecast), but this would mean to assume that a new professional forecast spreads out within the first ten working days, probably too strong an assumption: the work of Carroll uses quarterly data, which implicitly allows to have a longer time span in order for the forecast to reach the agents.

An alternative is to use the average of the value of the indexes over the following six months: this solution allows both for some delay (in terms of months) in the acquisition of the forecast by an agent and to have a smoother measure of unemployment expectations.

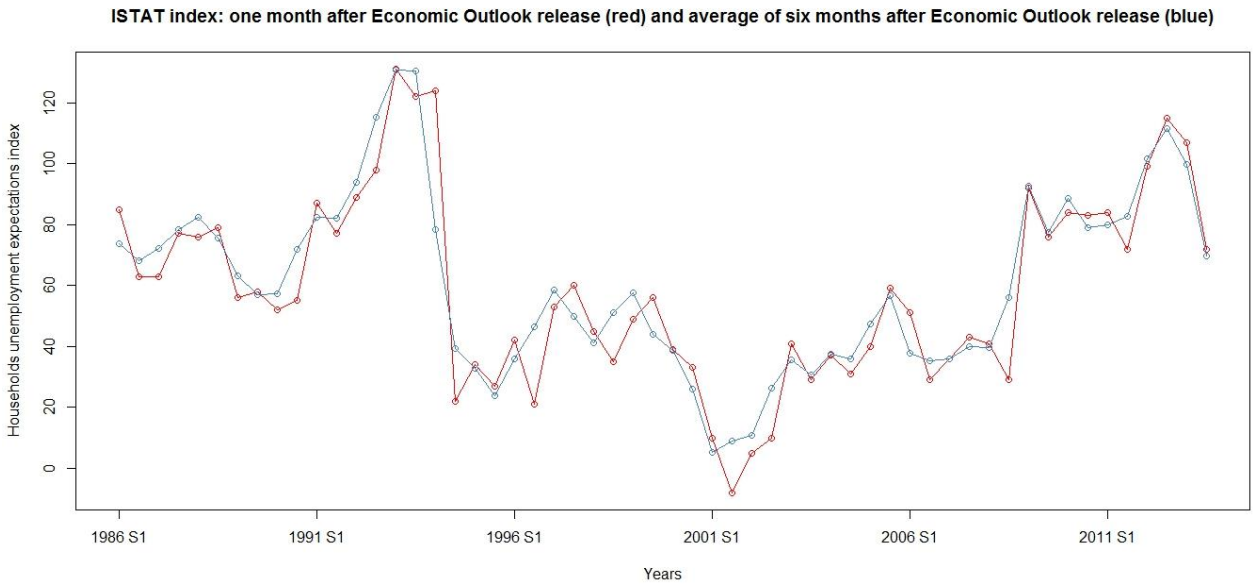


Figure 8: comparison of two households expectations indexes, one considering the value of the index the month after the Economic Outlook release (red), the other considering the average of the values of the index over the next six months (blue)

The two measures are not much different, as can be seen in Figure 8: their means are very similar (58.65 one-month version against 60.39 for the six-month version), the standard deviation of the difference between the two is 10.87 points, the maximum difference, in absolute value, is of 45.5. (Recall that the range of possible values is -200 to 200, that of the values actually recorded are between -8 and 131).

For the rest of the work, I will use the second measure (the average of the next six months). Moreover, I will refer to the period of the ISTAT household index: this means, for example, that "1995 S1", the first semester of 1995, refers to the average of the ISTAT households unemployment index from January 1995 to June 1995 and to the corresponding OECD Economic Outlook forecast, that is the one released in December 1994 supposed to spread out in the first semester of 1995.

THE MODEL

Carroll, in this basic "common-source" model, assumes that all individuals face the same probability of being "infected" but, making reference to other papers (in particular the one of Souleles (2001), who found that "there are highly statistically significant differences across demographic groups in forecasts of several macroeconomic variables"), did not exclude that there could be an heterogeneity in this probability across demographic groups.

So, I check if a similar conclusion could be drawn from Italian households microdata. Using again the *Indagine sui bilanci delle famiglie Italiane* data, in the 2011 survey[59] there are some questions concerning the expectations of the individual towards macroeconomic indicators in the near future. These questions are:

R1.3 According to you, on a range from 0 to 100, what is the probability that one year from now interest rates will be higher?

R1.4 According to you, on a range from 0 to 100, what is the probability that one year from now interest rates will be higher by more than one percentage point?[60]

R1.5 According to you, on a range from 0 to 100, what is the probability that investing now in the Italian Stock Exchange it is possible to have a positive return one year from now?

R1.6 According to you, on a range from 0 to 100, what is the probability that investing now in the Italian Stock Exchange it is possible to have a return higher than 10% one year from now?

R1.7 According to you, on a range from 0 to 100, what is the probability that one year from now housing prices will be lower?

R1.8 According to you, on a range from 0 to 100, what is the probability that one year from now housing prices will be lower by more than 10%?

I regress the answers to those questions on some demographic variables, that is gender, region, municipality dimension, employment status, educational level and age. Regressions are presented in Table 12 to Table 17. In every regression, at least four demographic regressors are statistically significant; in particular the educational level is a significant regressor for interest rates and stock market expectations, but not for housing prices. The reference groups are: Male, North, Employee, Less than 35 years old, Municipality dimension lower than 20000, "Less than high school" as education level.

[59] Which, I remember, refers to 2010.
[60] This questions are asked conditional on having reported a probability higher than zero in the previous one.

Table 12: regression of reported subjective probabilities to question R1.3 of *Indagine sui bilanci delle famiglie Italiane 2010* (Interest rates higher one year from now) on demographic variables. OLS

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	34.888	4.704	7.416	1.68E-13	***
D(Female)	-6.01	1.512	-3.974	7.28E-05	***
D(Center)	13.829	1.795	7.705	1.92E-14	***
D(South)	11.693	1.779	6.573	6.08E-11	***
D(Self-employed)	5.884	2.29	2.57	0.01024	*
D(Not employed)	-4.232	2.065	-2.049	0.04054	*
D(AGE 35-44)	2.738	4.243	0.645	0.51881	
D(AGE 45-54)	4.074	4.038	1.009	0.31314	
D(AGE 55-64)	1.617	3.994	0.405	0.68561	
D(AGE >65)	-1.369	4.247	-0.322	0.74714	
D(Municipality 20000-40000)	1.456	2.181	0.668	0.50439	
D(Municipality 40000-500000)	2.773	1.763	1.573	0.11585	
D(Municipality >500000)	5.66	2.851	1.986	0.04719	*
D(High school)	3.846	1.626	2.366	0.01808	*
D(Bachelor, Master or PhD)	6.266	2.232	2.807	0.00504	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34.44 on 2328 degrees of freedom (1792 observations deleted due to missingness)
Multiple R-squared: 0.07095, Adjusted R-squared: 0.06536 F-statistic: 12.7 on 14 and 2328 DF, p-value: < 2.2e-16

Table 13: regression of reported subjective probabilities to question R1.4 of *Indagine sui bilanci delle famiglie Italiane 2010* (Interest rates higher by more than one percentage point one year from now) on demographic variables. OLS

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	30.27041	5.12863	5.902	4.35E-09	***
D(Female)	-3.11026	1.68062	-1.851	0.0644	.
D(Center)	10.78951	2.01516	5.354	9.82E-08	***
D(South)	16.32735	1.90722	8.561	2.00E-16	***
D(Self-employed)	5.06983	2.38237	2.128	0.0335	*
D(Not employed)	-4.09933	2.22786	-1.84	0.0659	.
D(AGE 35-44)	-0.17307	4.58294	-0.038	0.9699	
D(AGE 45-54)	-0.48218	4.35128	-0.111	0.9118	
D(AGE 55-64)	-0.90114	4.32426	-0.208	0.8349	
D(AGE >65)	-1.37267	4.62639	-0.297	0.7667	
D(Municipality 20000-40000)	1.68164	2.39768	0.701	0.4832	
D(Municipality 40000-500000)	4.48546	1.99142	2.252	0.0244	*
D(Municipality >500000)	1.73802	3.12169	0.557	0.5778	
D(High school)	-0.09189	1.77053	-0.052	0.9586	
D(Bachelor, Master or PhD)	-4.19148	2.35778	-1.778	0.0756	.

Residual standard error: 31.32 on 1628 degrees of freedom (2492 observations deleted due to missingness)
Multiple R-squared: 0.0724, Adjusted R-squared: 0.06442 F-statistic: 9.076 on 14 and 1628 DF, p-value: < 2.2e-16

Table 14: regression of reported subjective probabilities to question R1.5 of *Indagine sui bilanci delle famiglie Italiane 2010* (Having a positive return in the Italian Stock Market one year from now) on demographic variables. OLS

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	29.1947	2.9409	9.927	2.00E-16	***
D(Female)	-6.6403	0.964	-6.888	7.56E-12	***
D(Center)	1.5934	1.1442	1.393	0.16389	
D(South)	-2.0794	1.1425	-1.82	0.0689	.
D(Self-employed)	0.9724	1.4646	0.664	0.50682	
D(Not employed)	0.2933	1.3171	0.223	0.82381	
D(AGE 35-44)	-3.084	2.6106	-1.181	0.23761	
D(AGE 45-54)	-2.5282	2.4641	-1.026	0.30503	
D(AGE 55-64)	-4.7549	2.4386	-1.95	0.05133	.
D(AGE >65)	-7.0211	2.6035	-2.697	0.00706	**
D(Municipality 20000-40000)	-3.3768	1.3641	-2.475	0.01339	*
D(Municipality 40000-500000)	-2.3437	1.1402	-2.056	0.03996	*
D(Municipality >500000)	2.0823	1.7789	1.171	0.2419	
D(High school)	4.6865	1.0385	4.513	6.78E-06	***
D(Bachelor, Master or PhD)	7.7565	1.428	5.432	6.26E-08	***

Residual standard error: 20.24 on 1986 degrees of freedom (2134 observations deleted due to missingness)
Multiple R-squared: 0.06894, Adjusted R-squared: 0.06238 F-statistic: 10.5 on 14 and 1986 DF, p-value: < 2.2e-16

Table 15: regression of reported subjective probabilities to question R1.6 of *Indagine sui bilanci delle famiglie Italiane 2010* (Having a return higher than 10% in the Italian Stock Market one year from now) on demographic variables. OLS

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	14.8916	2.6473	5.625	2.26E-08	***
D(Female)	-2.1413	0.8619	-2.484	0.0131	*
D(Center)	2.3213	1.0381	2.236	0.02551	*
D(South)	1.7818	1.0168	1.752	0.07993	.
D(Self-employed)	-1.7433	1.2528	-1.392	0.16429	
D(Not employed)	0.6156	1.1485	0.536	0.59203	
D(AGE 35-44)	-2.7167	2.3585	-1.152	0.24958	
D(AGE 45-54)	-2.4878	2.2257	-1.118	0.26388	
D(AGE 55-64)	-4.7563	2.2057	-2.156	0.03123	*
D(AGE >65)	-5.609	2.3646	-2.372	0.01783	*
D(Municipality 20000-40000)	-2.5579	1.2151	-2.105	0.03547	*
D(Municipality 40000-500000)	-2.1857	1.0154	-2.152	0.03154	*
D(Municipality >500000)	-1.4098	1.506	-0.936	0.34935	
D(High school)	2.8665	0.9239	3.103	0.00196	**
D(Bachelor, Master or PhD)	3.4668	1.2208	2.84	0.00458	**

Residual standard error: 14.58 on 1332 degrees of freedom (2788 observations deleted due to missingness)
Multiple R-squared: 0.03639, Adjusted R-squared: 0.02627 F-statistic: 3.593 on 14 and 1332 DF, p-value: 7.114e-06

Table 16: regression of reported subjective probabilities to question R1.7 of *Indagine sui bilanci delle famiglie Italiane 2010* (Housing prices lower one year from now) on demographic variables. OLS

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	27.7842	3.1344	8.864	2.00E-16	***
D(Female)	-3.0363	1.0185	-2.981	0.0029	**
D(Center)	2.6841	1.2367	2.17	0.03007	*
D(South)	-1.3683	1.1851	-1.155	0.24836	
D(Self-employed)	3.6351	1.6017	2.27	0.02332	*
D(Not employed)	-0.9542	1.409	-0.677	0.49835	
D(AGE 35-44)	1.5976	2.8415	0.562	0.57401	
D(AGE 45-54)	1.2657	2.6982	0.469	0.63905	
D(AGE 55-64)	-2.2615	2.6607	-0.85	0.39543	
D(AGE >65)	-2.0945	2.81	-0.745	0.45612	
D(Municipality 20000-40000)	-4.8752	1.4568	-3.347	0.00083	***
D(Municipality 40000-500000)	-3.8049	1.2051	-3.157	0.00161	**
D(Municipality >500000)	-5.8446	1.9076	-3.064	0.00221	**
D(High school)	-1.6464	1.1042	-1.491	0.13609	
D(Bachelor, Master or PhD)	-0.5158	1.5572	-0.331	0.74047	

Residual standard error: 24.05 on 2467 degrees of freedom (1653 observations deleted due to missingness)
Multiple R-squared: 0.02361, Adjusted R-squared: 0.01807 F-statistic: 4.261 on 14 and 2467 DF, p-value: 1.697e-07

Table 17: regression of reported subjective probabilities to question R1.8 of *Indagine sui bilanci delle famiglie Italiane 2011* (Housing prices lower by more than 10% one year from now) on demographic variables. OLS

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	21.14148	4.04564	5.226	1.99E-07	***
D(Female)	-3.30646	1.39791	-2.365	0.01815	*
D(Center)	10.57254	1.71884	6.151	9.97E-10	***
D(South)	6.59917	1.68245	3.922	9.18E-05	***
D(Self-employed)	4.13829	2.10065	1.97	0.04903	*
D(Not employed)	0.53557	1.96314	0.273	0.78504	
D(AGE 35-44)	3.72953	3.66746	1.017	0.30936	
D(AGE 45-54)	3.69128	3.45195	1.069	0.2851	
D(AGE 55-64)	1.86155	3.44226	0.541	0.58873	
D(AGE >65)	1.53885	3.69963	0.416	0.67751	
D(Municipality 20000-40000)	-6.76266	2.00646	-3.37	0.00077	***
D(Municipality 40000-500000)	-3.14292	1.65753	-1.896	0.05814	.
D(Municipality >500000)	-11.1122	2.59429	-4.283	1.96E-05	***
D(High school)	-0.06109	1.51899	-0.04	0.96792	
D(Bachelor, Master or PhD)	-2.97444	2.04841	-1.452	0.1467	

Residual standard error: 25.09 on 1431 degrees of freedom (2689 observations deleted due to missingness)
Multiple R-squared: 0.05549, Adjusted R-squared: 0.04625 F-statistic: 6.005 on 14 and 1431 DF, p-value: 1.184e-11

Consequently, the agent-based model developed in this thesis is based on the "common-source" model of Carroll, adding some other parameters. These parameters are:

- α_i, where $0 \leq \alpha_i \leq 1$: the probability the individual *i* in any period *t* obtains the latest professional forecast. If he obtains the forecast, the variable $F_{i,t}$ assumes value one, otherwise zero

$$p(F_{i,t}=1) = \alpha_i$$

- λ_i, where $0 \leq \lambda_i \leq 1$: the parameter according to which the individual *i* in any period *t* weights his previous expectation with the latest professional forecast obtained in that period, conditional on having obtained it ($F_{i,t}=1$). If the individual adopts the latest professional forecast ignoring his previous expectation, $\lambda_i=0$.
- β_i, where $0 \leq \beta_i \leq 1$: this persistency parameter has been added to allow for the hypothesis that, if no new information is acquired ($F_{i,t}=0$), the individual could assume that the rate would continue to follow the recent trend instead of not moving. For example, if in period *t* individual *i* receives from the news report a new professional forecast that changes his expectation by one percentage point ($\Delta u^e_{i,t,t-1} = 0.01$), if in the following periods he does not acquire new professional forecast, he assumes that $\Delta u^e_{i,t+1,t} = \beta_i \Delta u^e_{i,t,t-1} = \beta_i/100$, $\Delta u^e_{i,t+2,t+1} = \beta_i \Delta u^e_{i,t+1,t} = \beta_i^2 \Delta u^e_{i,t,t-1} = \beta_i^2/100$.

$$u^e_{i,t} = \begin{cases} \lambda_i u^e_{i,t-1} + (1-\lambda_i) u^f_t \ if \ F_{i,t} = 1 \\ u^e_{i,t-1} + \beta_i \left(u^e_{i,t-1} - u^e_{i,t-2} \right) if \ F_{i,t} = 0 \end{cases}$$

Where $u^e_{i,t}$ is the expectation of agent *i* in period *t* towards the near future, while u^f_t is the professional forecast in period *t* towards the near future. If the individual does not consider the trend, $\beta_i=0$

The model is composed by 1440 agents. In any of the 56 periods (one for each semester), for each agent there is a random draw from the uniform standard distribution U(0,1); if this draw is lower than α_i, the agents updates his expectation according to $u^e_{i,t} = \lambda_i u^e_{i,t-1} + (1-\lambda_i) u^f_{i,t}$, otherwise he assumes that the evolution of the macroeconomic indicator will continue to follow the recent trend $u^e_{i,t} = u^e_{i,t-1} + \beta_i \left(u^e_{i,t-1} - u^e_{i,t-2} \right)$. I also constrain $u^e_{i,t}$ to assume meaningful values, $0 < u^e_{i,t} < 1$, since for particular values of the parameters (α low and β high), these rules could lead some agents to have totally meaningless expectations[61].

[61] This because, with low values of α and high values of β, the agent rarely updates his expectations and believes that the unemployment rate will strongly follow the recent pattern. If, after some periods, he discovers that the unemployment rate is much different from what he expected (e.g. he thought unemployment rate was rising in a period of decreasing unemployment or viceversa), the high persistence parameters could lead the expectation to change (in absolute value) sharply also in the following periods, possibly reaching negative values or values above 1.

Then, an index is simulated adopting the same criteria of the ISTAT index. In order to have a simple model, we suppose each agent assumes that the value he expects for period *t* will be actually verified in period *t*: $u_{i,t}^{e} = u_{t+1}$.

Alternative but more complex models would take into consideration a time series with the current unemployment rate in any period and, more important, the agent should not only be informed about the professional forecast but also about the current unemployment rate in that period, an information that, in order to maintain the model coherent, should similarly spread out like a virus. This complicates the model: there should be two different "viruses" (one for the forecast and one for the current rate), possibly with two different probabilities, and in any period each agent could be "infected" by neither, one of the two or both.

Consequently, the agent compares his expectations of that period against the one of the previous one, calculating the change: $\Delta u_{i,t,t-1}^{e} = u_{i,t}^{e} - u_{i,t-1}^{e}$. Then, as a function of this change in expectation, he selects the modality for the question about his expectations towards the near future among "it will go up a lot", "it will go up a little", "it will remain stable", "it will go down a little" and "it will go down a lot".

This provides a further problem, since some thresholds have to be determined according to which the agent chooses the modality. But how to tell if, for a given agent, a 1 percentage point increase in the unemployment rate is perceived as an huge increase, leading to "it will go up a lot" as a response, or if it is perceived like a weaker increase ("it will go up a little") or no substantial increase at all ("it will remain stable")? The choice involves necessarily some degree of subjectivity.

The time series of the actual unemployment rate can help in providing a sensible range for the threshold. For the period 1985-2013, the time series of annual unemployment rate have a standard deviation equal to 1.83%, while the mean of the absolute change of the unemployment rate from one year to the following one is equal to 0.59%.

Next, I assume symmetry (i.e. a given percentage point change in absolute value is perceived equally, independently from its sign[62]) and that, for a given parameter *γ* to be calibrated, the agent would answer "it will go up a lot" if $\Delta u_{i,t-1,t}^{e} > \gamma_i$, "it will go up a little" if $\frac{\gamma_i}{2} \leq \Delta u_{i,t-1,t}^{e} \leq \gamma_i$, "it will remain stable" if

[62] This is a simplification, partly against the stream of literature in behavioral economics which sustains that losses "loom larger" than gains and in particular, according to experimental evidence, that "losses loom twice as much as gains" (Kahneman and Tversky (1979), Kahneman and Tversky (1992)). Anyway, I have tried to allow for this possibility further complicating the model and duplicating the parameters λ and γ, the two parameters according to which information are processed, with λ^{g} γ^{g} for gains and λ^{l} γ^{l} for losses. Anyway, when I calibrated the model looking for the parameters that maximize the correlation between simulated and actual data, the correlation did not increase so much and, more importantly, the parameters λ^{g} and λ^{l}, and γ^{g} and γ^{l}, turn out to be very similar, so I preferred to leave the model straightforward and be parsimonious on parameters.

$-\frac{\gamma_i}{2} < \Delta u^e_{i,t-1,t} < \frac{\gamma_i}{2}$, "it will go down a little" if $-\gamma_i \leq \Delta u^e_{i,t-1,t} \leq -\frac{\gamma_i}{2}$ and, finally, "it will go down a lot" if $\Delta u^e_{i,t-1,t} < -\gamma_i$.

The parameters α_i and λ_i, are heterogeneous as a function of the educational level of the individual, while, in order to leave the model parsimonious and ease the interpretation of the first two parameters (the crucial ones), β_i and γ_i are assumed to be homogeneous. Three categories of educational level are used: "Less than high school", "High school" and "Bachelor, Master or PhD".

THE NULL MODEL

Is such a model useful to explain the evolution of households expectations? In order to understand it, it is necessary to compare its explanatory potential with the one of other (and possibly simpler) models, the "null" models.

The simpler idea is to try to explain the evolution of households expectations with the evolution of OECD "professional forecasts". The sign of the correlation between the two series is positive, as expected: when professional forecasts go down and the unemployment rate is expected to decrease, reflecting optimism on the part of international institutions, also the value of households index goes down, indicating optimism also from households. On the other hand, the correlation coefficient is not so high: 20.13%. Moreover, running the regression of ISTAT index on OECD forecasts (Table 18), the coefficient on the forecasts is not statistically different from 0 even at the 10% level, and the R-squared is only 4.05%[63]. Forecasts alone do not appear to be a good candidate for describing households expectations.

Table 18: regression of ISTAT index on OECD unemployment forecasts. OLS
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	27.78	21.92	1.268	0.21
OECD forecast	320.38	212.19	1.51	0.137

Residual standard error: 28.82 on 54 degrees of freedom
Multiple R-squared: 0.04051, Adjusted R-squared: 0.02274
F-statistic: 2.28 on 1 and 54 DF, p-value: 0.1369

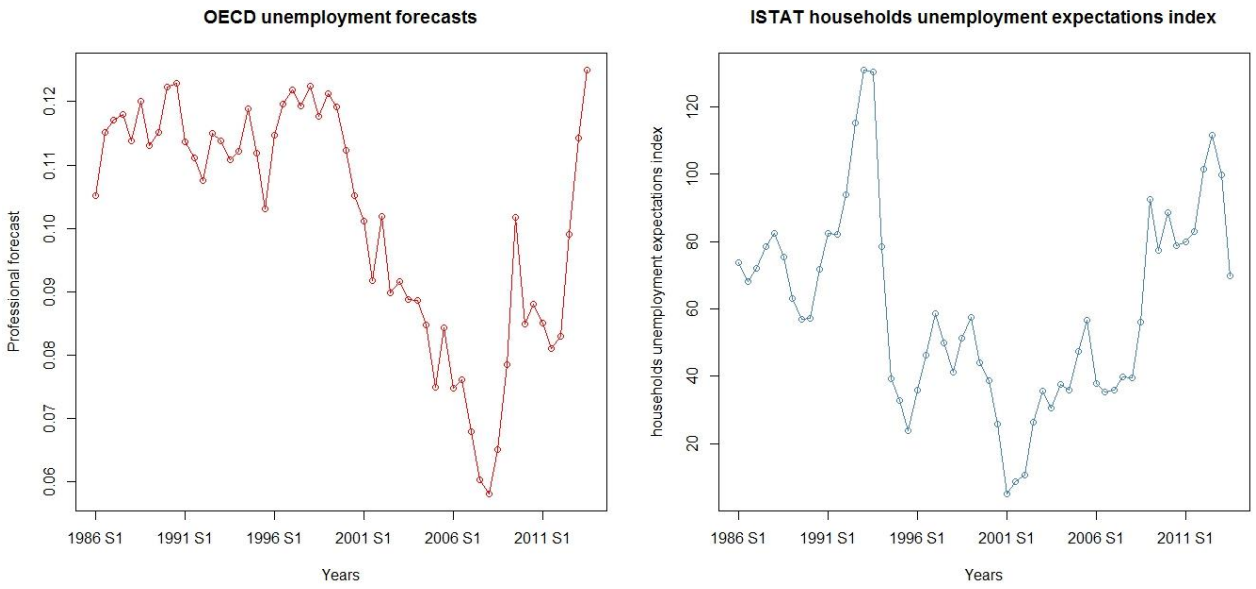


Figure 9: OECD unemployment forecasts (on the left) and ISTAT households unemployment expectations index (on the right)

[63] That is, actually the square of the correlation coefficient (20.13%).

The second option, a bit more complicated, is grounded on the idea that it is not possible to directly link the evolution of a variable defined in percentage points with an index defined with a different unit of measurement and that, more important, being the product of qualitative and not quantitative considerations, is a stepped function: a 1 percentage point increase in the expected unemployment rate change according to a given agent quantitative consideration do not necessarily imply that also his qualitative consideration has to change, while a smaller increase in another situation could lead the same agent to argue that the unemployment rate is no more "going to increase a little" but "going to increase a lot".

Consequently, in this model I assume that all agents are equal, they read the news report in any period and learn the new professional forecast and they assume it to be correct, without averaging it with their prior expectation. This, in terms of parameters values, means running the model letting $\alpha=1$, $\lambda=0$, $\beta=0$[64] and looking for the γ which maximizes the correlation between real and simulated index (which turned out to be $\gamma=0.012$, in a grid of values between 0.005 and 0.02 with 0.001 increments).

Table 19: regression of ISTAT index on simulated index with $\alpha=1$, $\lambda=0$, $\beta=0$ and $\gamma=0.012$. OLS

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	59.76304	3.69721	16.164	2.00E-16	***
simulated index value	0.11145	0.04124	2.702	0.00918	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.62 on 54 degrees of freedom F-statistic: 7.302 on 1 and 54 DF, p-value: 0.009184

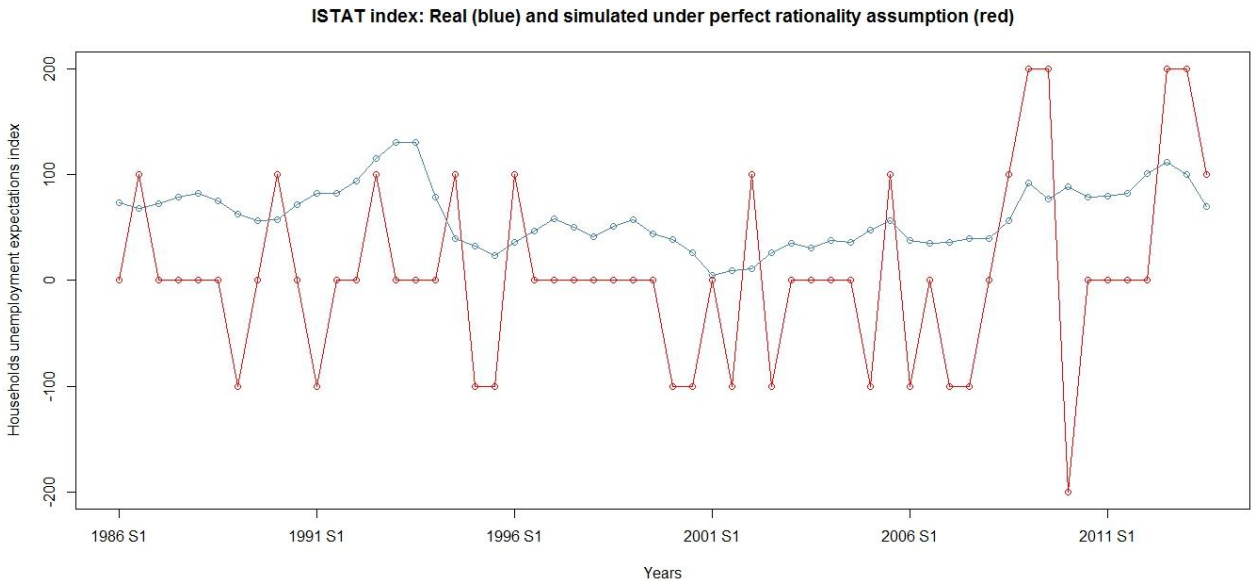


Figure 10: Simulated index (red) and ISTAT households unemployment expectations real index (blue)

[64] Even if the model works $\forall\beta \in [0,1]$: with $\alpha=1$, the agents always have an up-to-date forecast and never cares about the recent trend.

The simulated index, given the parameters, can assume only five values: -200, -100, 0, 100, 200. This because all the agents will receive the same information in each period (α_i=1 $\forall i$) and will react in the same way (β_i=0, λ_i=0, γ_i=0.012 $\forall i$), so in each period one of the modalities will be chosen by 100% of the agents. If everybody selects "It will go down a lot", the index will assume value -200 (-2*100); if everybody selects "It will go down a little", the index will assume value -100 (-1*100), and so on. The result is that the plot of the time series is everything but smooth (Figure 10): actually, it is characterized by sharp changes. Anyway, this is a better model than the one implied by the OECD forecasts only: the regressor is statistically significant at the 1% level, and the R-squared is equal to 11.91%.

INDEXES BY EDUCATIONAL GROUP

In order to look for a difference between households as a function of educational level, I exploit again ISTAT households confidence indexes. Data concerning unemployment expectations are collected since 1982, but since January 1995, together with the general index, other indexes for subsamples defined according to the educational level of the interviewee are available. These indexes are calculated using the same criteria of the general index, and three categories are available:

- Less than high school (*Nessun titolo, licenza elementare o media*)
- High school (*Diploma*)
- Bachelor, master o PhD (*Laurea e post-laurea*)

In Table 20 we report the statistics about ISTAT households expectations indexes. They are highly correlated: the lowest correlation coefficient is the one between "Less than high school" and "Bachelor, Master or PhD", equal to 0.9629. Anyway, it is immediately clear that the position indicators (mean and median) decrease as a function of educational level, while dispersion indicators (standard deviation) increase as a function of educational level.

Less educated households index never reached negative values between 1995 and 2013, meaning that, on average, they always expected rising unemployment: this is somewhat astonishing, in particular considering that between the very end of the 90ies and 2007 the unemployment rate has constantly decreased, almost halving (from about 12% to about 6%). Summarizing the findings, it appears that less educated individuals ("Less than high school") are more pessimistic (higher mean and median) and change less their opinion (lower standard deviation).

Table 20: Statistics of ISTAT households unemployment expectations index, 1995 S1-2013 S2, by educational level

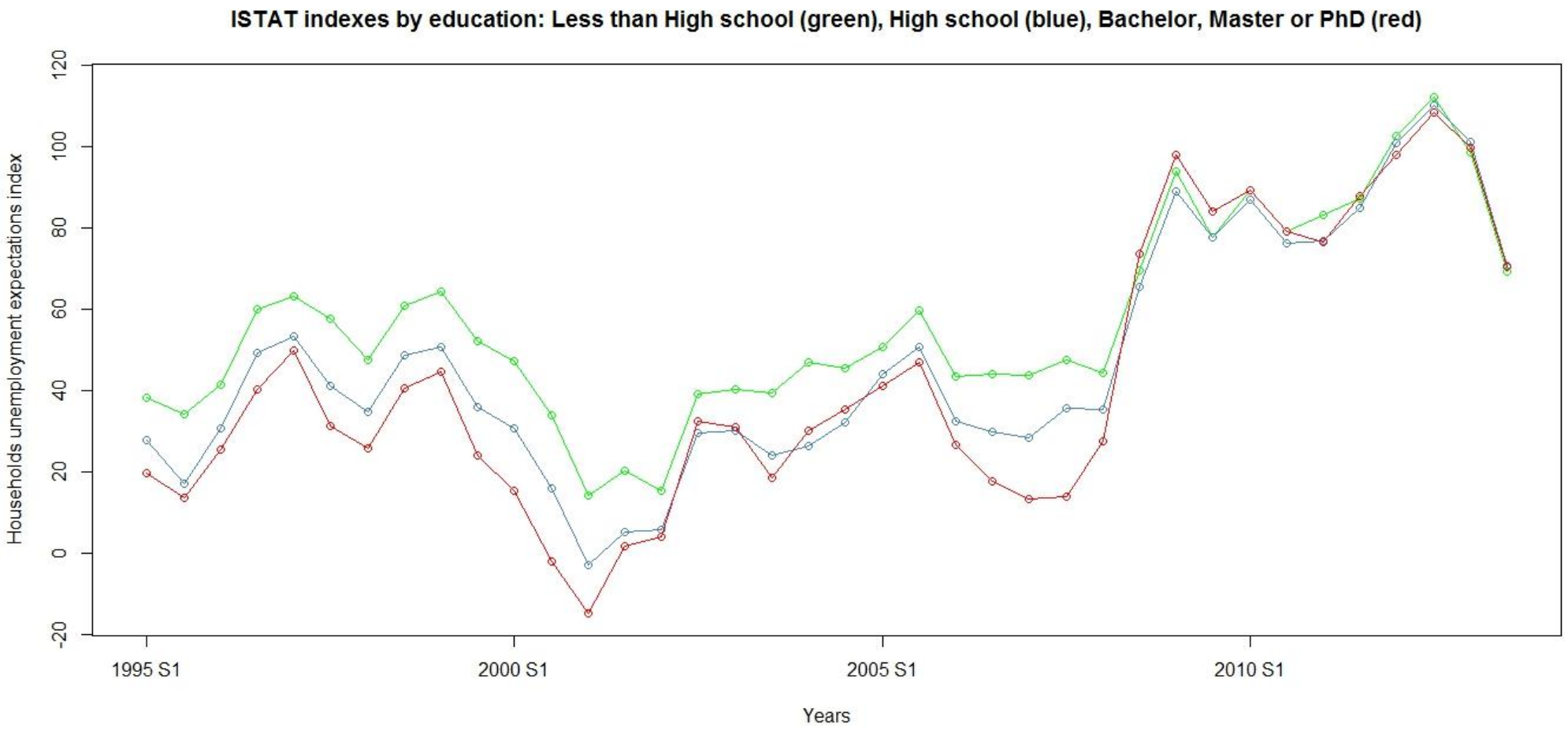
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev	Corr with Less…	Corr with High…	Corr with Bach…
Less than high school	14.17	41.96	49.17	56.78	69.46	112.2	23.80	1	0.98	0.96
High school	-2.833	29.71	35.75	46.9	69.38	110.2	28.56	0.98	1	0.99
Bachelor, Master or PhD	-14.83	18.79	31.83	42.57	72.83	108.5	32.64	0.96	0.99	1

Moreover, even if these general considerations are interesting and very useful, it is even more interesting to look at the evolution of in the indexes at the very beginning of the crisis. Looking at the pre-crisis period (2006 and 2007, periods 23 to 26 of Figure 11) the index for "Less than high school" is pretty stable, the one for "High school" draws a sort of U-shaped curve and, finally, the one for "Bachelor, Master or PhD" has a much more noticeable U-shaped pattern. The value of the index was definitely decreasing in the educational level: the average of the two semesters of 2007 (Periods 25 and 26) was 13.58 for "Bachelor, Master or PhD", around 32.09 for "High School" and around 45.67 for "Less then High school". They were all relatively "pessimistic", in the sense that on average they expected unemployment to rise, but to different extents. In 2008, the reaction is different: in the first semester (Period 27), the value of the index for "Bachelor, Master or PhD" starts increasing, almost doubling and coming back to the 2006 S1 level, while the other two do not make substantial moves. In the second semester (Period 28), all indexes grow, but to different extents: the value for "Bachelor, Master or PhD" increases by 46 points, starting to exceed the value for the other two educational levels. Comparing the second semester of 2007 and 2008, the index for "Bachelor, Master or PhD" has increased by 60 points, the one for "High school" by 30 points, the one for "Less than High school" by 22 points: the reaction of the "less educated" can be quantified to be between one half and a third of the one of the "more educated".

Table 21: ISTAT households unemployment expectations index, 2006 S1-2009 S2, by educational level

	2006 S1	2006 S2	2007 S1	2007 S2	2008 S1	2008 S2	2009 S1	2009 S2
Less than high school	43.33	44.00	43.83	47.50	44.33	69.50	94.00	77.67
High school	32.33	29.83	28.50	35.67	35.33	65.50	89.00	77.67
Bachelor, Master or PhD	26.67	17.67	13.33	13.83	27.50	73.67	97.83	84.17

Figure 11: ISTAT households unemployment expectations index, 1995 S1-2013 S2, by educational level



CALIBRATION OF HETEROGENEOUS PARAMETERS

In this section, I calibrate the different parameters by educational level. Even if expectation indexes by educational level are available only from the beginning of 1995, I let the model evolve for other 18 periods (18 semesters from 1986), exploiting the longer time series on OECD professional forecasts, before starting with the analysis, based on the real indexes and the last 36 periods of the simulated ones (actually, from 1995).

At a first stage, the calibration has been run letting all the four variables be heterogeneous and looking, for each educational level, at the values which maximize the correlation coefficient between the real and the simulated index. The interpretation is not easy, since there where nonlinearities in all the parameters but in α: the values of λ and β obtained were particularly high and, as explained in the section "The model", high values of β could lead to unrealistic expectations (below zero or too high) on the part of some individuals. This first-stage calibration and related statistics are presented in Appendix III.

Unfortunately, it is not easy to reduce the number of parameters and obtain a realistic model: β is realistically higher than 0, γ is necessary to convert the quantitative into qualitative considerations. Consequently, in this section I run the model by assuming an "heuristic" value of 0.01 for γ and a relatively low value of β (0.65), such as not to bring any problem of unrealistic expectations.

Then, keeping the values of β and γ constant, I calibrated the parameters α and λ for each educational group. Correlation alone is not a good criterion: looking at Table 20, it is possible to note that the indexes are all highly correlated (the lowest correlation is 0.96, between "Bachelor, Master or PhD" and "Less than High school"), so it is unrealistic to expect very different values. Actually, the combinations of α and λ that maximize the correlation and the simulated index are α=0.12 and λ=0 for "Less than High school", α=0.16 and λ=0 for "High school", α=0.14 and λ=0 for "Bachelor, Master or Phd". λ is always constant and α is nonlinear in the educational level. Moreover, these parameters are not able to capture the very different variances of the three indexes.

Consequently, the criterion used to select the optimal parameters is based on the correlation of the simulated index with the real one and a dispersion measure (the absolute value of the difference between the standard deviation of the real index and the standard deviation of the simulated one). For each couple of values (α, λ), ten simulations have been run, and the mean value of the correlation and the standard deviation across these simulations are considered for the criterion function:

$$\max_{\alpha,\lambda} \theta(\alpha,\lambda) = r_{s,r}(\alpha,\lambda) - 0.2 * |s_r - s_s(\alpha,\lambda)|$$

Where $r_{s,r}(\alpha,\lambda)$ is the mean of the correlation between the simulated and the real index, s_r the standard deviation of the real index and $s_s(\alpha,\lambda)$ the mean of the standard deviation of the simulated one. I look for the values of α and λ which maximize $\theta(\alpha,\lambda)$, by using an iterative procedure. I start with a two-dimensional matrix, where for each dimension there where 26 possible parameters values (between 0 and 0.5, by 0.02); then I refine the solution with another vector of 5 possible values for each parameter, this time by 0.01, centered around the first-stage solution.

Table 22: calibration of the model parameters, by education

	α	λ	β	γ
Less than high school	0.21	0.02	0.65	0.01
High school	0.24	0.06	0.65	0.01
Bachelor, Master or PhD	0.28	0.05	0.65	0.01

The results for α are quite different from those obtained by Carroll (2003), who found a value around 0.25 but for the quarterly[65] (and homogeneous), not biannual, infection probability for the US. I recall that the other parameters λ, β and γ are not present in the original paper by Carroll.

The most interesting parameter (α) is, as expected, increasing in the education level. An higher probability of learning the last professional forecast is related with an higher variability of the index. The intuition is that the higher the number of agents that "learn" the new professional forecast in each period, the more the index can change sharply as a consequence of the change of the pattern of "professionally forecasted" unemployment rate growth. On the other hand, if just a few agents have up-to-date information, the index necessarily has a smoother path: more periods are necessary to spread a new forecast.

The other parameter, λ, is nonlinear in the educational level. This complicates the interpretation. It could be that the less educated completely trust the forecasts, since they do not consider themselves sophisticated enough to develop a better prediction than trustworthy institutions, while the more educated trust forecasts because they know they are carried on in an accurate way but, at the same time, they know that not all variables are predictable and forecast may turn out not to be correct at all.

Table 23 Statistics of simulated households unemployment expectations index, 1995 S1-2013 S2, by educational level

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev
Less than high school	-42.44	-17.60	-6.05	-0.43	16.49	55.49	23.89
High school	-50.63	-22.77	-7.47	-2.07	19.69	65.92	28.75
Bachelor, Master or PhD	-55.61	-25.57	-7.81	-2.14	19.02	77.44	32.47

[65] On a biannual term, Carroll probability would be equal to 1-(0.75)^2=0.4375.

Table 24: Correlation matrix of simulated and real households unemployment expectations index, 1995 S1-2013 S2, by educational level

	Real index	Simulated: Less than high school	Simulated: High school	Simulated: Bachelor, Master or PhD
Simulated: Less than high school	0.76	1	0.99	0.98
Simulated: High school	0.74	0.99	1	0.99
Simulated: Bachelor, Master or PhD	0.73	0.98	0.99	1

Actually, the parameters for the three categories produce simulated indexes whose standard deviation is increasing as a function of educational level. Also, looking at the very beginning of the crisis in the simulated indexes, it is possible to see that in the second semester of year 2008 (period 28) we register an increase in all three indexes, with the absolute change that is increasing in the educational level, a result that is coherent with what is observed in real data.

Table 25: Simulated households unemployment expectations index, 2006 S1-2009 S2, by educational level

	2006 S1	2006 S2	2007 S1	2007 S2	2008 S1	2008 S2	2009 S1	2009 S2
Less than high school	-10.43	-11.83	-18.32	-29.15	-25.10	-8.56	18.40	45.99
High school	-18.91	-18.88	-23.82	-36.40	-33.58	-13.02	22.50	57.90
Bachelor, Master or PhD	-19.01	-20.72	-27.78	-44.44	-38.19	-11.64	33.37	71.10

One of the more marked differences between the evolution of the real index and the one of the simulated one can be found from 2009 S2 onwards, when the model performs poorly (the correlation coefficient in the last 10 semesters is even strongly negative: -0.53 for Less than High school, -0.46 for High school, -0.33 for Bachelor, Master or PhD). In the second semester of 2009, the simulated index rises sharply for all the three categories while the real one decreases a bit. This "anomaly" can be partly due to the pattern of the forecasts: in the second semester of 2009, according to the OECD Economic Outlook, an huge increase (from 7.85% to 10.18%: more than 2 percentage points) was predicted, to revert in the following semester to a "smoother" prediction of 8.50%.

Moreover, from 2011 there has been an exponentially growing attention of the mass media towards economic and financial indicators which until that moment where left to sophisticated users, like the well-known "spread" between the Italian and the German 10-year treasury bond interest rates. The latest value of the "spread" was always announced and interpreted[66] at any newscast, since it was considered the "litmus paper"[67] of the condition of our economy. Even if it is really disappointing that this model is not able to explain expectations development during the crisis, it could be that such an increased interest of the media and the public opinion towards economic subjects has changed the way expectations are developed. It may be, for example, that only educated agents had access to professional forecasts released

[66] I clearly remember a decrease in the "spread" attributed to the victory of the Italian football team against Germany the night before, in occasion of the Semi-finals of the European Championships.
[67] Other scientific term misused by Italian newscasts.

by OECD or other organizations, but also “Less than high school” agents could read the news on the spread dynamics. Moreover, looking at Figure 11 it is possible to appreciate that, if until 2008 the indexes tend to move in an almost parallel way but with different levels (with “optimism” increasing in education), from the beginning of the crisis the values of the indexes tended to converge to the same value, like if in this period also the expectations formation process had converged. The interpretation is similar to the one of Carroll (2003), who found that there was a decreasing standard deviation of individual expectations in a given period as a function of the newscasts coverage on that topic. Table 26 to Table 31 report the regression of the real indexes by education on the respective simulated ones, both with and without intercept, in order to estimate the explanatory potential of the model.

On the one hand, the regressions with the intercept (Table 26, Table 28, Table 30) have relevant R-squared values, all higher than 0.5, that is five times higher than the “best” null model: this confirms the explanatory power of this model. For each educational level, the intercept assumes a value similar (they differ only by some decimal digits) to the difference between the means of the real and the simulated index. Anyway, the coefficients on the simulated index are all far from 1: they are in a range between 0.73 and 0.76, values really similar to the correlation coefficients. This is not a surprise: I have calibrated the parameters so as to have the standard deviation of the simulated and real indexes as close as possible. Given that OLS regression coefficient is equal to $\beta_x = \frac{s_{xy}}{s_x^2}$, I get

$$\beta_x = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x s_y}\frac{s_y}{s_x} = r_{xy}\frac{s_y}{s_x} \approx r_{xy} * 1$$

In order to obtain an unitary regression coefficient, I would need a different criterion function, which considers the correlation and, instead of the standard deviation, the absolute deviation of the regression coefficient from 1. Anyway, this would mean obtaining a standard deviation for the simulated index much lower than the real one: I definitely prefer to sacrifice the unitary coefficient in order to have an index with similar variability.

On the other hand, the regressions without the intercept (Table 27, Table 29 and Table 31) have a very bad predictive power. This is due to the huge difference between the means of the simulated indexes and the real ones, reflected in intercepts between 43 and 59 when they are allowed. For “Less than high school” and for “High school”, the regression coefficient is statistically different from zero only at the 10% level, and the predictive power is lower than the null model. In case of “Bachelor, Master or PhD” the coefficient is statistically different from zero at the 5% level, but the regressions with the intercept are definitely to be preferred.

Table 26: regression of the real index on the simulated one, “Less than high school”

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	57.1069	2.5411	22.473	2e-16	***
Simulated Index	0.7579	0.1078	7.031	2.99e-08	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.66 on 36 degrees of freedom

Multiple R-squared: 0.5786, Adjusted R-squared: 0.5669

F-statistic: 49.43 on 1 and 36 DF, p-value: 2.991e-08

Table 27: regression of the real index on the simulated one, “Less than high school”, without intercept

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
Simulated Index	0.7137	0.4122	1.732	0.0917	.

Residual standard error: 59.89 on 37 degrees of freedom

Multiple R-squared: 0.07496, Adjusted R-squared: 0.04996

F-statistic: 2.998 on 1 and 37 DF, p-value: 0.09167

Table 28: regression of the real index on the simulated one, “High school”

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	48.4136	3.1883	15.185	2.00E-16	***
Simulated Index	0.7311	0.1121	6.522	1.4e-07	***

Residual standard error: 19.6 on 36 degrees of freedom

Multiple R-squared: 0.5416, Adjusted R-squared: 0.5288

F-statistic: 42.53 on 1 and 36 DF, p-value: 1.404e-07

Table 29: regression of the real index on the simulated one, “High school”, without intercept

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
Simulated Index	0.6075	0.3001	2.024	0.0502	.

Residual standard error: 52.62 on 37 degrees of freedom

Multiple R-squared: 0.09969, Adjusted R-squared: 0.07536

F-statistic: 4.097 on 1 and 37 DF, p-value: 0.05022

Table 30: regression of the real index on the simulated one, “Bachelor, Master or PhD”

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	44.1465	3.6559	12.076	3.21e-14	***
Simulated Index	0.7374	0.1139	6.476	1.61e-07	***

Residual standard error: 22.49 on 36 degrees of freedom

Multiple R-squared: 0.5381, Adjusted R-squared: 0.5253

F-statistic: 41.94 on 1 and 36 DF, p-value: 1.613e-07

Table 31: regression of the real index on the simulated one, “Bachelor, Master or PhD”, without intercept

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
Simulated Index	0.6456	0.2519	2.563	0.0146	*

Residual standard error: 49.85 on 37 degrees of freedom

Multiple R-squared: 0.1508, Adjusted R-squared: 0.1279

F-statistic: 6.571 on 1 and 37 DF, p-value: 0.01456

Figure 12: Simulated households unemployment expectations indexes, 1995 S1-2013 S2, by educational level

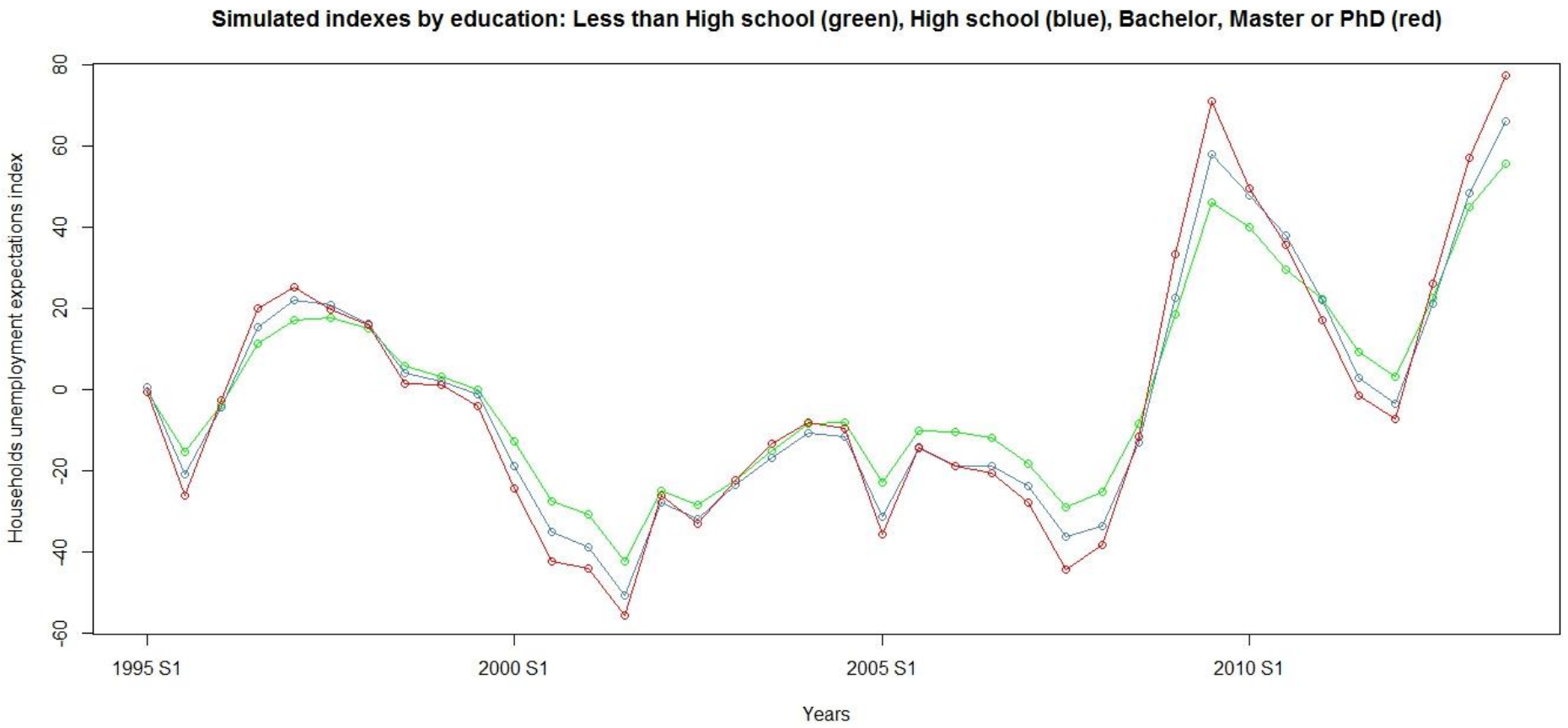
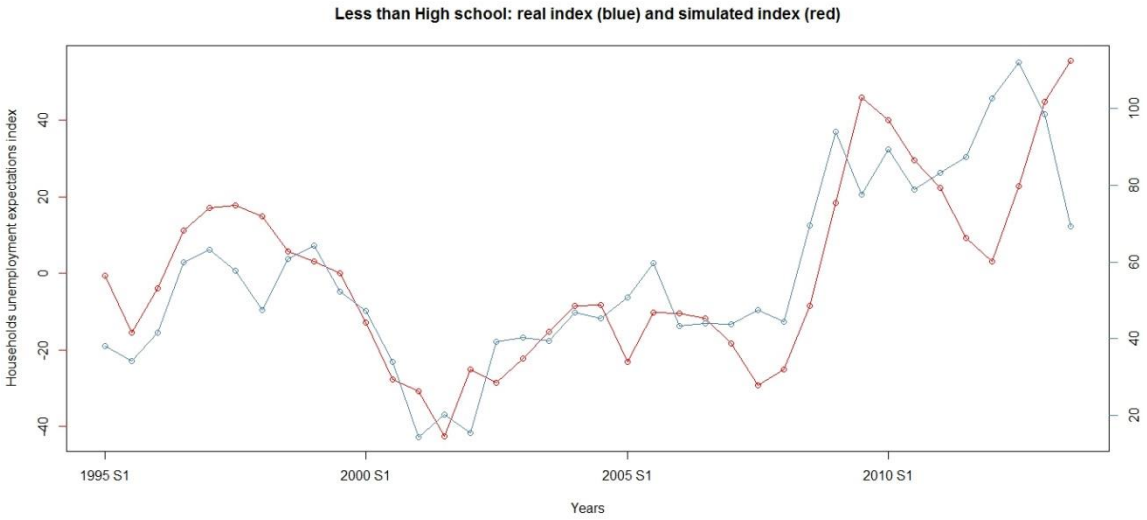
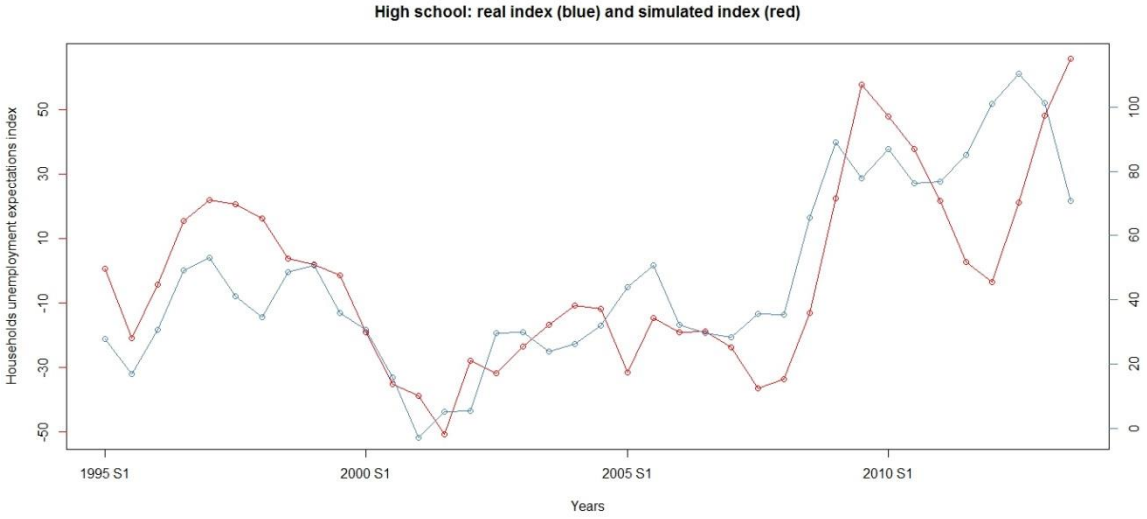
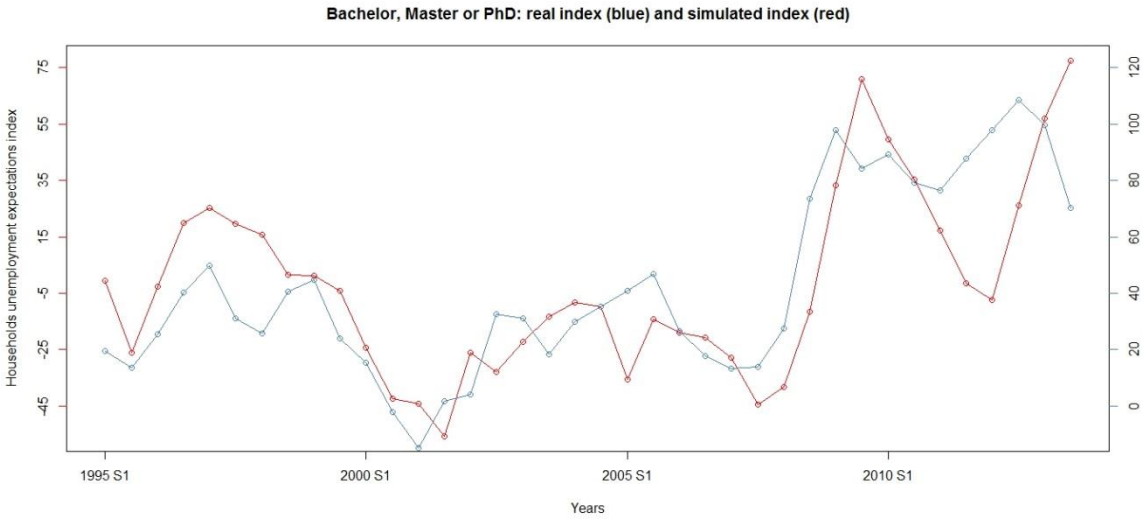


Figure 13: Comparison of simulated and real households unemployment expectations indexes, 1995 S1-2013 S2, by educational level







Note: observe that the scales for the two series are different: the left vertical axis refers to the simulated index, while the right one to the real index

A PROBLEM OF THE MODEL: POSITION INDICATORS

Anyway, even if the model is able to capture to some extent both the variability and the direction of the movements of real data indexes (the correlation coefficient between simulated and real data is never lower than 0.73), it is not able to capture at all the value of the position indicators (mean and median). On the one hand, as it is possible to see in Table 20, real indexes have positive values of the mean and of the median, "Less than high school" distribution never reaches negative values, while "High school" reaches negative values just once. On the other hand, the three simulated distributions differ in their variability measures, but are all centered around 0 (or values next to zero): this because the simulated index assumes positive values when the unemployment rate is expected to rise and negative values when it is expected to decrease. Why do real and simulated indexes differ so much in their position indicators?

The idea is that there could be a proportion of the population that behaves in a really different way with respect to the one described by the model, not trusting at all "official" forecasts and remaining stuck to their ideas of strongly or moderately rising unemployment, independently from what has happened recently to the unemployment rate or what trustworthy institutions are forecasting. Another interpretation of this assumption is that, among the ones who do not receive the latest forecast, a part forgets about the recent trend and, in absence of information, assumes a pessimistic opinion.

Such an arrangement could lead to an "overfitting" of the model on the data, increasing its capability of predicting the behavior of these indexes but, at the same time, dramatically deteriorating its general applicability in other situations. Anyway, the results are so interesting that it could be worthwhile to have a look at them.

In order to calculate the new parameters, I start from the values obtained in the section before and reported in Table 22. Then, I estimate the proportion *x* of individuals that should always select the modality "It will go up a lot" in order to equalize the means[68] of the simulated index and of the real one

$$2 * 100 * x + (1 - x)mean(simulated) = \ mean(real)$$

Of course, the standard deviation of the so transformed simulated index decreases: according to the usual formula for the linear transformation of a random variable, if $Y = a + bX, SD(Y) = b * SD(X)$. Here $a = 200x$ and $b = (1 - x) < 1$.

Consequently, starting from the so found value of *x,* I again look for the optimal parameters that maximize the (modified) criterion:

[68] The mean of the simulated index is obtained by considering the mean of the means among 10 simulation run with the given parameters.

$$\max_{\alpha,\lambda} \theta'(\alpha,\lambda) = r_{s,r}(\alpha,\lambda) - 0.2 * |(1-x) * s_r - s_s(\alpha,\lambda)|$$

Then, a new value of *x* is calculated: the procedure is repeated iteratively until *x* does not change by more than 0.005 in absolute value.

Table 32: calibration of the model parameters, by education

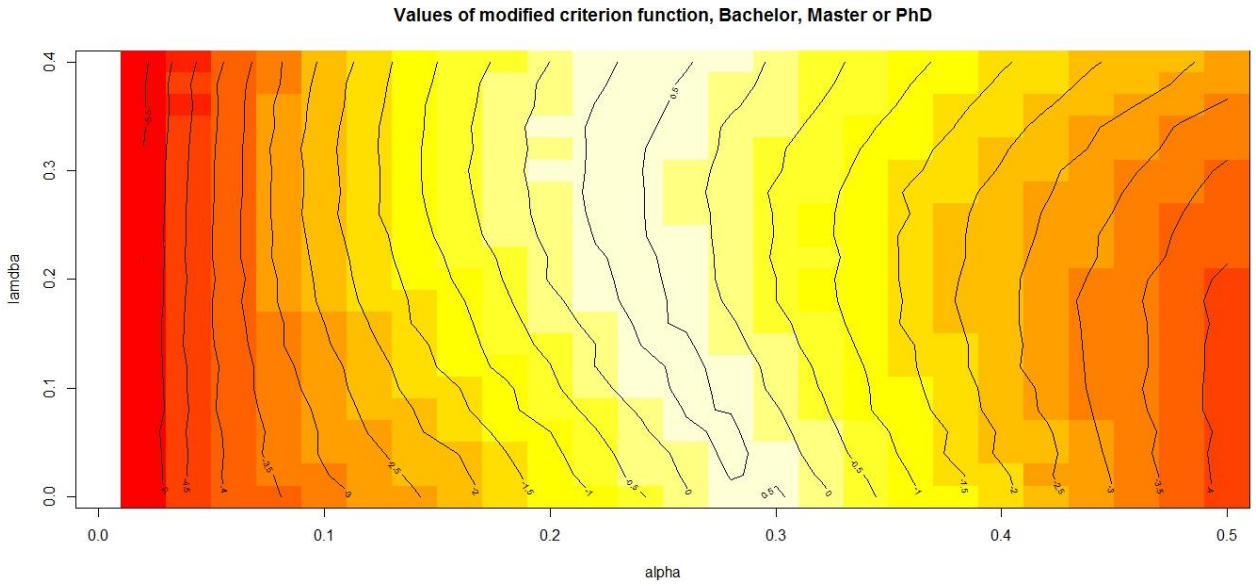
	α	λ	x	β	γ
Less than high school	0.26	0.40	0.293	0.65	0.01
High school	0.30	0.37	0.242	0.65	0.01
Bachelor, Master or PhD	0.36	0.36	0.217	0.65	0.01

The interpretation of the parameters is intuitive and straightforward:

- The probability that a given individual has access to up-to-date professional forecasts (α) is increasing in the educational level .
- λ, the parameter that determines the weight put on the individual previous forecast, conditional on having obtained a new professional forecast, is decreasing in the educational level. Less educated individuals weight more their previous expectations, maybe because they do not trust too much forecasts.
- The proportion of individuals in the population that always think in a pessimistic way (*x*), independently from forecast of the future or recent past trends, is decreasing in the educational level. There are less "irrationally pessimistic" individuals among more educated people. Moreover, this finding is someway consistent with results of regressions reported in Table 14 and Table 15 about expected returns on the Stock Market (an unambiguous sign of optimism) show that the proportion that thinks that one year from now there would be a return higher than 1% (question R1.5) or 10% (question R1.6) is increasing in the educational level.

An important difference between the values of Table 32 and Table 22 concerns the value of λ. According to the last calibration, agents put much more weight on their previous expectation. It is important to note, however, that the variable which affects more the value of the criterion function is α; λ has only a marginal effect. Looking, for example, at the level curves of the criterion function for the educational group "Bachelor, Master or PhD", which are shown in Figure 14, it could be noted that there is some trade-off between α and λ, but there is a narrow range of α for which the function assumes high values and, in this range, any level of λ provides similar results. The value of α seems to be more crucial than the value of λ, and this must be taken into account for the interpretation of the results.

Figure 14: values of $\theta'(\alpha, \lambda)$ as a function of α and λ, level curves. Educational level: Bachelor, Master or PhD



Also, looking at the values assumed by the simulated indexes at very beginning of the current crisis, it is possible to see that in the second semester of year 2008 (period 28) we register an increase in all three indexes, with the absolute change that is increasing in the educational level, a result that is coherent with what is observed in real data.

Table 33: Simulated households unemployment expectations index, 2006 S1-2009 S2, by educational level

	2006 S1	2006 S2	2007 S1	2007 S2	2008 S1	2008 S2	2009 S1	2009 S2
Less than high school	41.52	47.66	39.89	29.39	25.78	40.38	72.64	101.54
High school	30.36	33.99	27.07	14.87	11.16	29.40	69.21	102.74
Bachelor, Master or PhD	23.75	29.00	20.74	3.46	2.66	27.91	76.06	112.43

Table 34 Statistics of simulated households unemployment expectations index linearly transformed such as to make the mean of the simulated index equal to the mean of the real one, 1995 S1-2013 S2, by educational level

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev	Corr. w/ Real index
Less than high school - shifted	19.99	40.01	53.27	56.74	68.22	115.3	23.87	0.74
High school – shifted	2.378	26.76	41.74	46.81	63.06	116	28.23	0.75
Bachelor, Master or PhD - shifted	-8.05	21.49	36.02	42.57	55.82	125.5	32.61	0.74

Table 35: Comparison of statistics of real, simulated and shifted simulated households unemployment expectations indexes, 1995 S1-2013 S2, by educational level

Less than high school

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev
Real	14.17	41.96	49.17	56.78	69.46	112.2	23.80
Simulated	-42.44	-17.60	-6.05	-0.43	16.49	55.49	23.89
Simulated – shifted	19.99	40.01	53.27	56.74	68.22	115.3	23.87

High school

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev
Real	-2.83	29.71	35.75	46.9	69.38	110.2	28.56
Simulated	-50.63	-22.77	-7.47	-2.07	19.69	65.92	28.75
Simulated – shifted	2.378	26.76	41.74	46.81	63.06	116	28.22

Bachelor, Master or PhD

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev
Real	-14.83	18.79	31.83	42.57	72.83	108.5	32.64
Simulated	-55.61	-25.57	-7.81	-2.14	19.02	77.44	32.47
Simulated – shifted	-8.05	21.49	36.02	42.57	55.82	125.5	32.61

Table 36 to Table 41 report the results of the regressions of the real indexes by education on the respective transformed simulated ones, both with and without intercept, in order to have an estimate of the explanatory potential of the model.

Looking at the regressions with the intercept, the regression coefficients are all statistically significant at the 0.1% level and assume values between 0.73 and 0.76: for this reason, the intercept is different from zero ($\beta_0 = \bar{y} - \beta_1 \bar{x}_1 \approx (1 - \beta_1)\bar{y}$). It is important to note, however, that the intercepts are not highly statistically significant: the one for "Less than high school" is significant at the 5% level, while the other two are significant only at the 10% level. The R-squared are higher than 0.5, similarly to those of Table 26, Table 28, Table 30. This model is better than the null one, too.

Given that the intercepts are not statistically significant, I run the regressions without the intercept (Table 37, Table 39 and Table 41). The coefficients are positive and assume values much closer to the unitary value and are all statistically significant at the 0.1% level. The R-squared seems to suggests that the fit is extremely good (Table 37 reports a striking 0.92), but it is necessary to remember that, for technical reasons related to the calculation of the R-squared of a regression without the intercept, it is not possible to compare it with the R-squared of a regression with the intercept[69].

[69] Moreover, it would not make sense that a regression with the same the regressors plus one (the intercept) has a lower fit than another one with the same regressors and no intercept.

Table 36: regression of the real index on the shifted simulated one, “Less than high school”

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	14.9231	6.8697	2.172	0.0365	*
Simulated – shifted	0.7377	0.1118	6.597	1.12e-07	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 16.23 on 36 degrees of freedom
Multiple R-squared: 0.5473, Adjusted R-squared: 0.5347
F-statistic: 43.52 on 1 and 36 DF, p-value: 1.115e-07

Table 37: regression of the real index on the shifted simulated one, “Less than high school”, without intercept

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
Simulated – shifted	0.96205	0.04497	21.39	<2e-16	***

Residual standard error: 17.03 on 37 degrees of freedom
Multiple R-squared: 0.9252, Adjusted R-squared: 0.9232
F-statistic: 457.7 on 1 and 37 DF, p-value: < 2.2e-16

Table 38: regression of the real index on the shifted simulated one, “High school”

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	11.4421	6.0876	1.880	0.0683	.
Simulated – shifted	0.7576	0.1118	6.778	6.42e-08	***

Residual standard error: 19.19 on 36 degrees of freedom
Multiple R-squared: 0.5607, Adjusted R-squared: 0.5485
F-statistic: 45.95 on 1 and 36 DF, p-value: 6.421e-08

Table 39: regression of the real index on the shifted simulated one, “High school”, without intercept

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
Simulated – shifted	0.93814	0.05908	15.88	<2e-16	***

Residual standard error: 19.84 on 37 degrees of freedom
Multiple R-squared: 0.8721, Adjusted R-squared: 0.8686
F-statistic: 252.2 on 1 and 37 DF, p-value: < 2.2e-16

Table 40: regression of the real index on the shifted simulated one, “Bachelor, Master and PhD”

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	11.1535	6.0122	1.855	0.0718	.
Simulated – shifted	0.7380	0.1127	6.550	1.29e-07	***

Residual standard error: 22.35 on 36 degrees of freedom
Multiple R-squared: 0.5437, Adjusted R-squared: 0.531
F-statistic: 42.9 on 1 and 36 DF, p-value: 1.289e-07

Table 41: regression of the real index on the shifted simulated one, “Bachelor, Master and PhD”, without intercept

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
Simulated – shifted	0.90472	0.07015	12.9	2.93e-15	***

Residual standard error: 23.07 on 37 degrees of freedom
Multiple R-squared: 0.818, Adjusted R-squared: 0.8131
F-statistic: 166.3 on 1 and 37 DF, p-value: 2.927e-15

Figure 15: Simulated households unemployment expectations indexes, 1995 S1-2013 S2, by educational level

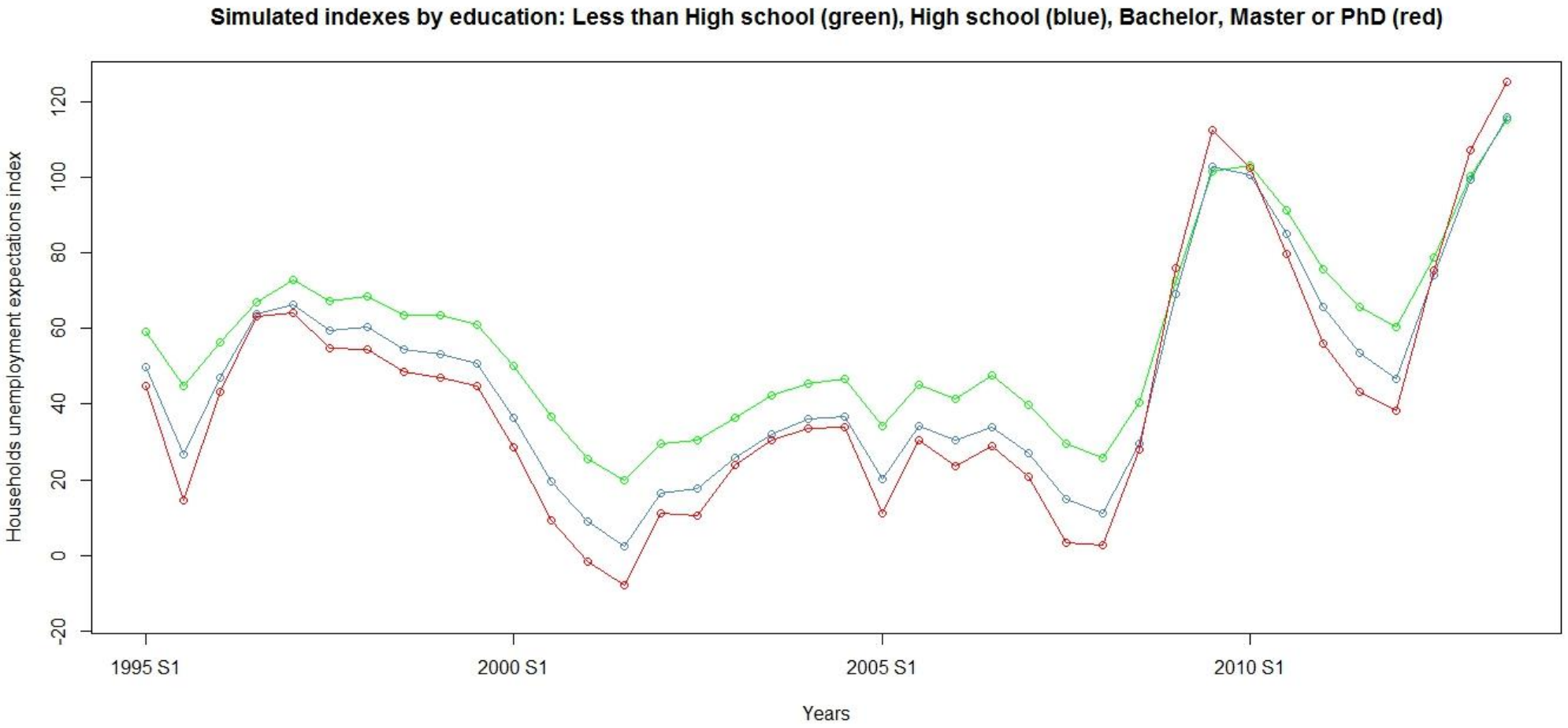
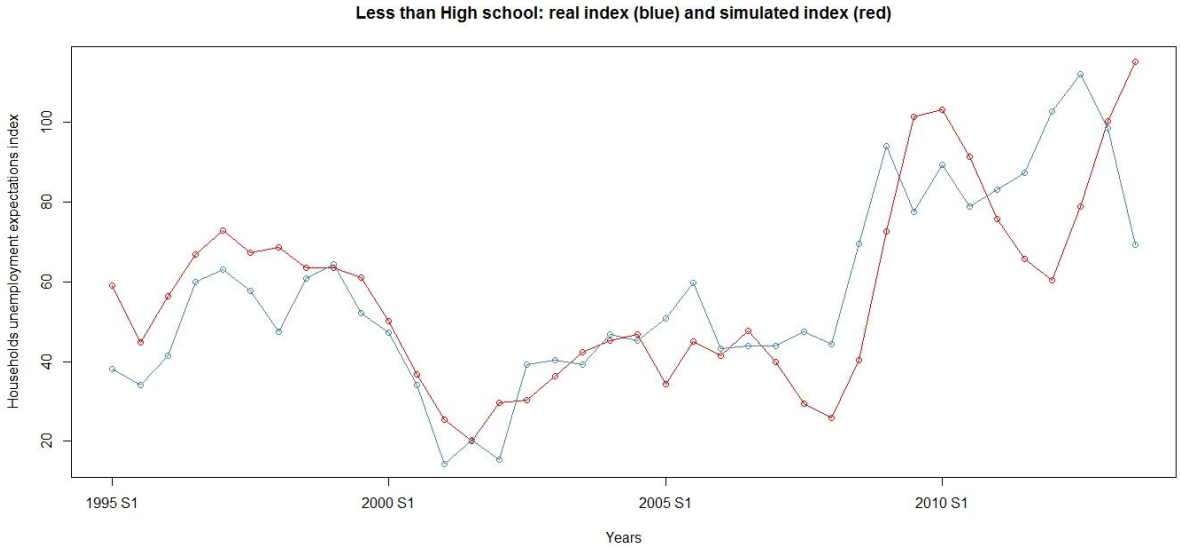
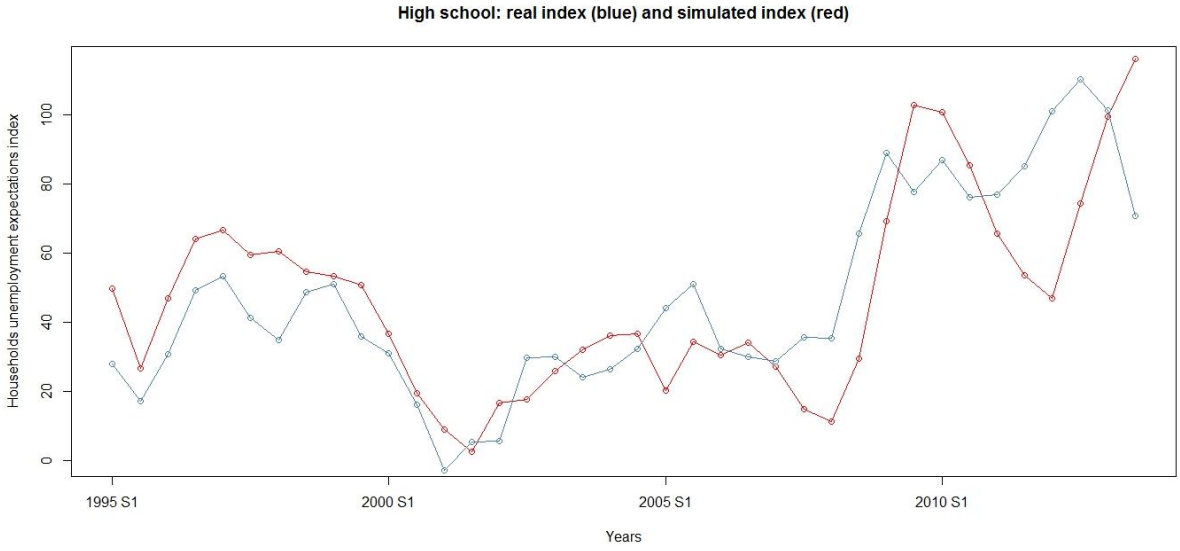
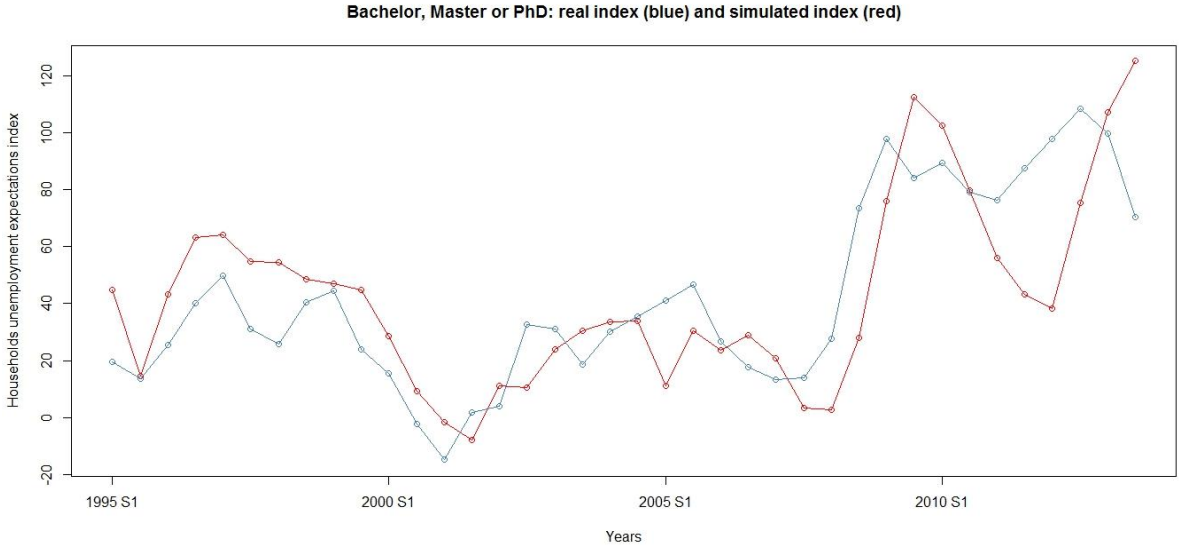


Figure 16 Comparison of simulated and real households unemployment expectations indexes, 1995 S1-2013 S2, by educational level







IV. CONCLUSIONS

In this thesis, I started from a review of the literature on precautionary savings, that predicts how saving behavior should react to a given shock, and I observed that the reaction of Italian households at the very beginning of the current crisis has been milder than what I would have expected, under the assumption of a fully rational expectation.

In order to look for an explanation of this different and apparently irrational behaviour, I find that an important driver of behaviour is the educational level. I develop an agent-based model under the hypothesis that individuals are not irrational but make a particular kind of choice due to incomplete or imperfect information. Hence, these decisions are not irrational, but may be described as rational ones based on the wrong ground.

The model is simple, but it is a good starting point for further extensions and is able to describe, with a potentially realistic story, the process that leads to expectations development as a function of educational level, with the more educated that are also the ones whose prediction is closer to the state of the art (the professional forecast). The results are interpretable in a very simple and intuitive way: the higher the educational level, the closer is the average expectation to the rational one, represented by the professional forecast. The probability of being informed of the latest professional forecast is increasing in the educational level: the more educated read more newspapers, and so they have an higher probability of learning the new forecast. There is also a lower proportion of “irrationally pessimistic” or of individuals who forget about the recent trend among the more educated, and the weight put on the individual previous forecast, conditional on having obtained a new professional forecast, is decreasing in the educational level: the more educated pay more attention to and trust more economic news than the less educated. The fact that the average expectations of the more educated are the one closest to the Rational Expectations Model prediction, even if this is a very intuitive and reasonable consideration, is another clue of the importance of education to provide the ability to understand economic phenomena and react “rationally”. This has effect on the business cycle, too: waves of excessive optimism and pessimism have frequently caused business cycle fluctuations (Milani 2011), turning out to be self-fulfilling prophecies.

I believe these results are interesting and that merit a deeper study. For example, looking at the households unemployment expectation indexes from 1995 onwards, the three sub-indexes by education are highly correlated but differ in the mean and standard deviation, supporting the hypothesis that expectations are developed differently, but in the last observations (since 2009-2010) the series appear to collapse into one unique curve, and the correlation of the simulated with the real indexes drop from around 0.75 to negative values. These two findings together could suggest that in periods of heightened uncertainty expectations are developed in different ways than during normal times, so it could be

interesting to study in depth these problematic periods, tentatively related to the spread crisis occurred in Italy in 2011.

A second point concerns the combination of micro and macrodata in order to calibrate the model more precisely. Considering some theoretical works (for example, Carroll (1997) on precautionary savings) an extension of the agent-based model could be developed, adding some rules that govern the behavior of the agent as a function of his expectation: this should be compared with actual behavior, to check if simulated expectations coupled with existing theoretical models could explain the observed behavior. A similar idea is applied in a totally different setup in Lettau (1997).

I adopted only one of the simplest agent-based models, which Carroll denoted as "common-source" model, according to which the agent can obtain the new professional forecast only from an unique source (newscasts), without taking into consideration, for example, interaction between the agents. Even if the fit is quite good richer models, which allow for the interaction between the agents, or featuring non-constant and endogenously defined parameters in the different periods, can explain more precisely the data and the expectation formation process.

Moreover, this kind of agent-based model could be adopted to study expectations development also for other macroeconomic indicators: the most trivial example is inflation, which is another driver of economic decisions, or more sophisticated indicators like long-term interest rates, which could drive important household investment decisions like housing expenditure or choice between fixed rate and variable rate mortgages. It would be very interesting to appreciate if the expectation formation process is different for different indicators, and if there is increasing divergence in the degree of "rationality" of consumers' expectations as a function of education in the degree of sophistication of the indicator considered.

REFERENCES

Bartiloro, Laura, and Cristiana Rampazzi. «Il risparmio e la ricchezza delle famiglie italiane durante la crisi.» *Questioni di Economia e Finanza*, n. 148 (February 2013): 1-31.

Branch, William A. «The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations.» *The Economic Journal*, n. 114 (2004): 592–621.

Caballero, Ricardo. «Earnings Uncertainty and Aggregate Wealth Accumulation.» *American Economic Review* 81 (1991): 859-871.

Cagetti, Marco. «Wealth Accumulation over the Life Cycle and Precautionary Savings.» *Journal of business & Economic Statistics* 21, n. 3 (2003): 339-353.

Carroll, Christopher D. «Buffer-Stock saving and the Life Cycle/Permanent Income Hypothesis.» *The Quarterly Journal of Economics* 112 (1997): 1-55.

Carroll, Christopher D. «Macroeconomic Expectations of Households and Professional Forecasters.» *The Quarterly Journal of Economics*, 2003: 269-298.

Carroll, Christopher D. «The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence.» *Brookings Papers on Economic Activity* 2 (1992): 62-156.

Carroll, Christopher D. «The Epidemiology of Macroeconomic Expectations.» *Working paper*, 2003: 1-25.

Deaton, Angus. «Saving and Liquidity Constraints.» *Econometrica* 59, n. 5 (1991): 1221-1248.

Engen, Eric M., and Jonathan Gruber. «Unemployment insurance and precautionary saving.» *Journal of Monetary Economics* 41 (2001): 545-579.

Giavazzi, Francesco, and Michael MacMahon. «Policy Uncertainty and Household Savings.» *The Review of Economics and Statistics* 94, n. 2 (May 2012): 517-531.

Gourinchas, Pierre-Olivier, and Jonathan A. Parker. «Consumption over the Life Cycle.» *Econometrica* 70 (2002): 48-89.

Hamilton, J. D. «A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.» *Econometrica* 57 (1989): 357-384.

Kahneman, Daniel, and Amos Tversky. «Advances in Prospect Theory: cumulative representation of uncertainty.» *Journal of Risk and Uncertainty* 5 (1992): 297-323.

Kahneman, Daniel, and Amos Tversky. «Prospect Theory, an Analysis of Decision under Risk.» *Econometrica* 47 (March 1979): 263-292.

LeBaron, Blake. «Agent-based computational finance: Suggested readings and early research.» *Journal of Economic Dynamics and Control* 24 (2000): 679-702.

Leland, Hayne E. «Saving and Uncertainty: The Precautionary Demand for Saving.» *The Quarterly Journal of Economics* 82, n. 3 (1968): 465:473.

Lettau, Martin. «Explaining the facts with adaptive agents: The case of mutual fund flows.» *Journal of Economic Dynamics and Control* 21 (1997): 1117-1147.

Lusardi, Annamaria, and Olivia S. Mitchell. «The Economic Importance of Financial Literacy: Theory and Evidence.» *NBER WORKING PAPER SERIES*, n. 18952 (2013).

Milani, Fabio. «Expectations Shocks and Learning as Drivers of the Business Cycle.» *The Economic Journal*, 2011: 379-401.

Roberts, John M. «Inflation Expectations and the Transmission of Monetary Policy.» *Federal Reserve Board FEDS working paper*, n. 1998-43 (1998).

Skinner, Jonathan. «Risky Income, Life Cycle Consumption, and Precautionary Savings.» *Journal of Monetary Economics*, n. 22 (1988): 237-255.

Souleles, Nicholas. «Consumer Sentiment: Its Rationality and Usefulness in Forecasting Expenditure; Evidence from the Michigan Micro Data.» *NBER WORKING PAPER SERIES*, n. 8410 (2001).

Zeldes, Stephen P. «Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence.» *Quarterly Journal of Economics* 104 (1989): 275:298.

APPENDIX I: MICRODATA ANALYSIS

The survey *Indagine sui bilanci delle famiglie italiane* collects data on about 8000 households and 20000 individuals each edition; about 300 different municipalities are represented. This is a panel study, with the same household tracked for different years, even if the attrition rate is relatively high (between 42% and 50% for the considered years).

These are the precise numbers: for 2010 there are 7951 households (of which 4621 are panel) and 19836 individuals; for 2008, 7977 households (of which 4345 panel) and 19907 individuals; for 2006, 7768 households (of which 3957 panel) and 19551 individuals.

Data for each survey is collected in different datasets; the drivers are NQUEST, the identity number of the household[70], and when information about the individuals are considered this is coupled with NORD, the ordinal number of the individual in the household (NORD=1 denotes the reference person for the household). For the purposes of this paper, I have used the datasets *QXXA*[71] (household structure), *LAVORO* (Job information), *CARCOMXX* (Individual characteristics), *ALLB1* (attachment, for employees only), *RFAMXX* (household income) *RISFAMXX* (household consumption and saving) and *RICFAMXX* (household wealth).

Now I will explain how the dataset to compare years 2008 and 2010 has been constructed; the one for years 2006 and 2008 has been constructed in a similar way, with a pair of minor differences that will be pointed out.

First of all, only panel households are considered (QUEST=3; the variable can be found in the *QXXA* dataset), and for each household only the reference person (NORD=1) is considered; as a further selection, it has been checked that the reference person was the same of the previous survey (NORDP=1; the variable can be found in the *CARCOMXX* dataset). Consequently, for the couple of years 2008-2010, the individuals that in the 2010 survey respond to the previous criteria are selected, and an unique dataset with all the variables collected in the datasets *QXXA, LAVORO, CARCOMXX*[72]*, RFAMXX, RISFAMXX* and *RICFAMXX* is created. Then, the same individuals are selected in the 2008 survey, creating an equivalent dataset.

Unfortunately, not all the variables have remained the same in the different surveys; consequently, to construct the dataset with information for both years 2008 and 2010, I was forced to consider only the

[70] Panel households have the same identity number (NQUEST) of previous surveys.
[71] Where XX denotes the last two digits of the survey year: e.g., for 2010 *Q10A.*
[72] For *QXXA*, *LAVORO*, *CARCOMXX*, the driver is NQUEST NORD; for *RFAMXX*, *RISFAMXX*, *RICFAMXX*, the driver is NQUEST.

variables which have maintained the same name and structure, and to check for the other ones whether they have simply changed name or if they do not exist in one of the two surveys.

For example, the sector in which the individual was (or has been, for retirees and unemployed) working in the 2010 survey was recorded under *NACE*, in the 2008 one under *APSETT*, two categorical variables where the different categories are identified with a number. Unfortunately, also the definition of the categories have changed between the two survey: for 2010 there are 21 narrower categories, for 2008 only 11. Using the definitions, I redefined *NACE* according to *APSETT* criteria:

NACE=1→APSETT=1	NACE=8→APSETT=5	NACE=15→APSETT=9
NACE=2→APSETT=2	NACE=9→APSETT=4	NACE=16→APSETT=9
NACE=3→APSETT=2	NACE=10→APSETT=5	NACE=17→APSETT=9
NACE=4→APSETT=2	NACE=11→APSETT=6	NACE=18→APSETT=11
NACE=5→APSETT=2	NACE=12→APSETT=7	NACE=19→APSETT=7
NACE=6→APSETT=3	NACE=13→APSETT=7	NACE=20→APSETT=8
NACE=7→APSETT=4	NACE=14→APSETT=7	NACE=21→APSETT=10

Unfortunately, this prevented me from using this variable as a definition of "civil servant"; NACE category number 15 concerns "public administration and defense, social security"[73], while the broader category APSETT=9 includes also Education and Health services: there are private workers also in the education field and, to a greater extent, in the Health sector, so using the workers with APSETT=9 as civil servants and the others as private workers would result in a very inaccurate definition of public sector.

So, in order to create a dummy "***PUBLIC***", I decided to use the variable PUBBLICO found in the attachment *ALLB1*, reserved to employees only. Given that all civil servants are employees, this means that all the self-employed are not civil servants: for this reason, I define the dummy PUBLIC equal to one for all those who have a value of PUBBLICO equal to 1[74], and 0 for everybody else.

Unfortunately, even this definition was not perfect for the purposes of this study: employees of companies like *Trenitalia* or *Poste Italiane*, which are Spa, are explicitly considered not to be part of the definition of Public Administration; anyway, from an uncertainty point of view, the employment situation is extremely stable. The inclusion of this workers with fewer precautionary motives in the control group of non-civil servants would probably bias our estimates of the diff-in-diff model towards zero.

Moreover, in the comparison between employees and self-employed workers, I excluded from the employees temporary workers ("temp"). This because the goal of the analysis is to compare the behavior of relatively stable employees with relatively unstable self-employed workers, but "temp" are in an employment situation that could be considered as more unstable of the one of self-employed. In order to

[73] In Italian, "*Amministrazione pubblica e difesa, assicurazione sociale obbligatoria*".
[74] This variable assumes value 1 if the answer is "Yes" and 2 if it is "No".

do so, I divide employees (those with variable *Q*=1. *Q* is in *CARCOMXX* and represents employment status (value 1 for employees, 2 for self-employed)) into "non-temp" employees (in *ALLB1, CONTRATT*=1 *lavoratore dipendente a tempo indeterminato*) and "temp" workers (*CONTRATT=2, lavoratore a tempo determinato*, *CONTRATT=3*, *lavoratore interinale*). I denote this variable as *QCONTRATT,* which assumes value 1 for "non-temp" employees, 2 for "temp" employees, 3 for self-employed workers.

The other variables constructed are:

- ***YEAR***: this variable assumes value 1 for observations in 2010 and 0 in 2008 [or 1 for 2008 and 0 for 2006]

- ***s***: the saving rate (*S/Y*)

- ***NCOMPEQ***: a measure of the dimension of the household used to construct the equivalent income. It is calculated through the modified OECD equivalence scale: the householder has a coefficient equal to 1, any other individual older than 14 0.5 and younger than 14 0.3. For each household the number of "equivalent adults" is calculated, summing up these coefficients, and household income is finally divided by this sum. The rationale is that two persons living together incur less expenses than two persons living separately, ceteris paribus. For further details, see Bartiloro and Rampazzi (2013)

- ***YEQ***: equivalent household income, defined as *Y/NCOMPEQ*. For further details, see Bartiloro and Rampazzi (2013)

Then, only active workers are of interest for our study: retirees are supposed to have a different saving pattern, unemployed have suffered too strong a shock in income to try to explain their behavior in terms of precautionary savings and, if any, they should draw from that buffer instead of increasing it. For this reason, all individuals for which the variable APQUAL assumes values 11 to 19 (those who are not employed, for any reason) at least in one of the two years are excluded.

Finally, it has been checked that the individuals have remained in the same group (treatment or control) for both periods: for the analysis of civil servants and private workers, the value of *PUBLIC* should be the same in the two periods, and similarly for the value of *QCONTRATT* for the comparison of self-employed against employees.

Individuals with extreme values of the dependent variable *s* in at least one of the two years have been excluded from the panel: only values $-1 \leq s \leq 1$ are accepted. These outliers in some cases strongly influenced regression results.

A final remark should be made about the construction of the variable PUBLIC in 2006: unfortunately, for that survey the variable PUBBLICO is not present. Anyway, until that year there was one question which asks for the number of employees in the firm (*DIMAZ*); one of the options was "not applicable because civil servant" (*DIMAZ*=7). I considered those with *DIMAZ*=7 as civil servants. Actually, I pointed out that the structure of this question was changed from year 2008, when it was split into two components: the first question asks if one was civil servant (*PUBBLICO*) and, in case of negative answer, the second one (*DIMAZ*) is submitted.

The only difference that must be pointed out is that the 2006 definition of civil servant is not so strict, not excluding explicitly employees of companies like *Trenitalia* or *Poste Italiane*; anyway, if they are registered as civil servants in 2006 but not in 2008, they are excluded from the regression because they do not remain in the same group for both periods. This discontinuity in the definition is one of the main motives leading me to prefer the diff-in-diff model that compares the employees against the self-employed.

APPENDIX II: DIFF-IN-DIFF MODEL WITH CIVIL SERVANTS AS TREATMENT GROUP

Table 42 Saving regression, baseline results. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.78E+00	1.10E-01	-25.186	2.00E-16	***
D(PUBLIC)	-4.50E-03	1.59E-02	-0.284	0.7765	
YEAR	-4.57E-03	9.49E-03	-0.481	0.63044	
D(temporary worker)	-3.99E-02	2.04E-02	-1.956	0.05055	.
D(Self-employed)	-1.81E-02	1.08E-02	-1.674	0.09432	.
log(Y)	2.93E-01	1.14E-02	25.734	2.00E-16	***
NCOMP	-4.31E-02	4.24E-03	-10.184	2.00E-16	***
W	-2.75E-08	8.26E-09	-3.336	0.00086	***
NPERC	4.27E-02	7.39E-03	5.776	8.49E-09	***
D(AGE 35-44)	1.99E-02	2.41E-02	0.824	0.41	
D(AGE 45-54)	1.04E-02	2.39E-02	0.435	0.66346	
D(AGE 55-64)	-7.91E-03	2.41E-02	-0.328	0.74306	
D(AGE >65)	3.15E-02	5.35E-02	0.588	0.55633	
D(North-East)	2.91E-02	1.16E-02	2.509	0.01215	*
D(Center)	1.84E-02	1.29E-02	1.431	0.15261	
D(South)	7.26E-02	1.39E-02	5.235	1.78E-07	***
D(Islands)	6.44E-02	1.57E-02	4.088	4.47E-05	***
D(Female)	4.93E-03	1.03E-02	0.481	0.63091	
D(High School)	-4.84E-02	9.78E-03	-4.946	8.03E-07	***
D(Bachelor, Master or PhD)	-7.38E-02	1.55E-02	-4.758	2.05E-06	***
YEAR:D(PUBLIC)	-9.63E-03	2.10E-02	-0.46	0.64572	

Residual standard error: 0.2227 on 2771 degrees of freedom
Multiple R-squared: 0.3203, Adjusted R-squared: 0.3153 F-statistic: 65.28 on 20 and 2771 DF, p-value: < 2.2e-16

Table 43: Saving regression, only households reporting equivalent household income higher than the median. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.03E+00	1.83E-01	-11.083	2.00E-16	***
D(PUBLIC)	2.89E-02	1.95E-02	1.488	0.1371	
YEAR	1.70E-02	1.24E-02	1.372	0.17038	
D(temporary worker)	1.42E-02	3.70E-02	0.383	0.70191	
D(Self-employed)	4.72E-02	1.36E-02	3.461	0.00055	***
log(Y)	2.18E-01	1.82E-02	11.95	2.00E-16	***
NCOMP	-4.79E-02	6.52E-03	-7.355	3.28E-13	***
W	-1.29E-08	8.15E-09	-1.585	0.11325	
NPERC	6.33E-02	9.93E-03	6.376	2.48E-10	***
D(AGE 35-44)	2.08E-02	3.22E-02	0.645	0.51902	
D(AGE 45-54)	6.00E-03	3.19E-02	0.188	0.85056	
D(AGE 55-64)	-4.20E-02	3.18E-02	-1.321	0.18664	
D(AGE >65)	-1.57E-02	6.02E-02	-0.261	0.79399	
D(North-East)	3.67E-02	1.36E-02	2.697	0.00709	**
D(Center)	1.84E-02	1.52E-02	1.211	0.22627	
D(South)	3.01E-02	2.08E-02	1.449	0.14748	
D(Islands)	4.21E-02	2.37E-02	1.778	0.07563	.
D(Female)	1.26E-02	1.27E-02	0.989	0.32295	
D(High School)	-4.67E-02	1.39E-02	-3.356	0.00081	***
D(Bachelor, Master or PhD)	-6.71E-02	1.78E-02	-3.762	0.00018	***
YEAR:D(PUBLIC)	-1.82E-02	2.50E-02	-0.729	0.46607	

Residual standard error: 0.1997 on 1375 degrees of freedom
Multiple R-squared: 0.1967, Adjusted R-squared: 0.185 F-statistic: 16.84 on 20 and 1375 DF, p-value: < 2.2e-16

Table 44 Saving regression, only households reporting equivalent household income lower than the median. Years 2006-2008. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.43E+00	2.27E-01	-15.159	2.00E-16	***
D(PUBLIC)	-4.47E-02	2.52E-02	-1.772	0.07657	.
YEAR	-2.54E-02	1.41E-02	-1.811	0.07038	.
D(temporary worker)	-3.97E-02	2.61E-02	-1.52	0.12883	
D(Self-employed)	-7.09E-02	1.73E-02	-4.107	4.24E-05	***
log(Y)	3.60E-01	2.41E-02	14.93	2.00E-16	***
NCOMP	-4.27E-02	7.01E-03	-6.087	1.49E-09	***
W	-7.60E-08	2.60E-08	-2.925	0.0035	**
NPERC	2.84E-02	1.15E-02	2.467	0.01376	*
D(AGE 35-44)	2.75E-02	3.52E-02	0.779	0.43592	
D(AGE 45-54)	2.29E-02	3.50E-02	0.654	0.51304	
D(AGE 55-64)	3.84E-02	3.57E-02	1.074	0.28316	
D(AGE >65)	1.23E-01	1.04E-01	1.19	0.23431	
D(North-East)	2.55E-02	1.94E-02	1.318	0.18788	
D(Center)	1.97E-02	2.15E-02	0.914	0.36083	
D(South)	9.84E-02	1.98E-02	4.963	7.80E-07	***
D(Islands)	8.50E-02	2.21E-02	3.853	0.00012	***
D(Female)	-6.34E-03	1.63E-02	-0.39	0.69678	
D(High School)	-4.67E-02	1.40E-02	-3.341	0.00086	***
D(Bachelor, Master or PhD)	-4.82E-02	3.21E-02	-1.5	0.13372	
YEAR:D(PUBLIC)	-1.15E-03	3.48E-02	-0.033	0.97367	

Residual standard error: 0.2383 on 1375 degrees of freedom

Multiple R-squared: 0.2465, Adjusted R-squared: 0.2356 F-statistic: 22.5 on 20 and 1375 DF, p-value: < 2.2e-16

Table 45: Saving regression, baseline results. Years 2008-2010. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.77E+00	9.99E-02	-27.728	2.00E-16	***
D(PUBLIC)	2.53E-03	1.37E-02	0.185	0.85328	
YEAR	-1.90E-02	8.95E-03	-2.124	0.03376	*
D(temporary worker)	-3.95E-02	1.68E-02	-2.356	0.01852	*
D(Self-employed)	-2.88E-03	1.02E-02	-0.282	0.77771	
log(Y)	2.93E-01	1.03E-02	28.392	2.00E-16	***
NCOMP	-3.81E-02	3.89E-03	-9.786	2.00E-16	***
W	-2.63E-08	6.23E-09	-4.23	2.41E-05	***
NPERC	2.59E-02	6.99E-03	3.709	0.00021	***
D(AGE 35-44)	-1.02E-02	2.19E-02	-0.466	0.64139	
D(AGE 45-54)	-3.80E-03	2.15E-02	-0.177	0.85931	
D(AGE 55-64)	-1.25E-02	2.17E-02	-0.578	0.56357	
D(AGE >65)	3.47E-02	3.65E-02	0.951	0.34158	
D(North-East)	4.52E-02	1.07E-02	4.219	2.53E-05	***
D(Center)	9.53E-03	1.22E-02	0.781	0.43488	
D(South)	7.67E-02	1.30E-02	5.915	3.69E-09	***
D(Islands)	5.72E-02	1.46E-02	3.929	8.72E-05	***
D(Female)	1.63E-02	8.93E-03	1.82	0.06888	.
D(High School)	-4.48E-02	9.18E-03	-4.884	1.10E-06	***
D(Bachelor, Master or PhD)	-6.70E-02	1.33E-02	-5.031	5.17E-07	***
YEAR:D(PUBLIC)	5.79E-03	1.80E-02	0.321	0.74798	

Residual standard error: 0.2155 on 3083 degrees of freedom

Multiple R-squared: 0.3304, Adjusted R-squared: 0.326 F-statistic: 76.06 on 20 and 3083 DF, p-value: < 2.2e-16

Table 46: Saving regression, only households reporting equivalent household income higher than the median. Years 2008-2010. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.95E+00	1.73E-01	-11.246	2.00E-16	***
D(PUBLIC)	3.69E-03	1.77E-02	0.208	0.83537	
YEAR	-4.70E-02	1.20E-02	-3.935	8.70E-05	***
D(temporary worker)	2.10E-02	2.97E-02	0.709	0.47813	
D(Self-employed)	3.63E-02	1.30E-02	2.8	0.00517	**
log(Y)	2.15E-01	1.72E-02	12.501	2.00E-16	***
NCOMP	-3.94E-02	6.39E-03	-6.164	9.03E-10	***
W	-1.52E-08	5.93E-09	-2.567	0.01036	*
NPERC	3.76E-02	9.79E-03	3.839	0.00013	***
D(AGE 35-44)	-6.95E-04	3.20E-02	-0.022	0.98268	
D(AGE 45-54)	-1.70E-02	3.09E-02	-0.549	0.58331	
D(AGE 55-64)	-3.10E-02	3.12E-02	-0.993	0.32082	
D(AGE >65)	1.02E-02	4.51E-02	0.226	0.82151	
D(North-East)	5.53E-02	1.28E-02	4.315	1.70E-05	***
D(Center)	5.05E-03	1.46E-02	0.346	0.72947	
D(South)	3.75E-02	1.91E-02	1.964	0.04966	*
D(Islands)	3.26E-02	2.20E-02	1.478	0.1395	
D(Female)	1.84E-02	1.13E-02	1.625	0.10442	
D(High School)	-4.45E-02	1.35E-02	-3.295	0.00101	**
D(Bachelor, Master or PhD)	-6.14E-02	1.70E-02	-3.611	0.00032	***
YEAR:D(PUBLIC)	3.66E-02	2.25E-02	1.627	0.10396	

Residual standard error: 0.1984 on 1531 degrees of freedom
Multiple R-squared: 0.1747, Adjusted R-squared: 0.1639 F-statistic: 16.2 on 20 and 1531 DF, p-value: < 2.2e-16

Table 47: Saving regression, only households reporting equivalent household income lower than the median. Years 2008-2010. Dependent variable: reported saving rate

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.66E+00	2.09E-01	-17.515	2.00E-16	***
D(PUBLIC)	-1.98E-02	2.08E-02	-0.955	0.33997	
YEAR	4.00E-03	1.30E-02	0.308	0.75825	
D(temporary worker)	-4.04E-02	2.14E-02	-1.884	0.05971	.
D(Self-employed)	-3.13E-02	1.67E-02	-1.874	0.06106	.
log(Y)	3.83E-01	2.22E-02	17.274	2.00E-16	***
NCOMP	-4.14E-02	6.38E-03	-6.479	1.24E-10	***
W	-1.13E-07	4.16E-08	-2.719	0.00662	**
NPERC	1.71E-02	1.06E-02	1.618	0.10591	
D(AGE 35-44)	-1.45E-02	2.99E-02	-0.485	0.62758	
D(AGE 45-54)	5.59E-03	2.96E-02	0.189	0.85016	
D(AGE 55-64)	1.01E-02	3.01E-02	0.334	0.73816	
D(AGE >65)	1.03E-01	6.40E-02	1.616	0.10626	
D(North-East)	3.57E-02	1.78E-02	2.01	0.04461	*
D(Center)	1.80E-02	2.04E-02	0.881	0.37856	
D(South)	1.04E-01	1.88E-02	5.516	4.06E-08	***
D(Islands)	8.19E-02	2.04E-02	4.012	6.31E-05	***
D(Female)	4.47E-03	1.38E-02	0.325	0.74531	
D(High School)	-4.76E-02	1.26E-02	-3.772	0.00017	***
D(Bachelor, Master or PhD)	-3.91E-02	2.30E-02	-1.696	0.09008	.
YEAR:D(PUBLIC)	-2.31E-02	2.84E-02	-0.815	0.41493	

Residual standard error: 0.2259 on 1531 degrees of freedom
Multiple R-squared: 0.2799, Adjusted R-squared: 0.2705 F-statistic: 29.76 on 20 and 1531 DF, p-value: < 2.2e-16

APPENDIX III CALIBRATION WITH α, λ, β AND γ HETEROGENOUS

In this section, I present the results of the model calibrated without the homogeneity assumption for β and γ. I looked for the values which maximize the objective function, by using an iterative procedure. I start with a four-dimensional array, where for each dimension there were 11 possible parameters values (with the exception of γ ∈ $\{0.008,0.009,0.010,0.011,0.012\}$): for α and λ the possible values considered are between 0 and 0.5, by 0.05, while for β between 0.5 and 1, always by 0.05.

In Table 48, I report the values of the parameters which maximize the correlation coefficient between the real and the simulated index. In Table 49, I report the values of the parameters which maximize the criterion function $\theta(\alpha,\lambda) = r_{s,r}(\alpha,\lambda) - 0.2 * |s_r - s_s(\alpha,\lambda)|$. In Table 50 I report the values of the parameters which maximize the modified criterion function $\theta'(\alpha,\lambda) = r_{s,r}(\alpha,\lambda) - 0.2 * |(1-x) * s_r - s_s(\alpha,\lambda)|$.

Table 48: calibration of the model parameters such as to maximize the correlation coefficient, by education

	α	λ	β	γ
Less than high school	0.10	0.00	0.70	0.012
High school	0.15	0.25	0.85	0.008
Bachelor, Master or PhD	0.15	0.05	0.75	0.011

Table 49: calibration of the model parameters such as to maximize $\boldsymbol{\theta(\alpha,\lambda)}$, by education

	α	λ	β	γ
Less than high school	0.15	0.45	0.85	0.009
High school	0.20	0.45	0.95	0.008
Bachelor, Master or PhD	0.25	0.30	0.85	0.009

Table 50: calibration of the model parameters such as to maximize $\boldsymbol{\theta'(\alpha,\lambda)}$, by education

	α	λ	x	β	γ
Less than high school	0.25	0.35	0.287	0.90	0.008
High school	0.25	0.40	0.247	0.80	0.009
Bachelor, Master or PhD	0.35	0.45	0.227	0.65	0.009

In all the three circumstances, it is difficult to capture any pattern but in α, which appears to be always increasing in education. λ is non-linear in Table 48, decreasing in education in Table 49 and increasing in Table 50. β and γ are nonlinear, but in Table 50 where the first is decreasing and the second increasing in education, and, as a further consideration, the values assumed by β are quite high, in particular in Table 49.

As far as the values of the "trend parameter" β are concerned, it must be pointed out that the expectations indexes are highly autocorrelated: 0.86 for "Less than high school", 0.89 for "High school", 0.89 for "Bachelor, Master or PhD". This, in most of the cases reported in Table 48, Table 49 and Table 50, is probably reflected in the high values of the trend parameter, since a way to obtain a simulated index with

an high autocorrelation coefficient is to have the majority of agents keeping an opinion very similar to the year before (high β) and the rest of the agents that update their expectation according to the last forecast, possibly correcting the direction of the movement of the index.

Anyway, as already highlighted, high values of β carry the risk of having a given proportion of agents with meaningless expectations. This could be solved adopting richer models which allow, for example, for non-constant λ and β but that makes them endogenous as function of $\Delta u^{e}_{i,t-1,t}$, or for a non-linear function for the expectation development.

Having constructed a model with relatively simple (linear) rules in order to ease the interpretation of the results, I decided to run the simulation with α and λ heterogeneous but β and γ homogeneous assuming the lowest possible value of β found in Table 48, Table 49 and Table 50 (0.65): this to avoid meaningless expectations. In further extensions of this model that solve in another way the problem, the assumption could be relaxed.

For γ, I assumed a value of 0.01 for several reasons. First of all, 0.01 is a value between the minimum and the maximum present in Table 48, 49 and 50. Secondly, this constitutes an easy "rule of thumb" for individuals: like a good priced €19.99 seems much cheaper than a good priced €20.00, similarly a percentage point change seems much more than a 0.9 percentage points change. Finally, the average annual unemployment rate absolute change in the period studied has been equal to 0.59 percentage points. For these reasons, assuming simmetry, 0.5 percentage points seems a sensible threshold between the answers "it will remain stable" and "it will go up a little", and 1 percentage points seems a sensible threshold between the answers "it will go up a little" and "it will go up a lot".

APPENDIX IV: R CODE

In this appendix I report an example of the code used to simulate the index (Less than High School). For data inputs, I will use the names employed in my dataset, explain what kind of data I am referring to.

Similarly, simulations for other educational groups could be performed by substituting "`exp$LTHS`" (the time series of ISTAT households unemployment expectations index for "Less than High School" educational level) with "`exp$HS`" (the time series of ISTAT households unemployment expectations index for "High School" educational level) for High School or "`exp$BMP`" (the time series of ISTAT households unemployment expectations index for "Bachelor, Master or PhD" educational level) for Bachelor, Master or PhD, and eventually changing the parameters that control the simulation (`alpha1` for α, `lambda1` for λ, `beta1` for β, `gamma1` for γ). exp$OECD is the time series of OECD professional forecasts, defined on a range from 0 to 100 (and not 0 to 1)

```
exp<-read.csv("ocse_istat.csv", header=T, sep=";")

#I DEFINE THE VALUES FOR THE FOUR PARAMETERS
alpha1<-0.21
lambda1<-0.02
beta1<-0.65
gamma1<-0.01

#I CONSTRUCT THE TIME SERIES OF "PROFESSIONAL FORECASTS"
disoecd<-exp$OECD/100

#THE PROBABILITY OF BEING INFECTED BY THE VIRUS FOR THE EDUCATIONAL GROUP IS EQUAL TO ALPHA
pinfection1<-alpha1

#I ASSUME THAT ALL AGENTS START IN PERIOD 0 FROM THE SAME EXPECTATION (0.105)
tmat<-matrix(0.105, nrow=1440, ncol=1)
tdis<-rep(0.105, 1440)

r<-rep(0, 1440)
```

```
#FOR EACH PERIOD, FOR EACH INDIVIDUAL, I SIMULATE THE EXPECTATION
for (semester in 1:56) {
  for (agent in 1:1440) {
r[agent]<-runif(1)
    if (r[agent]<=pinfection1)
tdis[agent]<-lambda1*tdis[agent]+(1-lambda1)*disoecd[semester]
else tdis[agent]<-max(0,min(1,tdis[agent]+beta1*(tmat[agent, semester]-tmat[agent, max(1,semester-1)])))}
tmat<-cbind(tmat,tdis)
}

#FOR EACH PERIOD, GIVEN THE SIMULATION, I CALCULATE THE AVERAGE EXPECTATION
averageexp<-rep(0,56)
for (semester in 1:56) {
averageexp[semester]<-mean(tmat[,semester+1])
}

#FOR EACH PERIOD, FOR EACH INDIVIDUAL, I CALCULATE THE CHANGE IN EXPECTATION
tdiff<-matrix(0, nrow=1440, ncol=57)

for (semester in 2:57) {
  for (agent in 1:1440) {
tdiff[agent, semester]<-(tmat[agent, semester]-tmat[agent, semester-1])
}
}

#I VERIFIY IF, IN EACH PERIOD, A GIVEN AGENT WOULD CHOOSE THE MODALITY "IT WILL GO UP A LOT"
upalot<-matrix(0, nrow=1440, ncol=57)

for (semester in 2:57) {
  for (agent in 1:1440) {
if (tdiff[agent, semester]>gamma1)
upalot[agent, semester]<-1
else upalot[agent, semester]<-0
}
}

#I VERIFIY IF, IN EACH PERIOD, A GIVEN AGENT WOULD CHOOSE THE MODALITY "IT WILL GO UP A LITTLE"
up<-matrix(0, nrow=1440, ncol=57)
```

```
for (semester in 2:57) {
  for (agent in 1:1440) {
if (tdiff[agent, semester]>=(gamma1/2)&tdiff[agent, semester]<=gamma1)
up[agent, semester]<-1
else up[agent, semester]<-0
}
}

#I VERIFIY IF, IN EACH PERIOD, A GIVEN AGENT WOULD CHOOSE THE MODALITY "IT WILL GO DOWN A LITTLE"
down<-matrix(0, nrow=1440, ncol=57)

for (semester in 2:57) {
  for (agent in 1:1440) {
if (tdiff[agent, semester]<=(-gamma1/2)&tdiff[agent, semester]>=(-gamma1))
down[agent, semester]<-1
else down[agent, semester]<-0
}
}

#I VERIFIY IF, IN EACH PERIOD, A GIVEN AGENT WOULD CHOOSE THE MODALITY "IT WILL GO DOWN A LOT"
downalot<-matrix(0, nrow=1440, ncol=57)

for (semester in 2:57) {
  for (agent in 1:1440) {
if (tdiff[agent, semester]<(-gamma1))
downalot[agent, semester]<-1
else downalot[agent, semester]<-0
}
}

#GIVEN THE MODALITIES, I SIMULATE THE INDEX
index1<-rep(0,56)

for (semester in 1:56) {
index1[semester]<-(sum(upalot[,semester+1])*2+sum(up[,semester+1])-
sum(down[,semester+1])+sum(downalot[,semester+1])*(-2))
}
```

```
index1<-index1*100/1440

#I PRESENT SOME PLOTS
par(mfrow=c(2,2))
plot(index1[19:56], type="o", labels=F, tick=F, main="SIMULATED INDEX (Less than High School)",
xlab="Years", ylab="Households unemployment expectations index")
axis(side=1, at=c(1,11,21,31), labels=c("1995 S1", "2000 S1", "2005 S1", "2010 S1"))
axis(side=2, labels=TRUE)
plot(exp$LTHS[19:56], type="o", labels=F, tick=F, main="REAL INDEX (Less than High School)",
ylab="Households unemployment expectations index", xlab="Years")
axis(side=1, at=c(1,11,21,31), labels=c("1995 S1", "2000 S1", "2005 S1", "2010 S1"))
axis(side=2, labels=TRUE)
plot(averageexp[19:56], type="o", labels=F, tick=F, main="HOUSEHOLDS AVERAGE EXPECTATIONS (Less than High
School)", ylab="Households average unemployment rate expectation", xlab="Years")
axis(side=1, at=c(1,11,21,31), labels=c("1995 S1", "2000 S1", "2005 S1", "2010 S1"))
axis(side=2, labels=TRUE)
plot(disoecd[19:56], type="o", labels=F, main="PROFESSIONAL FORECASTS", ylab="OECD unemployment rate
forecasts", tick=F, xlab="Years")
axis(side=1, at=c(1,11,21,31), labels=c("1995 S1", "2000 S1", "2005 S1", "2010 S1"))
axis(side=2, labels=TRUE)
```