

PhD in Economics – I cycle

On the Effects of the Introduction of the Deficit Reduction Plans  
on the Dynamic of the Public Health Expenditure in Italy

Heterogeneous Expectations and Stock Market Participation

Candidate:

Hermes Morgavi

Tutor:

Pierpaolo Benigno



# On the Effects of the Introduction of the Deficit Reduction Plans on the Dynamic of the Public Health Expenditure in Italy

Hermes Morgavi

## Abstract

The Italian health system was modified in 2001 from a centralized system to a decentralized one. This reform helped to reduce the dynamic of the health expenditure in the subsequent period. Nevertheless, this reduction was not uniform in all the Regions. Some Regions were able to better manage the increased autonomy to reorganize more efficiently the Regional health system (which already was administered regionally) and were able to reduce the dynamic of public health expenditure while some others were less able to achieve the same reduction.

The initial approach of bailing out the subsequent deficit generated by these Regions generated in the Regions an expectation of bailing out. This softened the Regional budget constraint and every two years was necessary again to bail out the debts of the Regions.

The introduction of the deficit reduction plans was an important instrument of governance that significantly reduced the expectation of bailing out. This paper analyzes the driving factors of public health expenditure in Italy and quantifies the effect of each factor and particularly the reduction effect of the deficit reduction plans on the dynamic of the Regional public health expenditure.

## Section 1. Introduction

The last reform of the Italian National health service (Servizio Sanitario Nazionale, SSN) decentralized the organization and the administration of the SSN. The Central Government defines the general framework, determines the Essential Assistance Level (Livelli Essenziali di Assistenza, LEA) and guarantees the necessary financial resources consistently with National public finance objectives and monitors the fulfilment of the LEA and of the financial objectives by the Regions. On the other hand, Regions internally organize the health care system and guarantee the respect of the LEA in their territories. Every Region has absolute freedom to best organize its own Regional Health System (Servizio Sanitario Regionale, SSR) in order to fulfil the LEA in the most efficient and effective way, consistently with the financial resources provided by Central Government and its own Regional resources.

**The Republic safeguards health as a fundamental right of the individual and as a collective interest, and guarantees free medical care to the indigent. (art.32 Cost.)**



### Central Government

- defines the Essential Assistance Level (LEA)
- guarantees the necessary financial resources to guarantee the LEA with an efficient and appropriate use of the resources and compatibly with the public finance objectives



### Regions

- organize the Regional health system (RHS)
- guarantee the supply of health service in their territory

Figure 1 – Scheme of the allocation of the competences of Central Government and Regional Governments on health

The idea behind this system is to acknowledge the historical differences between the Regions and differentiate the health care organization on the territory giving to the Regions the responsibility to best organize assistance on their territory. The main idea was that the Regions have better knowledge of local problems than the Central Government has and that levels of Government closer to citizenship can be better controlled by the body of the citizens itself.

As pointed out by (Ambrosiano & Bordignon, 2007), several European countries are involved in the same process of decentralization and this can indeed jeopardize the stability of the Euro-area. In fact, decentralization requires great attention in setting up effective rules for local finance in order not to fall into a Soft Budget Constraint problem (SBC). The organization of the SSN after the reform of 2001 was a clear example of SBC. (Bordignon & Turati, 2003), in fact, underlined that the bailing-out expectation during the years subsequent the 2001 reform generated a SBC problem and, consequently, persistent deficits of the SSN. These deficits were concentrated for the great part in four Regions in which expenditures systematically overturned incomes. To manage this problem the Deficit Reduction Plan (DRP) were introduced in 2006. This instrument allows differentiating the intervention of the Central Government, beside the Regions involved, on regional public health expenditure forcing the so-called “non-virtuous Regions” to reduce the dynamic of their public health expenditure without affecting the politics of the so-called “virtuous Regions”.

The discriminant between virtuous and non-virtuous Regions can be given by an efficient organization of the SSR that controls the dynamic of the public expenditure coherently with the dynamic of the incomes. The alternative explanation can be given by an incorrect allocation of the financial contribution from the Central Government.

The data on the SSRs for the years 2002-2010 allow studying the driving factors of the dynamic of the regional public health expenditures and the effects of the introduction of the deficit reduction plans. The aim of this paper is to identify these factors and quantify their impact on the dynamic

of the public health expenditure and in particular the effect of the introduction of the deficit reduction plans.

The paper is organized as follows. In **Section 2**, following (Bodenheimer, 2005), I will introduce the possible factors driving the Regional health expenditure in Italy. In **Section 3** I will analyze the data I used to identify which of the previous factors actually drive the dynamic of the health expenditure in Italy. In **Section 4** I will analyze which of the factors are significant to explain the dynamic of public health expenditure in Italy and quantify the effects of each of the significant driving factor. Finally, in **Section 5** I will comment the results.

## **Section 2.**

The factors driving public health expenditure dynamics can be grouped into three macro areas:

1. **External factors.** These do not pertain to the organization of the health system but are due to causes that are outside the scope of the SSN. The demand for health may be driven, among the others, by the increase of life expectancy and by the economic growth. These causes cannot be controlled by the administration of the health system, which must operate adequately taking into account of these factors as exogenous.
2. **Internal organization of the SSR.** The inefficiencies of the administration can be an important driving factor. Regions better organized can more effectively contain the costs of the SSR, while worse organized Regions may find it hard to control the dynamic of the Regional health expenditure. The Regional Government tries to force the bureaucrats to reduce the internal inefficiencies in order to reduce the financial resources devolved to the SSR without reducing the health services offered. On the contrary, bureaucrats maximize their individual utility maximizing their budget.

3. **Relations between SSR and providers.** These factors attain to the principal-agent problem between the administration of the Regional health system and the providers of the health services. The SSR tries to reduce the expenditure in health services, while the providers try to maximize their profit, for private providers, or their budget, for public providers.
  
4. **Relation between Central Government and Regional Governments.** Another principal-agent problem is given by the relation between Central Government and Regional Government. The Central Government defines the LEA and contributes with enough financial resources to guarantee in their territory the respect of the LEA. The Regions tend to ask more and more resources in order not to reduce their internal inefficiencies and expand their budget. On the contrary the Central Government tends to force Regions to reduce as much as possible the inefficiencies of the SSR in order to reduce the financial resources devolved to the Regions without reducing the health services offered.

### **External factors**

Public health expenditure can be driven by the dynamic of the economy. The increased wellness may induce greater interest in health and therefore the citizenship may prefer to expand health expenditure. This is consistent with the results of (Hall & Jones, 2007) who found out that the growth of health spending is a rational response to the growth of income. The higher is the income per capita, in fact, the higher is the expected utility of an additional year of life and therefore the more each individual is willing to spend in health services to extend life expectancy. Hence, two possible income factors can drive the regional health expenditure: the national income and the regional one. The growth of the national income can induce a general growth of the public health expenditure as it can raise the standards of health required by the citizenship. This factor affects in the same way all the Regions and rises the health expenditure in all the Regions. Moreover, each Region can decide to offer additional health services with respect of the LEA at the expenses of the

Regional balances. Therefore if the economic growth were one of the driving factors of the public health expenditure, it would be even more relevant to study the impact of the increase in the income per capita in each Region on the regional public health expenditure.

The second possible external factor is given by the age profile of the population. Some studies pointed-out that people of seventy-five years or older tend to spend around five times more than people of age between twenty-five and thirty-four in health services<sup>1</sup>. On the other hand some other studies found out that the increase in the number of individuals older than seventy-five years can explain only three or four per cent of the variability of the health expenditure. It must be noticed, in fact, that the number of people older than seventy-five years does not vary much over the time<sup>2</sup>.

Technologic innovation is considered by many authors an important factor driving health expenditure<sup>3</sup>. Many authors showed that, counterintuitively, technological innovation has increasing effects on health expenditure. According to these authors, in fact, technological innovation tends to lower prices but increase expenditure. The impact of innovation may not be the same among Regions. The new technologies, in fact, may have different diffusion among Regions due to differences of quality of the specialists, of the hospitals and in the investment in innovation made by the Regions.

Other relevant aspects are linked to social and geographical characteristics of the Regions. Social factors as instruction, different attention to food quality, importance given to the sport may be, in fact, relevant on the health of the population and in turn on health expenditure. The geography of the territory is important for the organization of the SSR. For example, mountainous areas may have more transportation problems than flat areas.

---

<sup>1</sup> See for example (Reinhardt, 2003)

<sup>2</sup> See, for example, (Aaron, 1991), (Newhouse, 1993), (Reinhardt, 2003)

<sup>3</sup> See, for example, (Gelijns & Rosenberg, 1994), (Chernew, Hirth, S., Ermann, & Fendrick, 1998), (Fuchs & Sox, 2001)

## **Internal organization of the SSR**

Administrative costs are an important component of the non-market costs of the health sector. They are direct costs for the Regions and may in turn induce additional costs due to the inefficiencies of the bureaucracy. As pointed out by several authors<sup>4</sup>, bureaucracy may tend to expand as much as it can to increase its power. On the other hand, the administrative personnel may be important in reducing the inefficiencies of the health system and therefore reduce the public health expenditure.

Copayments are not only a way to increase the incomes of the SSR. In fact, the more patients pay out-of-pocket for the services they have access to, the more they are aware of the costs of the health system and responsible of the increased cost due to inappropriateness or to moral hazard. Because of the public accounting rules, out-of-pocket payments represent an immediate reduction of the public health expenditure and, at the same time, an instrument to awaken the citizenship on the importance of reducing the inefficiencies of the SSR and the inappropriate access to health service. On the other hand, the presence of copayment may induce citizens to postpone their accesses to an appropriate fruition of health services. This, in some cases, may have the effect of worsening some minor health problems, which could be easily cured, making them become more serious. This mechanism can, at the end of the story, increase the final cost. Moreover, if copayment are too high, some citizen may choose to pay directly out of pocket in the private sector.

The Regional political cycle can have either a positive or a negative effect on the Regional public health expenditure. The Regions, in fact, may have a positive incentive to increase public health expenditure, as it is a social relevant argument and a possibility of patronages. On the other hand, Regional governments may also be interested to show to the citizenship positive results in

---

<sup>4</sup> See, for example, (Downs, 1967), (Garston, 1993), (Niskanen & Niskanen, 2007)

containing the public health expenditure to be certified virtuous by the committee for the respect of the commitments<sup>5</sup>.

### **Relations between SSR and private sector providers.**

Another potential important aspect of the dynamic of the expenditure is given by the presence of competitive market. Some scholars support the idea that the creation of a competitive market in the health sector can reduce the price of health services through competition between the health care providers<sup>6</sup>. Nevertheless, it must be noticed that if on one hand competition has in general a positive effect on the level of prices, on the other hand the health sector is characterized by indirect demand and asymmetry of information and hence it is not said that competition may have positive effect in the overall health expenditure. Increased competition may, on the contrary, raise the number of health services sold as a consequence of increased induced demand.

Increases in public health expenditure may be driven also by providers' market power. The higher is the power of one provider the more bargaining power it has in defining its budget. Providers' market power can create a mechanism of expansion of the share of health services bought from private providers. On the other hand, it is possible that the private sector may be more efficient than the public one. In this case, a higher presence of the private sector can reduce the Regional public health expenditure.

### **Relation between Central Government and Regional Governments**

Wages and salaries in the public health sector are defined through a collective contract procedure in which Region have only a marginal role. This procedure involves several actors and takes a long period to be completed (see Table 2 for a comparison of the period of validity of the collective

---

<sup>5</sup> See (Soroka & Wlezien, 2005) for an analysis of how public opinion may affect political decision on public expenditure

<sup>6</sup> See, for example, (Enthoven, 2003), (Liebowitz, 1994), (Keeler, Melnick, & Zwanziger, 1999), (Robinson, 2004)

contracts with the year in which these were signed), usually after one or two years after the end period of validity of the collective contract<sup>7</sup>. The public accounting rules do not allow the public firms to book the correspondent additional costs introduced by the new collective contracts until the end of the collective contract procedure. Only starting from 2007 it have been imposed to the health firms to earmark, in absence of the collective contract, enough funds to meet the additional costs of the new contract as if wages and salaries were to increase of the same percentage of the programmed inflation. This is coherent with the fact that the central Government contribution to the FSN is comprehensive of the cost of the new contract as if wages and salaries increased of the same percentage of the programmed inflation. If the cost of the collective contract is higher, the central Government gives the correspondent extra funds; vice versa, if the cost of the collective contract is lower, the central Government takes back the extra funds.

This implies that, until 2007, in absence of valid collective contract, the health firms did not book any cost for the renewal of the collective contract. Only in the year of conclusion of the collective contract procedure, Regions and health firms booked the current and the outstanding additional costs introduced by the collective contracts. Hence, in that year one observes an anomalous increase in public expenditure. In order to better analyze the expenditure data, I therefore corrected the original data to take in account of this shift forward of the costs of personnel and family doctors. As it can be seen in Figure 7 this leads to significant correction in the years before 2007.

The impact of these corrections are even more evident in the first differences of the logarithms (see Figure 9). Analyzing, in particular, the corrected series it seems to be present the effect of another factor that can be called the last year effect. When bargaining the increases in wage and salaries, the

---

<sup>7</sup> See the ARAN website (<http://www.aranagenzia.it/>) for more information on the public contract procedure

social representatives (ARAN<sup>8</sup>, unions, etc.) must fix the increases of wage and salaries for the first and the following years of the period. Of course, unions are more interested to the increase at the end of the period – the final increase. The quantification of the additional costs of the collective contract is made in terms of outstanding costs and final costs. The public balance rules impose the Government to budget balance for the current year and the following four, though less attention is given to the following years; this also because the following years' balances do not have any authorization value. Therefore, given the available financial resources, unions are more prone to accept small increases in the first year in exchange of higher final increases. The effect of this mechanism is that in the final year of effectiveness of the contract the Dependent personnel expenditure and the Family doctors expenditure have higher increases than in the previous years. This effect can be seen in Table 3 where the increases in wages and salaries in each year for the dependent personnel of the SSN are reported.

Finally, but most important, the expansion can be driven by the lack of controls and of cost-containing measures. As said, in the principal-agent problem between Central Government and Regional Governments, the first tries to force the latter to reduce the regional inefficiencies in order to reduce the total cost of the SSN while Regions tend to ask as many financial resources as they can from the Central Government to increase their budgets.

Every three years the Central Government, together with the Regions, fixes the total amount of the contribution of the Central Government to the FSN. This amount is then allocated to the Regions according to a set of demand of health indicators. In this bargaining model between Central Government and Regional Governments a key role is played by the asymmetry of information on the efficient cost of the SSRs. The Central Government, in fact, can only observe the realized cost

---

<sup>8</sup> The ARAN (Agenzia per la Rappresentanza Negoziabile delle Pubbliche Amministrazioni) is the agency entitled to participate with the Unions to the bargaining process as representative of the Public Administration side

of each Region, which is given by the efficient cost plus the cost of the inefficiencies. It must be notice that the efficient cost may vary for each Region because of the different composition of the populations or because of geographical, social and economical differences between Regions. Therefore the total cost per capita of the SSR of a generic Region  $r$ , at time  $t$ , is given by

$$c_{r,t} = ec_{r,t} + \zeta_{r,t} \quad (1)$$

Where  $ec_{r,t}$  is the efficient regional cost of the SSR and  $\zeta_{r,t}$  is the cost of the inefficiencies. The efficient cost can be decomposed in a common component and a regional one. That is:

$$ec_{r,t} = \alpha_{0,t} + \alpha_{r,t} \quad (2)$$

Where  $\alpha_0$  is the efficient cost in the hypothetical case of no differences between Regions and  $\alpha_r$  is an additional component which is typical of Region  $r$  and it is due to exogenous factors.

The Central Government cannot observe the efficient cost but only the total one; therefore Regions have incentives to lie on the efficient regional cost to obtain more resources from the Central Government. It may therefore happen that some Regions exceed their budget and create deficit claiming that the financial resources transferred from the Central Government are not enough to cover the efficient cost.

Once these Regions generated the deficits, the Central Government has two possible choices: the first choice is to bailout the deficits, reducing other expenditures; the second choice is not to bailout and force the Regions to cover their deficits with Regional resources. If it chooses to force the Regions to cover their deficits, these may claim not to be able to cover because they do not have enough resources and that the deficit is not due to Regional inefficiencies but to insufficient of financial resources. It must be noticed at this point that, because of the European community rules,

these extra deficits are accounted for the deficit/GDP ratio and increase the amount of National debt. Hence if these deficit are not covered the Central Government risks to receive sanctions from the European Commission while the Regions that generated the deficit can be sanctioned only by the Central Government. This leads back to the initial problem of the correct identification of the efficient cost and adequate contribution of the Central Government to the FSN.

This aspect is the key of the time inconsistency problem. In fact, the Central Government can allocate low level of resources to the Regions, consistently with a correct estimation of the efficient cost, to force them to reduce the inefficiencies costs. Nevertheless, at the end of the year it may be induced to give extra-resources under the pressure of the Regions. Resources that are taken from other public functions. As pointed out by (Bordignon & Turati, 2003), this scheme can generate a mechanism of expectation of bailing-out that may induce the Regions to expand their public health expenditure in hope of extra resources and Central Government to bail-out at least part of the deficits<sup>9</sup>.

In order to break this mechanism, in 2006 the instrument of the Deficit Reduction Plan was introduced. This instrument changes the prospective of the Regions deciding whether to contain the inefficiencies of the SSR or not.

### **The Deficit Reduction Plans**

Regions with high and significant deficit in the Regional health sector are obliged to undertake a deficit reduction plan. The definition of high deficit is given in the Pact for Health, (Patto per la salute, PpS) as

---

<sup>9</sup> In Figure 2 - Figure 4 the extensive forms of the scenarios presented by (Bordignon & Turati, 2003) are shown

those Region whose deficit is greater than 5 per cent<sup>10</sup> of the resources of the FSN allocated for the to the SSR<sup>11</sup>.

The Regions undertaking Deficit Reduction Plan must demonstrate the concrete will and administrative and management skills to undertake a reorganization plan of the Regional health system. Otherwise, the Central Government can use the “subsidiary powers<sup>12</sup>”

The deficit reduction plan must be composed of the following sections:

- **National legislation.** This first section is necessary mostly for the citizenship. It is a section which synthesizes the current National legislation on health; that is the boundary of Regional interventions;
- **Analysis of Regional context.** This section contains a demographic analysis of the population of the Region and an analysis of the demand of health services, the Regional legislation on health and the organization of the Regional health service. This was thought to be the most important section of the reduction plan as it must contain an analysis of the main components

---

<sup>10</sup> Originally the definition of Regions with high and significant deficit comprehended the Regions whose deficits were greater than 7 per cent of the Central Government Contribution to the FSR

<sup>11</sup> See L. 191/2009

<sup>12</sup> This power is defined in the art. 117 and art. 120 of the Italian Constitution which, respectively, quote:

- art 117: “The Regions and the autonomous provinces of Trent and Bolzano take part in preparatory decision-making process of EU legislative acts in the areas that fall within their responsibilities. They are also responsible for the implementation of international agreements and EU measures, subject to the rules set out in State law which regulate the exercise of subsidiary powers by the State in the case of non-performance by the Regions and autonomous Provinces”;
- art 120: “The Government can act for bodies of the regions, metropolitan cities, provinces and municipalities if the latter fail to comply with international rules and treaties or EU legislation, or in the case of grave danger for public safety and security, or whenever such action is necessary to preserve legal or economic unity and in particular to guarantee the basic level of benefits relating to civil and social entitlements, regardless of the geographic borders of local authorities. The law shall lay down the procedures to ensure that subsidiary powers are exercised in compliance with the principles of subsidiarity and loyal co-operation”

driving the regional public health expenditure. This section is crucial to underline the inefficiencies of the system and the contradiction of the Regional legislation and of the administrative organization;

- **Tendential forecast for the following three years.** In this section the Region must provide a forecast of the expenditure aggregates<sup>13</sup> and of the revenues aggregates<sup>14</sup> for the following three years, given the current National and Regional legislation without any other intervention. The Revenues are calculated, given the total National contribution from the Central Government and using the current allocation between Regions. The fiscal capacity of the Region is estimated by the Ministry of economics, Department of fiscal policies, while the other revenues are estimated based on the current values. In this way it is possible to estimate the deficit for the following three years. This gives the measure of the intervention necessary to budget balance at the end of the three years;
- **Plan of interventions.** The Region must provide a series of interventions to reduce the public expenditure and increase the public revenues. For each intervention, the Region must provide a date before which it commits itself to approve the intervention, a quantification of the impact on public expenditure or public revenues and a SWOT analysis. The complex of the

---

<sup>13</sup> In order to make the analysis on health expenditure easier to read and compare among regions, six expenditure aggregates were defined. All the reports on of the Deficit reduction plans are based on this classification. The expenditure aggregates used are Personnel, Goods and Services, Pharmaceutical expenditure, Home family doctors, Health services from private sector and Other expenditures.

<sup>14</sup> In order to make the analysis on health revenues easier to read and compare among regions, three revenues aggregates were defined. All the reports on of the Deficit reduction plans are based on this classification. The revenues aggregates are Regional health fund, Mobility and Other revenues. The first is the allocation to the Region of the FSN; the second is the net payment of health services bought by citizens of other Region in the Region's territory and of health services bought by citizen of the Region in other Regions' territory; the latter is composed by the revenues from patient copayments, revenues from health service sold to the private sector and other revenues.

intervention must achieve to budget balance in all the three years, given the possible treats and the temporary funds provided by the Central Government;

- **Programmatic forecast.** The Region must provide a forecast of the expenditure aggregates and revenues aggregate for the following three years that takes in account of the interventions on the revenues and on the expenditure side. The result must budget balance for all the three years, taking in account of the possible treats and of the temporary fund provided by the Central Government.

Moreover, the legislation on the reduction plans also provides that the Region must increase the Regional additional tax on income (IRPEF) and on productive activities (IRAP) up to the maximum level fixed by law. The additional tax on IRPEF and IRAP a must remain at the maximum level until the end of the period of the deficit reduction plan. In addition, a ceil is fixed on new recruitments. The Region can hire, in fact, only a fixed percentage of the employees who retire. These can be considered as the political cost the Regional Government must pay to have access to the extra funds. These are, in fact, the most expensive measures in political terms and, on the other end, are the most reliable measures to reduce the Regional deficit in the short term and also the ones whose effects is immediate and the easiest to verify by the Ministries.

In general, the Regions may have incentives to underreport the costs and over report the entrances to show better financial results and shift the necessary measures forward in time, to the next legislation. This is particularly true for the Regions with high health deficits and even more for the Regions undertaking a deficit reduction plan, especially about the forecasts of the costs and the financial results of the years in which they are involved in the DRP. Hence, an auditing society is chosen to guarantee the quality of the financial data and to provide quarterly reports on the respect of the deficit reduction plans. These must contain an analysis of the effect of the measures already

adopted and the estimation of the financial results for the following years, updated on the basis of the quarterly data.

Every quarter and at the end of the year the respect of the plan and of the time schedule of the interventions to adopt is verified by a commission composed of representatives of the Ministry of Economics, Ministry of health, State-Region Committee and of the Regions, called Committee for the respect of the commitments. This meeting are necessary to eventually redefine the adopted measures or to find new ones to correct possible problems emerged during the realization of the deficit reduction plan. In these meetings can be decided to give to the Regions part, or the whole, of the financial resources provided in the deficit reduction plan, if they respected the commitments. In case the Region, instead, does not fulfil the commitments of the Deficit reduction plan a Commissioner is appointed by the Council of ministers.

In case the Region does not fulfil the intermediate objectives, the Council of ministers can impose the Regional additional tax on IRAP and IRPEF to increase above the maximum level fixed by law. The ratio of it is similar to the previous case: the increase in the additional taxes is the most expensive measure in political terms and is the most reliable measure to reduce the deficit in the short term and the easiest to verify by the Ministries. The additional taxes are increased in measure such that to cover the negative effect of non-fulfilment of the intermediate objectives and is kept so high until the Committee does not verify it is not necessary anymore.

Moreover, all the non-mandatory transfers from Central Government are blocked and at the same time all the top manager of the health firms are automatically dismissed. These last measures are thought to reduce the possibility of expenses for the Regions and to break the eventual political link between Regional Governments and managers of the health firms.

The deficit reduction plan can be therefore thought as a temporary reduction of autonomy only for the Regions that cannot adequately contain their public health expenditure, with no effects on the

virtuous Regions. Notice that with respect to the analysis made by (Bordignon & Turati, 2003), this introduces an additional scenario that was not considered by the authors. That is the possibility of a partial bailing-out such that, in case one Region creates deficit, the utility for the Central Government is greater in case of bailing-out with deficit Reduction plan than in case of not bailing out and less than in case of no deficit. At the same time, for the Region the utility of bailing-out with deficit reduction plan is greater than in the case of not bailing-out and less than in the case of no deficit. This scheme can therefore induce the Regions to contain the expenditure instead of creating deficit and then be obliged to undertake a deficit reduction plan<sup>15</sup>.

I wrote above that the introduction of the deficit reduction plan has no effect on the virtuous Regions. This may not be entirely true. Virtuous Regions, in fact, could take advantage of the deficit generated by the non-virtuous Regions, in the sense that every time the Government raised the level of contribution to finance the SSN these extra resource were allocated between all the Regions with the same distribution of the initial contribution. Therefore, virtuous Regions received extra resources because of the deficit generated by non-virtuous Regions. If virtuous Regions expected Central Government to bailout, they could expand their health expenditure at expenses of their Regional resources, counting on future extra financial resources from the Central Government. In this sense, the introduction of the deficit reduction plan may have what can be called a new climate effect also in virtuous Regions.

### **Section 3.**

The expenditure data used are taken from the General Economic Relation (Relazione Generale sulla Situazione Economica del Paese). Thought this document is published by the Ministry of

---

<sup>15</sup> See Figure 5 for a representation in the extended form of the scenario created by the introduction of the deficit reduction plans

Economics, the data for the health system are provided by the Ministry of health, which is responsible of collecting, managing and for the quality of these data.

As written in Section 2, the original data were corrected to take in account of the timing of the collective contracts. It can be seen that the public health expenditure constantly increased in the period 2002-2010 with tendentially decreasing rate (see Figure 6 - Figure 9). It also grew in per capita terms, passing from around 1,400 euro to around 1,800 euro per capita (see Figure 10).

If one compares the dynamic of the public health expenditure in the period before the introduction of the deficit reduction plans with the dynamic in the following period, it appears quite clear that all the Regions had much lower increases in the period 2007-2010 than in the previous period (see Figure 11). Analyzing the differences between virtuous, non-virtuous and autonomous Regions it can be seen that (see Figure 13), non-virtuous Regions tended to have on average higher public health expenditure per capita than virtuous Regions in all the period 2002-2010, with values substantially similar to the public health expenditure per capita of the autonomous Regions. It must be noticed also that the non-virtuous Regions tended to have higher increases than virtuous and autonomous Regions in the per capita expenditure in the first years of the period but had on average systematically lower increases since 2006.

Therefore, from this initial analysis of the data, it seems that the non-virtuous Regions had in the period before 2007 higher public health expenditure per capita and higher annual increases than virtuous ones. In the following period the Regions contained the dynamic of the public health expenditure more than the previous period. This may be compatible with a general change of climate or a different level of attention on the public health expenditure. If this explanation was verified, the introduction of the deficit reduction plans may be just a signal from the Central Government that the approach is changed. On the other hand, it can be thought that the deficit reduction plans effectively had significant reduction effects on the dynamic of the public health

expenditure and it also reduced the expectations of additional financial resources for the virtuous Regions. The fact that the reduction of the dynamics starts in the year 2006 seems to work in favour of the change of climate explanation but it is also compatible with the fact that the non-virtuous Regions had to demonstrate with effective measures the will to undertake a plan of reorganization of the SSR, before the deficit reduction plan started.

### **External factors**

To analyze the effect of the economic growth on the health expenditure, I used the real national and Regional GDP and the National and Regional GDP per capita in absolute value and the percent variation of the National and Regional GDP ( $RGDPp_{c,r,t}, \Delta RGP_{r,t}, GDPp_{c,t}, \Delta GDP_t$ )<sup>16</sup>. The two variables at National level can be used to analyze the general effect of the economic growth, while the ones at Regional level can be used to analyze the impact of the different economic growth of the Regions. In order to avoid endogeneity problems, I used the variation of the GDP of the previous year to analyze the effects of the dynamic of the economy on the dynamic of the public health expenditure. It may be preferable to use some proxy that is more correlated with the dynamic of the economics in the same year; nevertheless it must also be thought that the public expenditure choices are usually made at least one year before and do not change much during the year.

The GDP per capita is instead a standardized indicator of the level of the economy of the Regions with respect to the numbers of citizen. Better than the Regional GDP it measures the income of the individuals and therefore the possibility that higher incomes may influence the required standards of health assistance and in this way the public health expenditure.

As pointed out before, it cannot be said a priori whether this variable has a positive or a negative effect on the dynamic of public health expenditure. An higher income per capita may in fact imply more demand for health and therefore higher increases in public health expenditure; on the contrary

---

<sup>16</sup> The data are taken from the ISTAT web site (<http://dati.istat.it/>).

an higher income per capita may be induce richer people to ask for better services in the private sector, with shorter waiting lists and therefore lower increases in the health expenditure.

The age distribution effect is analyzed using the percentage of people of age 75 or older resident in the Region ( $PO75_{r,t}$ ) and the percent variation of people of age 75 or older resident in the Region ( $\Delta PO75_{r,t}$ )<sup>17</sup>.

Among the possible geographic indicators, I have chosen the percentage of hilly areas and the percentage of mountainous areas on Regional territory ( $PHR_r, PMR_r$ ).

Several indicators of the quality of the environment, the healthy behaviours and lifestyle are collected by the ISTAT. Among the indicators of the environment, I have chosen:

- the percentage of population served by depuration plans of refluent water ( $PDR_{r,t}$ ) and its variation ( $\Delta PDR_{r,t}$ );
- the quantity of domestic water consumed per capita ( $DW_{r,t}$ ) and its percent variation ( $\Delta DW_{r,t}$ );
- the quantity of urban wastes ( $UW_{r,t}$ ) and its percent variation ( $\Delta UW_{r,t}$ );
- the percentage of urban wastes that are recycled ( $RUW_{r,t}$ ) and its variation ( $\Delta RUW_{r,t}$ );
- the percentage of public green ( $PG_{r,t}$ ) and its variation ( $\Delta PG_{r,t}$ );
- the percentage of buses and trolleybuses among the vehicles present on the territory ( $BT_{r,t}$ ) and its variation ( $\Delta BT_{r,t}$ );
- the percentage of cars among the vehicles present on the territory ( $CR_{r,t}$ ) and its variation ( $\Delta CR_{r,t}$ );
- the demand of public transportation per capita ( $PT_{r,t}$ ) and its percent variation ( $\Delta PT_{r,t}$ );
- the number of times the limit of pm10 for the protection of human health is exceeded ( $PM10_{r,t}$ ) and its percent variation ( $\Delta PM10_{r,t}$ ).

These are indicators of the attention given in the territory to the quality of the environment. It is not *a priori* clear how these factors affect the dynamic of the Regional public health expenditure.

---

<sup>17</sup> The data are taken from the ISTAT web site (<http://dati.istat.it/>).

Higher quality of the environment may, in general, have some impact on the quality of life and health and in turn have some effect of the public health expenditure. On the other hand high quality of the environment may be associated with more attention on health and therefore induce more demand for health services.

Among the indicators of healthy behaviours and lifestyle I used:

- the percentage of the number of individuals of age 3 or older having an adequate breakfast ( $BRK_{r,t}$ ) and its variation ( $\Delta BRK_{r,t}$ );
- the percentage of smokers among the individuals of age 14 or older ( $SMK_{r,t}$ ) and its variation ( $\Delta SMK_{r,t}$ );
- the ratio between the number of vehicles involved in car accidents over the number of vehicles present in the Region ( $CAC_{r,t}$ ).

Also for these indicators, it is not clear whether these have positive or negative effect on public health expenditure. For these indicators, in fact, it can be said, as said before, that a healthier behaviour can have positive effect on health and therefore reduce the demand for health services. Nevertheless, a healthier behaviour may be associated with more attention on health and in this way induce more demand for health services.

For the price effect, I used the deflator of the value of production of the sector “Health and social assistance”<sup>18</sup> ( $HSP_t$ ). This is not exactly equal to the price index faced by the health firms when buying drugs, goods and services but no better indicators are available. There is no Regional detail for this indicator; therefore, I used the National indicator for all the Regions. It may seem that a Regional indicator is not necessary because all the Regions should face, in principle, the same prices. It is not so in reality, as even contiguous Regions can have significant differences of prices. This probably depends on two factors: the aggregation of the demand and the waiting time for payment. The first factor is due to the organization of the purchases inside the Region: most of the

---

<sup>18</sup> In particular, I used the ratio between the production expressed in current prices and the production expressed in prices of the previous year. See <http://dati.istat.it/>.

Regions tend to centralize the purchases of goods and services of group of health firms, if not of all the health firms, into one aggregate purchaser. This has more bargaining power with respect to the providers and the supplying firms and in this way can realize lower prices. Notice that the dimension of the aggregate purchaser is not the only relevant variable, the bargaining ability of the structure and the timing of purchasing is as well important. The waiting time for payment can have in some Regions even a bigger impact. Unfortunately, no systematic data are collected but a recent analysis by Confartigianato<sup>19</sup> showed that they vary between 87 days in Friuli Venezia Giulia and 873 days in Calabria with a National average value of 113 day (these values are to be compared with the European standard recently fixed at 30 days). This phenomenon is probably responsible of significant raises in the purchasing costs of goods and services as providers and supplier can internalize the waiting time risk and raise the prices. This rational reaction can in turn induce Regions to expand the waiting time for payment because of lack of financial resources.

### **Internal organization of the SSR**

In order to study the weight of the administrative costs on the dynamic of the health expenditure I used the percentage of expenditure for administrative personnel among the total Regional public health expenditure ( $ADM_{r,t}$ ) and its variation ( $\Delta ADM_{r,t}$ ). The idea is that higher budget for the bureaucrats may be associated with higher power to influence the decisions on public health expenditure. Of course, as said, this can have either a positive or a negative effect on the dynamics of the health expenditure. In fact, higher power on the bureaucrats may be due to a rent seeking mechanism of the bureaucracy and therefore tend to increase the public health expenditure dynamics; on the other hand, more power on the bureaucrats may be associated with more controlling power on the public expenditure and therefore tend to reduce the dynamic of public health expenditure. The variable used to proxy the administrative costs does not fully capture the

---

<sup>19</sup> See [http://www.confartigianato.it/UfficioStudi\\_publ.asp](http://www.confartigianato.it/UfficioStudi_publ.asp)

dimension of the administrative costs. These, in fact, contain, for example, also expenditure in goods and services. Unfortunately, there are no data that can help distinguish the final usage of each expenditure voice.

I used the percentage of copayment among the total amount of the Regional health expenditure ( $CP_{r,t}$ ) and its variation ( $\Delta CP_{r,t}$ ) to test whether these measures could reduce the dynamic of the Regional health expenditure.

To study the effect of the Regional political cycle I used three variables. The first one ( $EL_{r,t}^{t-1}$ ) is a dummy variable that, for each Region, is equal to one in the year before the year of the Regional elections and zero otherwise. The second one ( $EL_{r,t}^t$ ) is a dummy variable that, for each Region, is equal to one in the year of the Regional elections and zero otherwise. The third one ( $EL_{r,t}^{t+1}$ ) is a dummy variable that, for each Region, is equal to one in the year after the year of the Regional elections and zero otherwise<sup>20</sup>.

Mobility is an important variable to analyze the impact of the quality of the Regional health service on public health expenditure. Citizens in fact choose to go to another Region to be cured only if the quality of the service of the other Region is considered by the patient better than the quality of her Region's one. Therefore, the net mobility can be thought as an indicator of the quality of the Regional health service. I have used the percentage of net mobility on total health expenditure ( $MB_t$ ) and its variation ( $\Delta MB_t$ ) to analyze the impact of the quality of the SSR on the on the Regional public health expenditure.

---

<sup>20</sup> See Table 6 for the years in which each Region had the Regional elections.

### **Relations between SSR and private providers.**

Increased expenditure may be driven also by providers' market power. The higher is the power of a provider, the more bargaining power it has in defining its budget. Market power can create a mechanism of expansion of the share of health services bought from private providers. On the other hand, it is possible that the private sector may be more efficient than the public one. In this case, a higher presence of the private sector can reduce the Regional public health expenditure. In order to analyze the importance of the market power on the dynamic of the public health expenditure, I used the percentage of health services bought from the private sector on total Regional public health expenditure ( $PS_{r,t}$ ) and its variation ( $\Delta PS_{r,t}$ ).

Unfortunately, I do not have knowledge of any concentration indicators of private providers of health services. This would be an important variable to analyze the structure of the health services market and verify whether the presence of a free market can have reduction effect on dynamic of the public health expenditure<sup>21</sup>.

### **Relation between Central Government and Regional Governments**

In order to study the effect of the deficit reduction plans I introduced two variables. The first is a dummy variable that is equal to one for all the Regions if the year is greater or equal than 2007 and zero otherwise ( $P07_t$ ). This variable can catch the effect of what I called the general change of climate. That is, if the correspondent parameter is significant, this could be interpreted in favour of the thesis that the introduction of the deficit reduction plans affected the dynamic both of the virtuous and the non-virtuous Regions. The correspondent value of the estimate could be interpreted as the quantification of the effect of the general change in the climate on the dynamic of the public health expenditure of both the virtuous and non-virtuous regions.

---

<sup>21</sup> For an analysis of the effects of concentration on health expenditure see (Gaynor & Town, 2011)

I also added a dummy variable that is equal to one in the last year of validity of a collective contract and zero otherwise. This variable is necessary to quantify what I called the last year effect ( $LYE_t$ ). If the correspondent parameter were significant, it could be interpreted in favour of the thesis that the distortion toward the last year of validity of the collective contracts to concede higher increment in wages and salaries generated significant differences in the dynamic of the public health expenditure. Otherwise, it would mean that these higher increases were absorbed, reducing other expenditures.

In order to study the differences between virtuous, non-virtuous and autonomous Regions, I introduced a set of three dummy variables: this first ( $VR_t$ ) is equal to one for virtuous Regions and zero otherwise; the second ( $NVR_t$ ) is equal to one for non-virtuous Regions and zero otherwise; the third ( $AUT_t$ ) is equal to one for autonomous Regions and zero otherwise. If the correspondent parameters were significant this would mean that virtuous, non-virtuous or autonomous Regions have different trend from other Regions. I considered non-virtuous the Regions that undertook a deficit reduction plan in the period 2006-2010 for excessive deficit. The Regions Piedmont, Apulia and Sardinia undertook a different plan because they did not fulfil other requirements but had fulfilled the financial equilibrium: these Regions can be considered virtuous for the purpose of this paper. The definition used for autonomous Regions is slightly different from the definition given in the Italian Constitution. I considered autonomous the Region that do not receive any contribution from the Central Government for their Regional health services and I considered the Autonomous Provinces of Bolzano and Trento separately, instead of the Region Trentino Alto Adige, coherently with what is done in all the document on public health expenditure. The two provinces have in fact the same autonomy of legislation and organization of the autonomous Regions (see Figure 15 and Table 7 for the classification of the Regions and autonomous provinces into the three categories).

I also introduced three sets of interrelation effects. The first set of interrelations between each variable and the dummy variables regarding virtuous, non-virtuous and autonomous Regions. The second set of interrelation between each variable and the introduction of the deficit reduction plans. The third set of interrelation between each variable the dummy variables regarding virtuous, non-virtuous and autonomous Regions and the introduction of the deficit reduction plans jointly.

The first set of interactions allows testing whether the different importance of some of the driving factors can explain the differences between virtuous, non-virtuous Regions and autonomous Regions. These are important to understand the areas in which the Regions should intervene to reduce the dynamic of the Regional public health expenditure.

The second set of interrelations instead is useful to analyze the general effect of the introduction of the deficit reduction plans on the driving factors. If the introduction of the deficit reduction plans had had a general effect of change of climate, it would be important to understand in which way it had modified the driving factors of the public health expenditure.

Finally, the third set of interrelations is important to understand the different impact the introduction of the deficit reduction plans had on virtuous, non-virtuous and autonomous Regions. It is particularly important to test whether the deficit reduction plans helped to correct the problems of the non-virtuous Regions and let them converge to a virtuous path. This would make the difference between temporary and permanent reduction of the regional public health expenditure.

## **Section 4. Results**

In order to analyze the driving factors of the public health expenditure, I performed a stepwise regression using the BIC to select the regressors. In Table 1 the results of the stepwise regression are shown.

These show that virtuous, non-virtuous and autonomous Regions do not have statistically significant differences in the trend of the public health expenditure. All the Regions present a positive trend, which makes the public health expenditure of about 5% per year. The differences between virtuous and non-virtuous Regions, instead, seem to be due mostly to the internal organization of the SSRs.

---

<i>Costant</i>	0,05*** 0,003
$\Delta CP_{r,t} \times NVR_r$	-28,36*** 3,342
$\Delta EL_{r,t}^{t-1} \times NVR_r$	-0,04*** 0,01
$EL_{r,t}^{t+1} \times NVR_r$	0,03*** 0,01
$\Delta ADM_{r,t} \times NVR_r$	3,44*** 0,99
$\Delta BT_{r,t} \times P07_t$	10,24*** 3,03
$\Delta PS_{r,t} \times VR_r \times P07_t$	-2,17*** 0,61
$PS_{r,t} \times NVR_r \times P07_t$	-0,15*** 0,04
$\Delta ADM_{r,t} \times NVR_r \times P07_t$	-7,30*** 1,21
$\Delta CP_{r,t} \times NVR_r \times P07_t$	44,93*** 7,10
$\Delta ADM_{r,t} \times AUT \times P07_t$	-6,76*** 2,24
$R^2$	0,48

---

Table 1 – Results of the Stepwise regression. To each significant coefficient is associated the correspondent significance level. If the coefficient is significant at the 10 per cent level then one asterisk is associated; if the coefficient is significant at the 5 per cent level then two asterisks are associated; if the coefficient is significant at the 1 per cent level then three asterisks are associated

The first significant interrelation is related to the effect of copayments on the dynamic of the Regional public health expenditure. The sign of the coefficient indicates that an increase in the percentage of copayment on Regional health expenditure is associated with a decrease of public health expenditure.

A first explanation can be given by a demand effect subsequent to the fact that the non-virtuous Regions have lower average income per capita than the remaining Regions. The real GDP per capita of the non virtuous Regions was in fact in the period 2002-2009 around 5.000 euro less than the real GDP per capita in the virtuous Region and 4.000 less than the real GDP per capita in the autonomous Regions (see Table 4). An increase of copayments can induce individuals with lower income to reduce the demand of health services more than individuals with higher incomes. This explanation is nevertheless not fully coherent with the fact that the parameter of the Regional GDP per capita or of the percent variation of the Regional GDP. On the other hand it must be noticed that the GDP per capita or the percent variation of the GDP do not capture the income inequalities in the Region. The Regions have in fact different internal economic inequalities (see Table 5) with five of the seven non-virtuous Regions being in the six most unequal Regions.

Another possible explanation may be an incorrect choice of the copayment level. When the level of copayment is too high, citizen may prefer the private providers to the public ones. If the cost of the private providers is not much higher than the copayment and offers the same service with more efficiencies and less waiting times, the patient can choose to pay a bit more for better services. This explanation is not coherent with the fact that non-virtuous Regions tend to have lower percentage of copayment on Regional public health expenditure and lower percentage of copayment on Regional GDP (see Table 9 and Table 10) than the remaining regions. In reality, to better test this argument, it would be necessary to compare the prices in the private sector of the services for which citizen contribute to the expenditure. Unfortunately, these kind of data are not available at the moment.

Finally, another explanation may be that non-virtuous Regions may have more difficulties to reduce the internal inefficiencies of the SSR. In this case, the Regions may deliberately choose to use copayments to reduce Regional public health expenditure because it is the most effective instrument to reduce the public health expenditure dynamics and increase revenues.

The second difference between non-virtuous Regions and the other ones is the effect of the political cycle on the dynamic of the public health expenditure. The values of the coefficients indicate that the year before the election non-virtuous Regions tend to increase their public health expenditure of 2,8% more than the trend; the year after the election instead they tend to reduce the dynamic of public health expenditure of around 4,3%. These estimates are coherent with the political cycle theory. Nevertheless, it must be noticed that the final value in the years after the Regional election is lower than in absence of political cycle but it needs three years after the election to have any cumulative saving (see Figure 12 for the effect of the political cycle on the public health expenditure using the estimated coefficient). In this sense, it can be said that the controls on the Regions already present before the introduction of the deficit reduction plan had some positive effect, though limited.

The last significant interrelation useful to analyze the differences between virtuous and non-virtuous Regions shows that, unlike the other Regions, in non-virtuous Regions, an increase in the percentage of expenditure for administrative personnel is associated with an increase of Regional public health expenditure. This is coherent with the bureaucracy effect and suggests that part of the inefficiencies of the non-virtuous Regions is due to an inefficient bureaucracy that instead of containing the public health expenditure tends to expand the costs.

In synthesis, public health expenditure in the non-virtuous Regions tends to grow more than in other Regions mostly due to inefficient bureaucracies. The containing measures had weak effects and mostly subject to the political cycle: increasing the public health expenditure the year before the

Regional elections and containing the costs the year after the elections. The result is that in the years before the introduction of the deficit reduction plans the non-virtuous Regions had higher public health expenditure per capita and higher trend than the other ones (see Figure 13 and Figure 14).

It can be said that the introduction of the deficit reduction plan had a general reduction effect on the dynamic of the public health expenditure. The introduction of the deficit reduction plan, in fact, shows a significant interaction with the variation of the percentage of busses and trolley-busses. As said, the percentage of busses and trolley-busses is one of the indicators of the quality of environment. It must be noticed that the percentage of busses and trolley-busses decreased on average over the period 2007-2010. This can be interpreted as the natural reaction to stricter budget constraints. After the introduction of the deficit reduction plans, Regions may have made general cuts to public expenditure in the quality of the environment, instead of reducing only the public health expenditure. The general climate of more attention on containing the costs has induced the Regions to reduce the expenditures in the areas regarding the quality of the environment.

Virtuous and non-virtuous Regions seem to have had similar approaches of containing the expenditure after the introduction of the deficit reduction plans. Both of them in fact have significant negative interrelation between the introduction of the deficit reduction plans and percentage of health services bought from the private sector or its variation. The two variables capture the same phenomenon in a different way. That is the containment of the public health expenditure was made more on the internal costs than on the external costs. Both virtuous and non-virtuous Regions experienced increases in the expenditure for health services bought from the private sector with the virtuous Regions having higher increases than non-virtuous Regions after the introduction of the deficit reduction plans (see Figure 16 and Table 8). The variation of the percentage of health services from the private sector for the virtuous Regions and the percentage of health services from the private sector for the non-virtuous Regions have similar dynamic capturing

the measures of containment of the internal costs of the SSR, though the two variables are different in scale and sign.

Moreover, the Regions undertaking deficit reduction plans changed the relation between the percentage of copayment and the dynamic of the public health expenditure. Before the introduction of the deficit reduction plans, in fact, the increases of the percentage of copayment was associated with lower increases in the public health expenditure; while, after the introduction of the deficit reduction plans, an increase in the percentage of copayment is associated with higher increases in the public health expenditure. It must first be noticed that most of the positive increase in the percentage of copayment for the Regions undertaking deficit reduction plans occurred in 2007 when the deficit reduction plans started and most of the non-virtuous Regions had high increases of expenditure. This suggests that the increases in the copayment percentage was thought by the Regions undertaking a deficit reduction plans as a necessary measure to cover the increasing costs of the SSR. After having reduced the trend of the expenditures, the Regions could also reduce the weight of the copayment. These are, in fact, together with taxes the most political expensive measures a Region can adopt. However, while non-virtuous Regions cannot reduce taxes until they do not budget balance, they can reduce the weight of the copayment if they adopt alternative measures with the same effect on the regional balance.

The introduction of the deficit reduction plans changed the relation between an increase in the percentage of the expenditure for administrative personnel and the dynamic of the public health expenditure for the Regions undertaking a deficit reduction plan. Before the introduction of the deficit reduction plan, in fact, an increase in the percentage of the expenditure for administrative personnel was associated with higher increases in public health expenditure; while, after the introduction of the deficit reduction plan, an increase in the percentage of the expenditure for administrative personnel is associated with a reduction of the public health expenditure. The

expenditure for administrative personnel in the Regions undertaking deficit reduction plans decreased in 2007 both in absolute value and in percentage probably because of the ceiling on new hirings; it started increasing again then at slower pace than the rest of the personnel. This can be interpreted as a change toward a more efficient bureaucracy by increasing the productivity of the administrative personnel. A similar effect can be observed in the autonomous Regions with slightly lower impact on public health expenditure.

In synthesis, the differences between virtuous and non-virtuous Regions seem to be due to the different weight of the political cycle and to the different efficiency level of the administration of the SSR. The introduction of the deficit reduction plans put more attention on containing costs. The Regions seem to have reduced the dynamic of the expenditure by containing the internal costs in the areas regarding the quality of the environment, confirming in this way the explanation of a general change of climate. The introduction of the deficit reduction plans seemed to have reduced the inefficiencies of the bureaucracy of the SSRs of the non-virtuous Regions and the internal costs of the SSRs, while the weight of the political cycle remains an issue. The introduction of the deficit reduction plans, then, can be said to have had general positive effects in containing the Regional public health expenditure and more effectively in the Regions undertaking deficit reduction plans. These can be considered permanent reductions of the dynamic of the Regional public health expenditure – at least in part – as the Regions undertaking deficit reduction plans seemed to have solved part of the inefficiency problems of the administration of their SSRs that drove the excessive expansion of their public health expenditure. The autonomous Regions experienced, as well, a reduction of the dynamic of the public health expenditure, obtained by increasing the efficiency of the administrative personnel but with smaller effects on public health expenditure.

Finally, some observation must be made on the goodness of fit of the regression. The  $R^2$  of the regression is just 0,48. This means that the regressors can explain less than half of the variability of

the dynamic of the regional public health expenditure. This can be in part due to the fact that in some of the variables used for the stepwise regression there are several missing values, which reduce the prediction ability of these variables. Moreover, most of the variables used as indicators of the quality of the environment and on healthy behaviour are weak proxies of the measured phenomena.

The deflator of the production of the sector Health and social assistance, surprisingly, is not significant. This is probably because, as said before, this is not exactly the correct indicator of the prices faced by the health firms when buying goods and services.

Moreover, it must be noticed that public health expenditure is not the whole health expenditure and that in the public expenditure mechanism there are many non-market components that may not be captured with the available data.

Finally, it must be said that in this analysis only the short-term effects were studied, while health expenditure may have some long-term components that cannot be studied with such a short panel data. This may account for the non significance of most of the variables that are generally considered the main driving factors of the health expenditure.

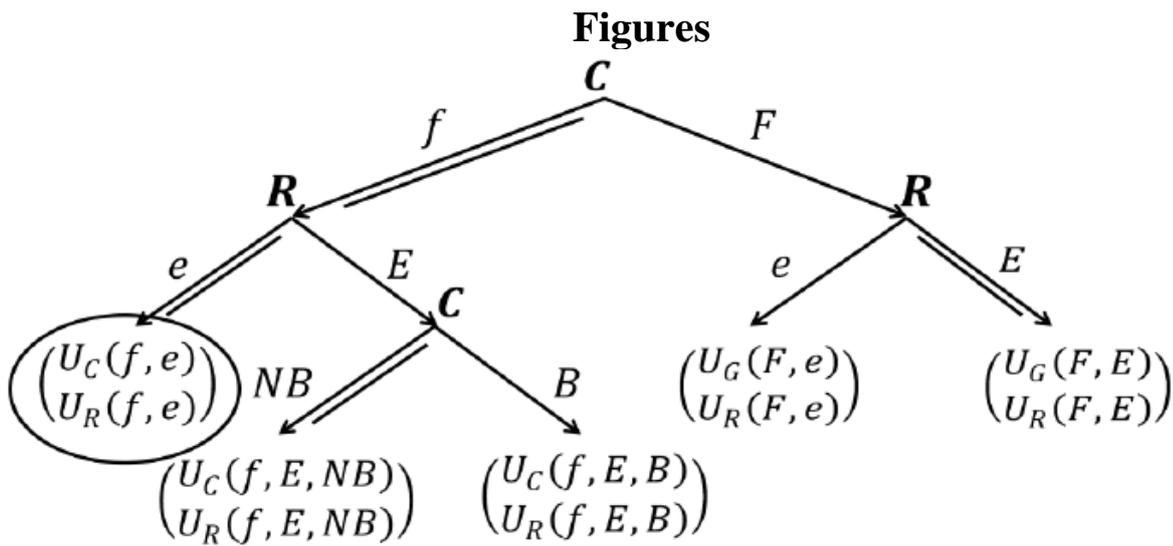
These last two comments may make clear why the regression could not explain the good part of the variability of the regional dynamic of the public health expenditure.

## **Section 5. Conclusions**

In the present work, I have analyzed the driving factors of the Regional public health expenditures in Italy. The relevance of the argument is crucial in all the developing country facing increases in the public health expenditure that is going to be non-sustainable in the long term for some countries.

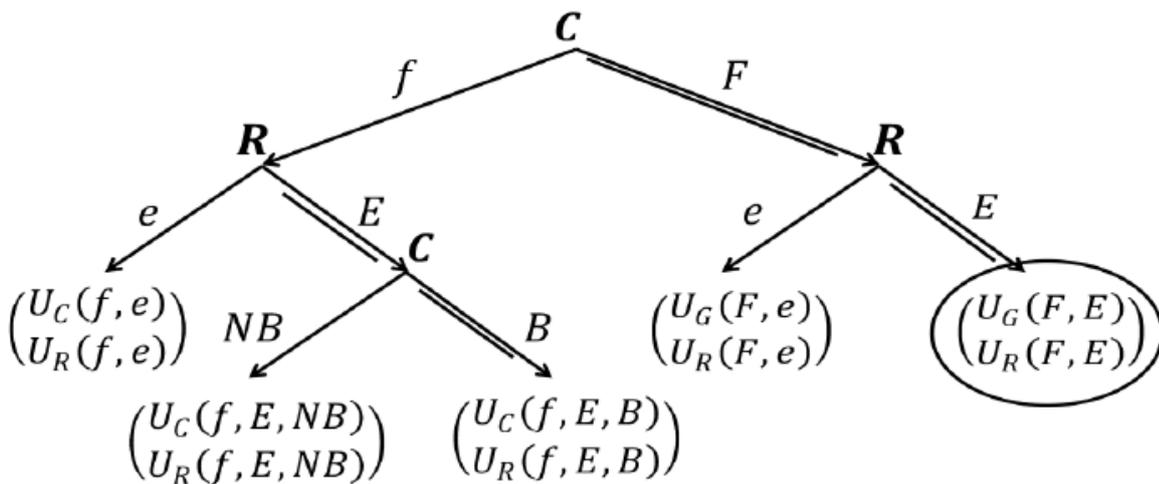
The results of the analysis show that, in the short-term, the public health expenditure is mostly

driven by the inefficiencies of the internal organization of the SSRs and the dead-weight cost of the political cycle and of the bureaucracy. None of what are considered the driving factors of the health expenditure seems to have statistically significant effects on the dynamic of the public health expenditure in Italy, at least in the short-term. This is probably because of the quality of the data available and the lack of variables to measure some non-market phenomena. Moreover, the excessive weight of the internal inefficiencies may cover the effect of some relevant factors or some of them may have long-term effects, which cannot be captured with a short-term analysis like the present one. Italy has introduced in 2006 an important instrument to intervene more effectively in the most problematic Regions. The innovation had positive effects in containing the costs in all the Regions and more effectively in the Regions undertaking deficit reduction plans, with permanent reduction of the dynamic of the public health expenditure.



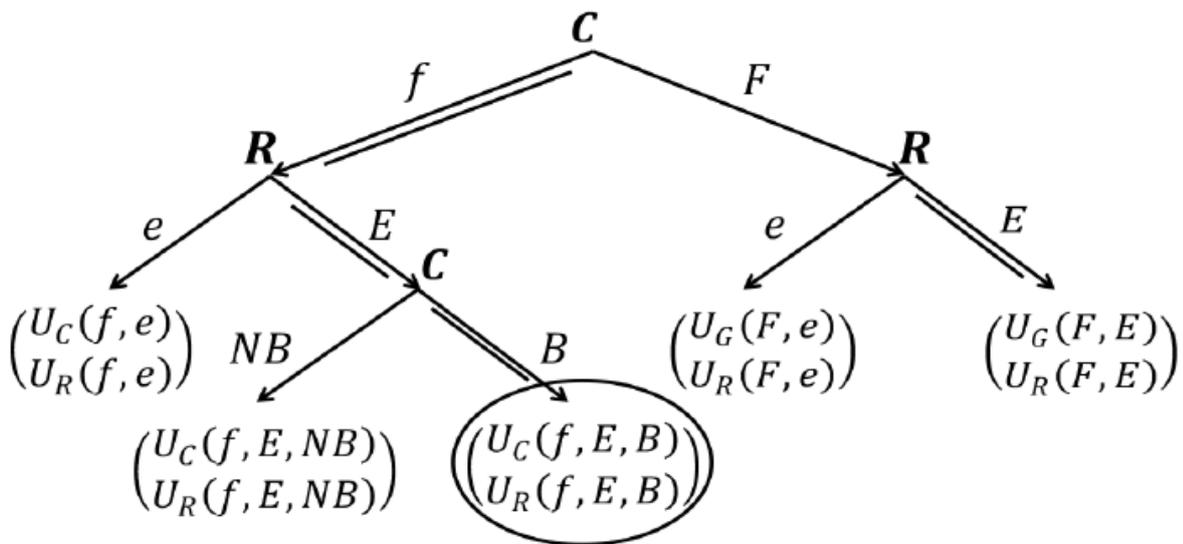
$$U_C(f, E, NB) > U_C(f, E, B)$$

Figure 2 – Scenario 1 of the three presented by (Bordignon & Turati, 2003). If the utility of the Central Government in case of non bailing-out the health deficits created by the Regions is greater than the utility in case of bailing-out, the Central Government chooses low level of financial resources and the Regions choose low level of expenditure.



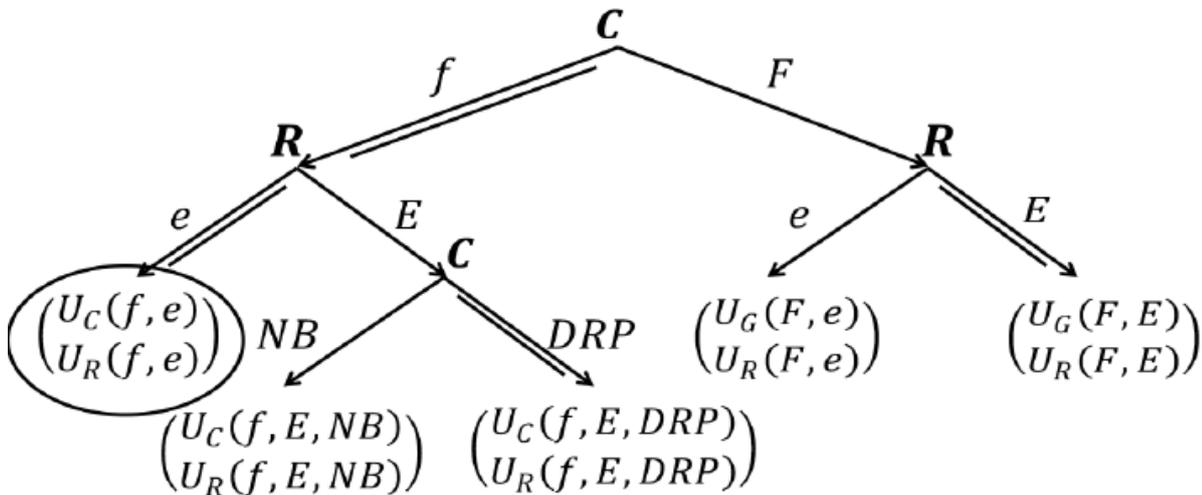
$$U_C(f, E, NB) < U_C(f, E, B) < U_C(F, E)$$

Figure 3 – Scenario 2 of the three presented by (Bordignon & Turati, 2003). If the utility of the Central Government in case of non bailing-out the health deficits created by the Regions is lower than the utility in case of bailing-out and this is lower than the utility in case of high level of financial resources, the Central Government chooses high level of financial resources and the Regions choose high level of expenditure.



$$U_C(f, E, NB) < U_C(F, E) < U_C(f, E, B)$$

Figure 4 – Scenario 3 of the three presented by (Bordignon & Turati, 2003). If the utility of the Central Government in case of non bailing-out the health deficits created by the Regions is lower than the utility in case of high level of financial resources and this is lower than the utility in case of bailing-out, the Central Government chooses low level of financial resources, the Regions choose high level of expenditure and Central Government chooses to bail-out.



$$U_C(f, E, DRP) > U_C(f, E, NB)$$

$$U_R(f, e) > U_R(f, E, DRP)$$

Figure 5– The new scenario created by the introduction of the deficit reduction plans. If the utility of the Central Government in case of partial bailing-out with deficit reduction plans is greater than the utility of non bailing-out and the utility of the Regions in case of low level of expenditure is greater than the utility of the Regions in case of partial bailing-out with deficit reduction plans, the Central Government chooses low level of financial resources and the Regions choose low level of expenditure.

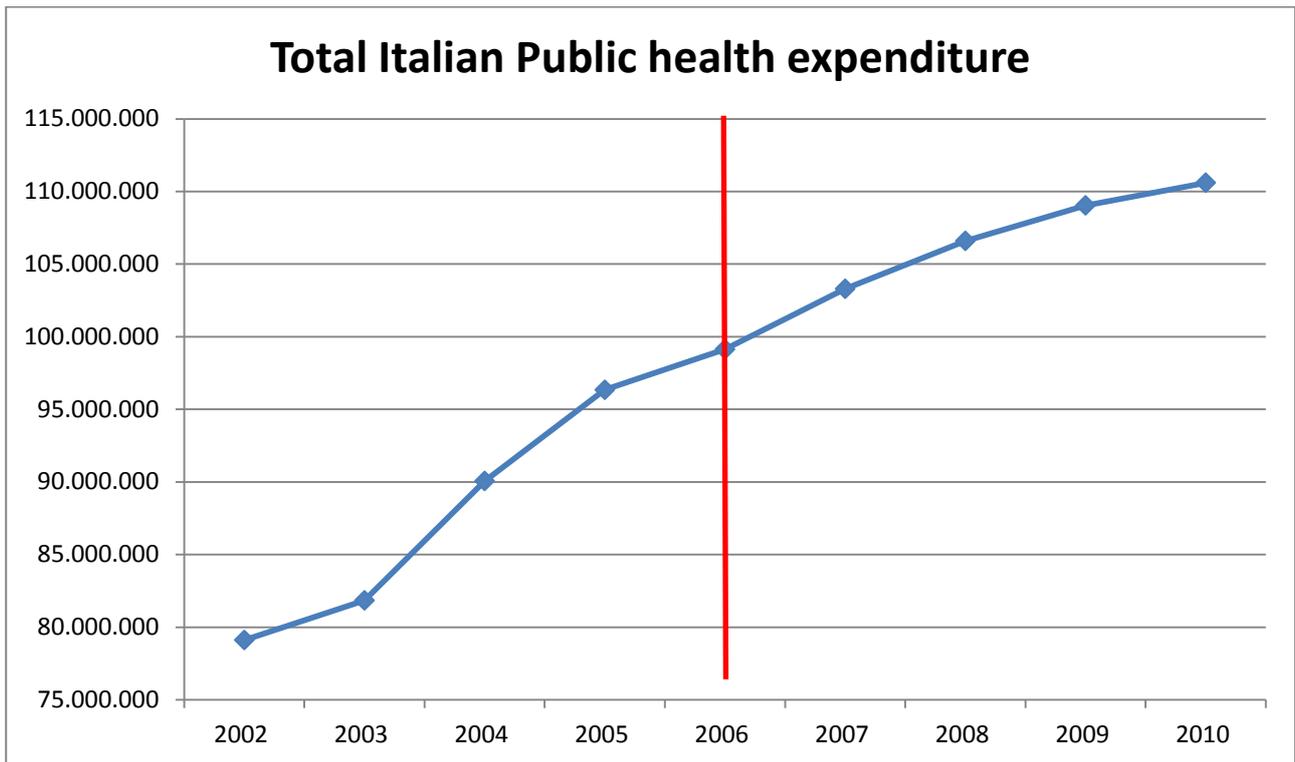


Figure 6 – Total Italian health expenditure, years 2002-2010 – Current values

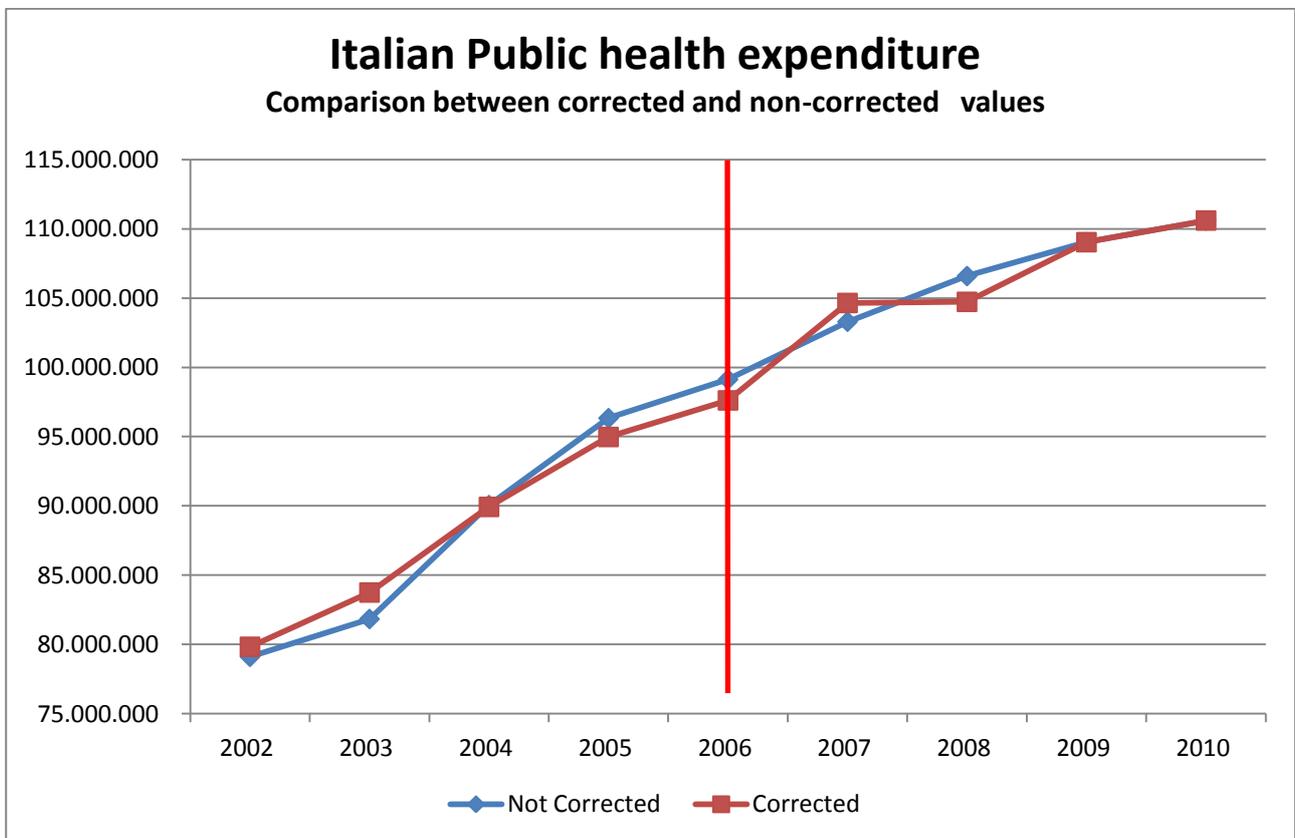


Figure 7 – Total Italian health expenditure, years 2002-2010 – Comparison between current data and corrected data

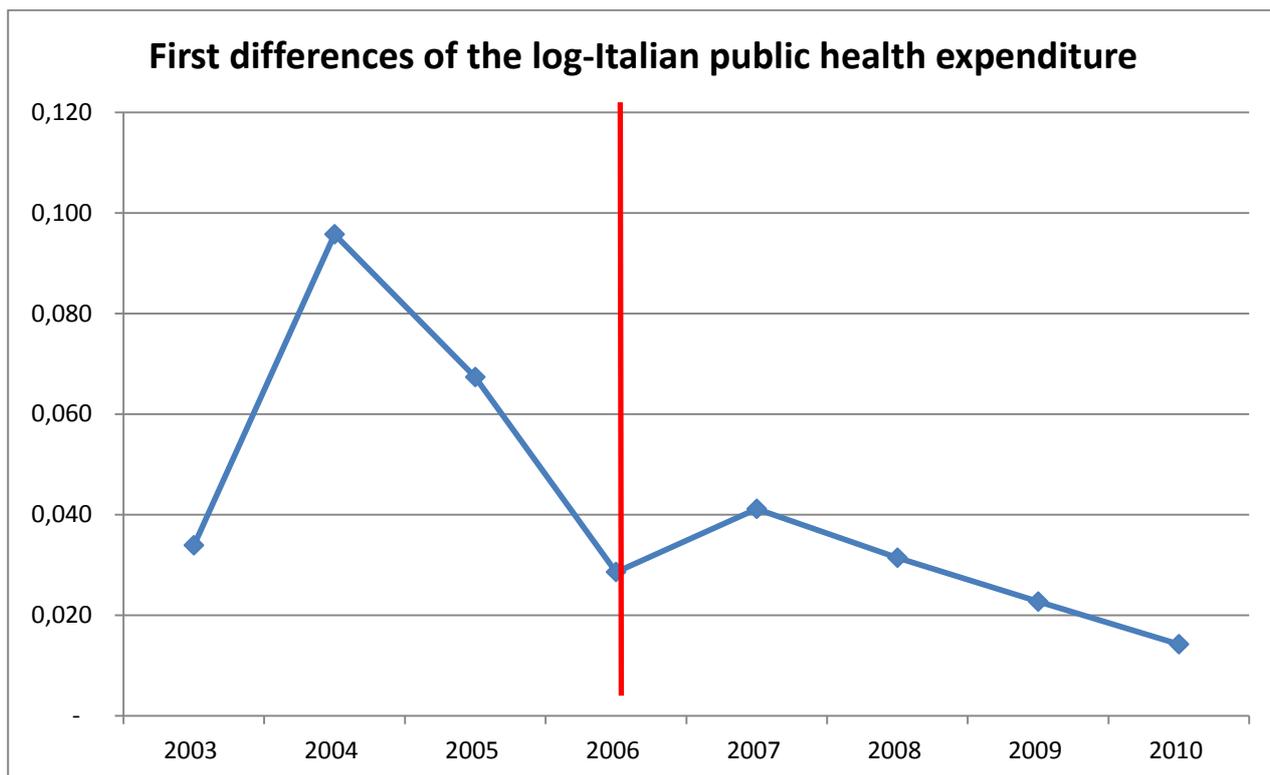


Figure 8 - First differences of the logarithm of Italian public health expenditure (years 2003-2010)

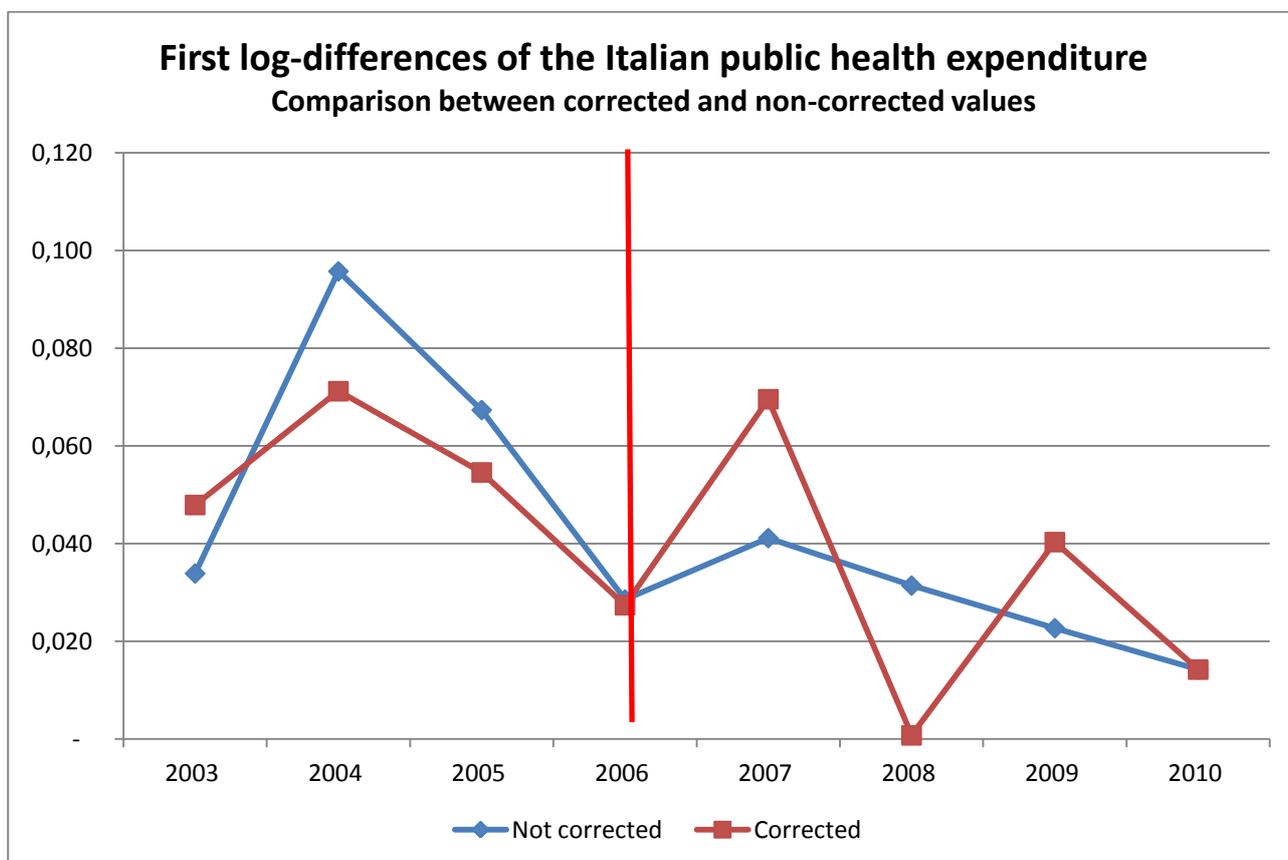


Figure 9 - First differences of the logarithm of Italian public health expenditure - Comparison between corrected and not corrected values (years 2003-2010)

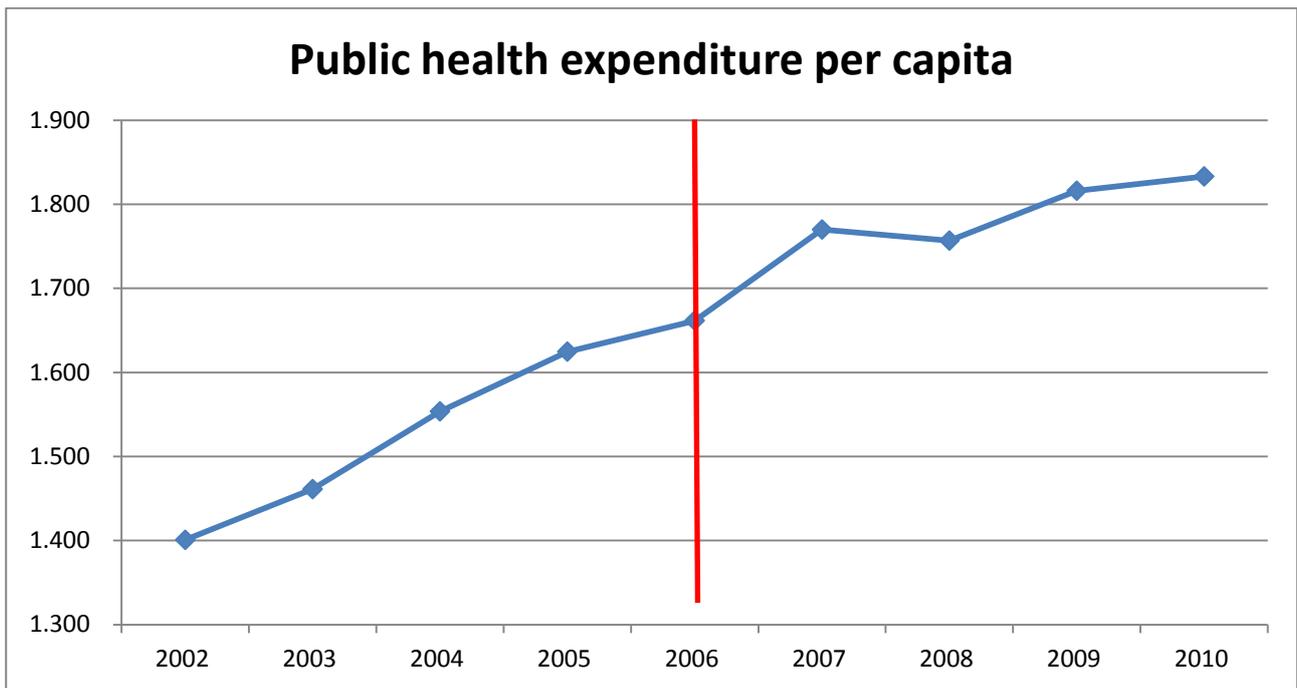


Figure 10 – Italian public health expenditure per capita

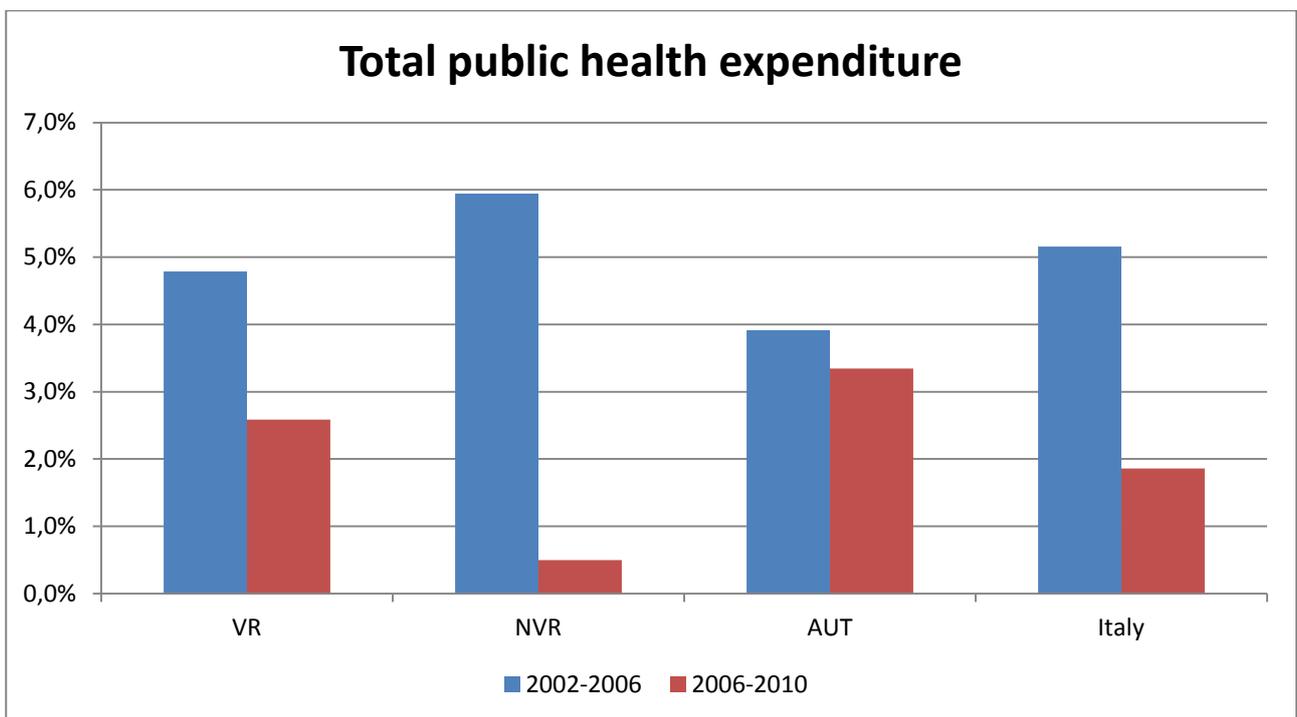


Figure 11– Comparison of the percent variation of total health expenditure in the period 2002-2006 and 2006-2010. Comparison between virtuous, non-virtuous and autonomous Regions

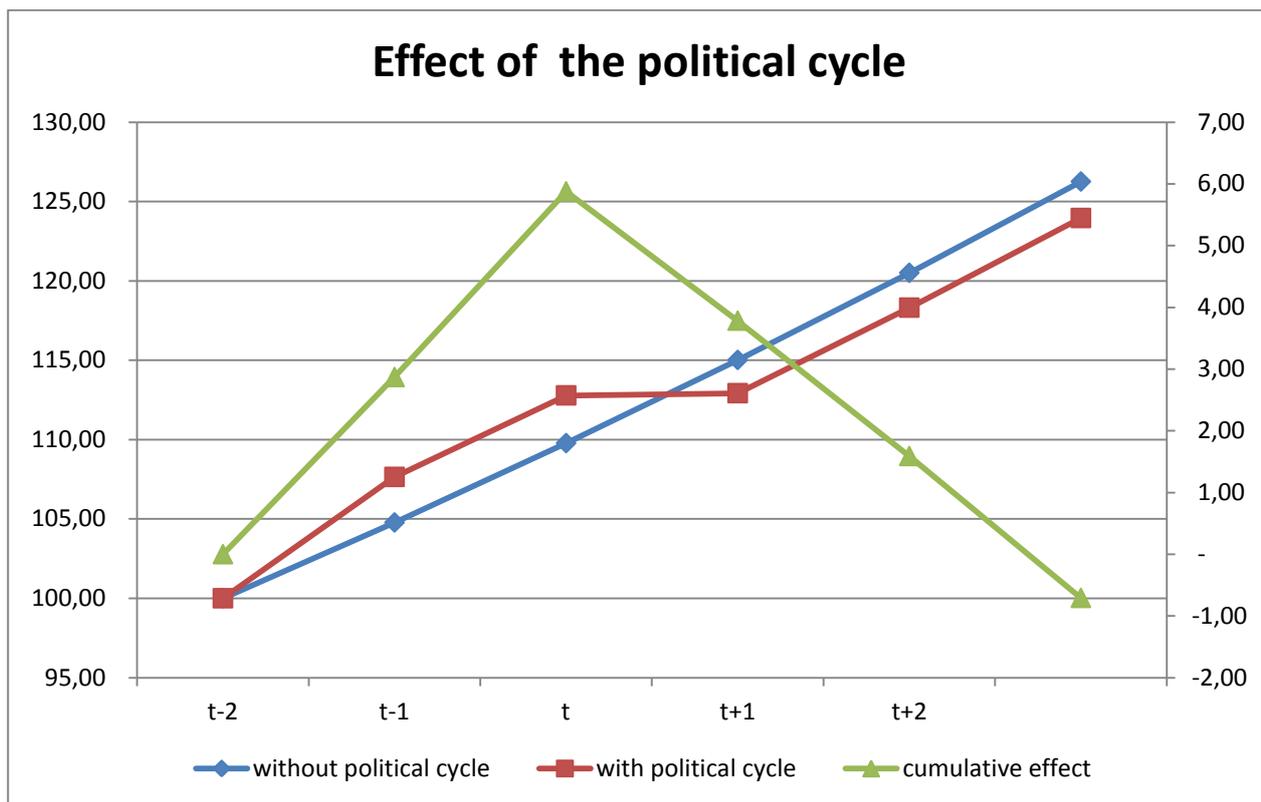


Figure 12 – Cumulated and current effect of the political cycle using the estimated values

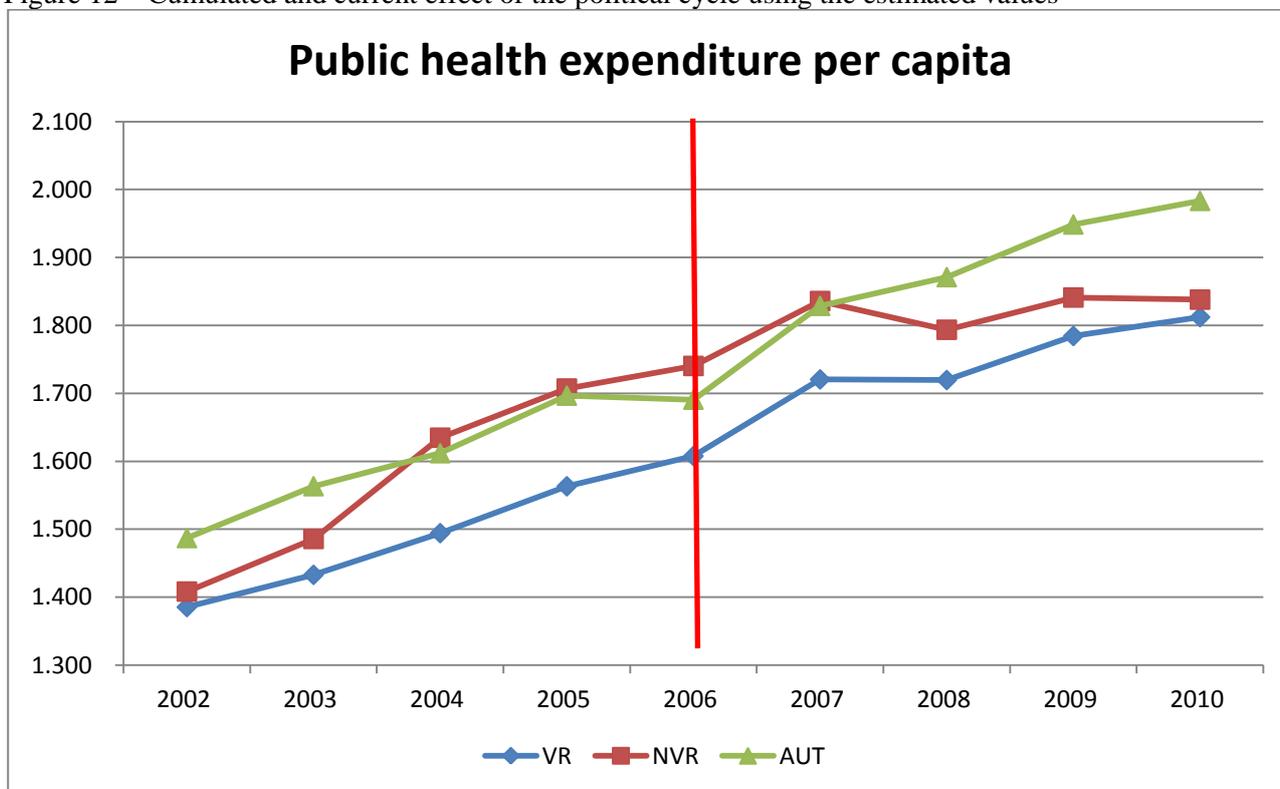


Figure 13 - Public health expenditure per capita - Comparison between virtuous, non-virtuous and autonomous Regions

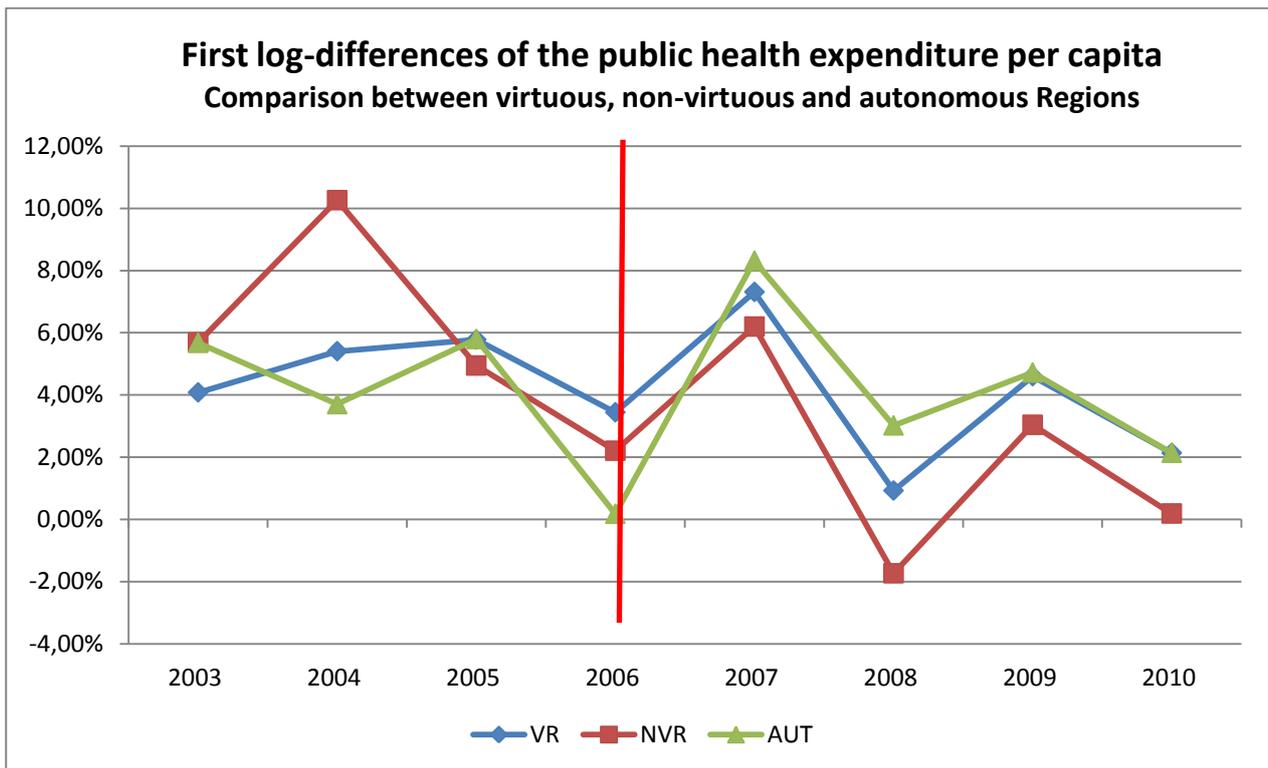


Figure 14 First log-differences of the public health expenditure per capita - Comparison between virtuous, non-virtuous and autonomous Regions



Figure 15 – Classification of the Regions in virtuous (in blue), non-virtuous (in red) and autonomous (in green).

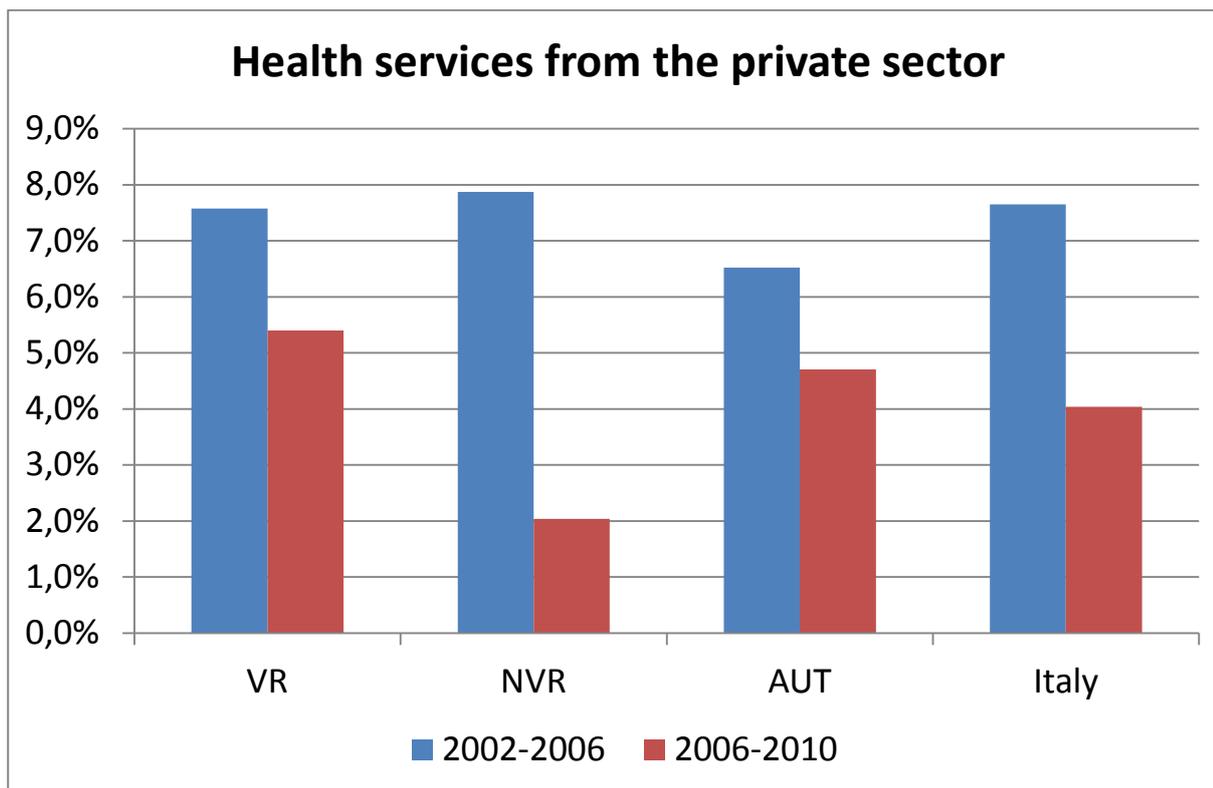


Figure 16 – Comparison of the percent variation of the expenditure for health services bought from the private sector in the period 2002-2006 with the percent variation of the expenditure for health services bought from the private sector in the period 2006-2010. Comparison between virtuous, non-virtuous and autonomous Regions

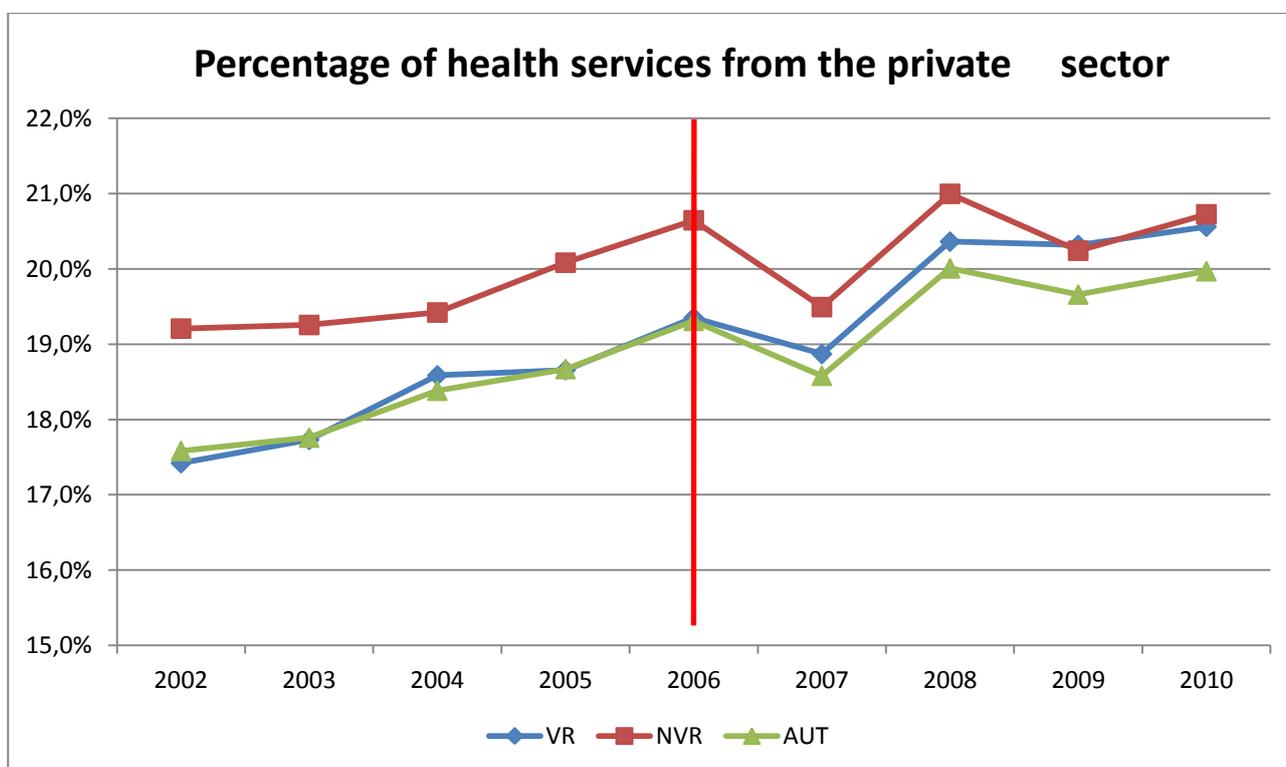


Figure 17 – Percentage of expenditure for health services from the private sector among total health expenditure (years 2002-2010)

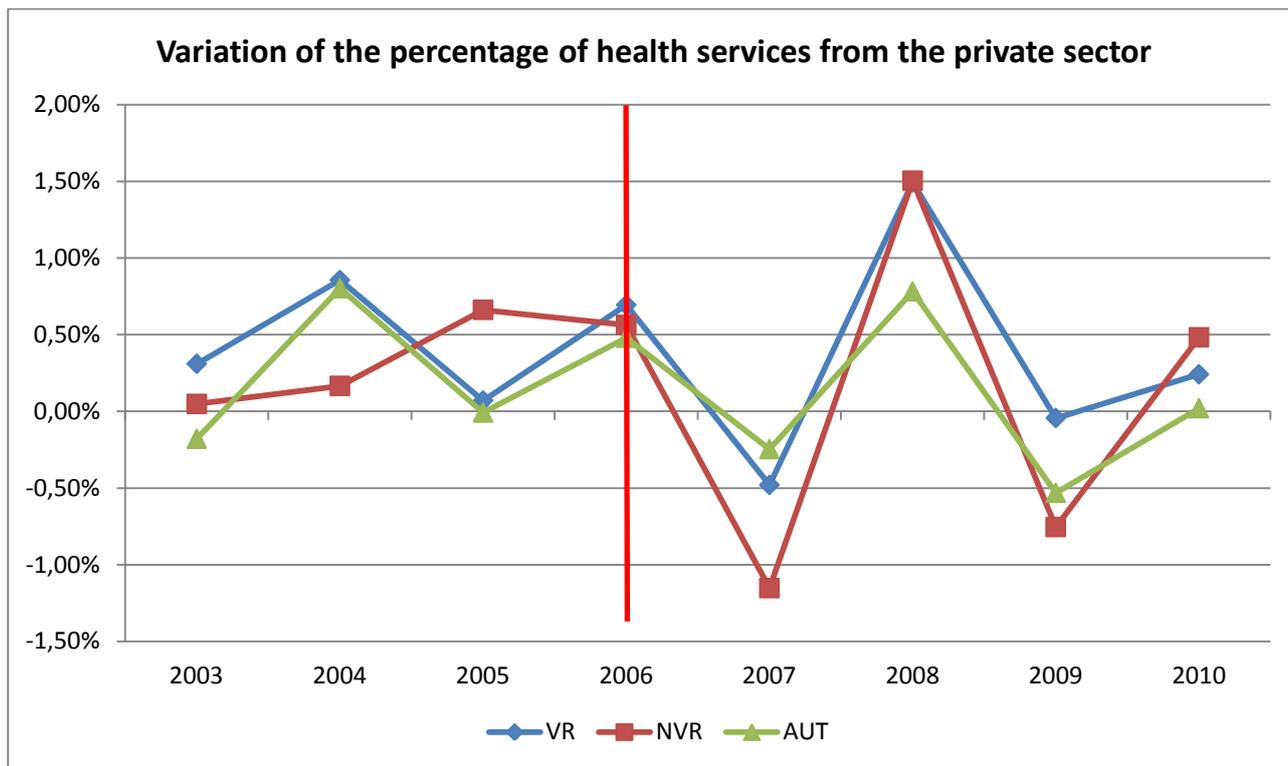


Figure 18– Variation of the percentage of expenditure for health services from the private sector among total health expenditure (years 2003-2010)

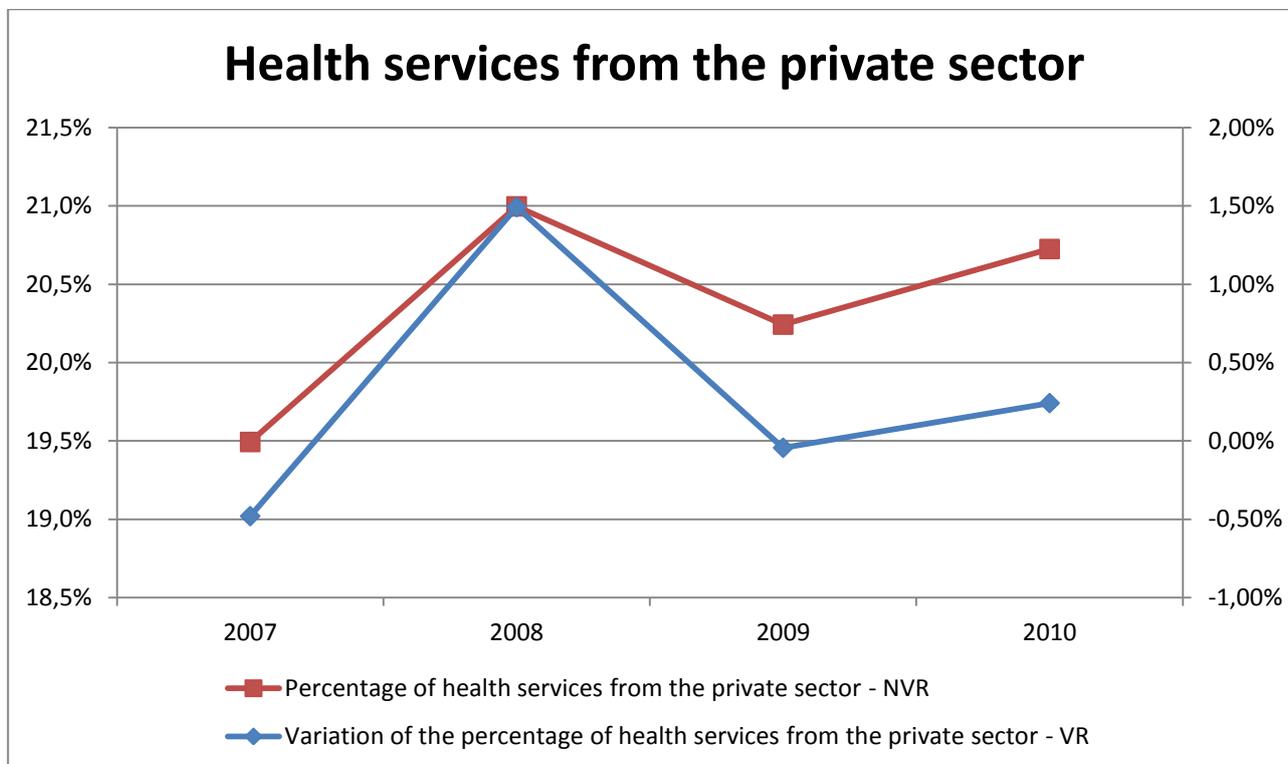


Figure 19 – Comparison of the dynamic of the percentage of the expenditure for health services from the private sector among total health expenditure for non-virtuous Regions (left axis) and the variation of the percentage of the expenditure for health services from the private sector among total health expenditure for virtuous Regions (right axis) in the years 2007-2010.

## Tables

<b>Collective contract</b>	<b>Period of validity of the collective contract</b>	<b>Year in which the contract was approved</b>
Employees	2002-2003	2004
Doctors	2002-2003	2005
Managers	2002-2003	2005
Family doctors	2003-2004	2005
Employees	2004-2005	2006
Doctors	2004-2005	2006
Managers	2004-2005	2006
Family doctors	2005-2006	2007
Employees	2006-2007	2008
Doctors	2006-2007	2008
Managers	2006-2007	2008

Table 2 – Comparison of the period of validity of the collective contracts with the year in which these were signed

	CCNL 2002-2003		CCNL 2004-2005		CCNL 2006-2007	
	2002	2003	2004	2005*	2006	2007
Employee	2,46%	4,31%	1,90%	3,11%	1,70%	3,15%
Doctors	2,46%	4,31%	1,90%	3,11%	1,70%	3,15%
Managers	2,46%	4,31%	1,90%	3,11%	1,70%	3,15%

Table 3 – Increases in wages and salaries in each year for the dependent personnel of the SSN.

\* Part of the increase in the year 2005 (0,7%) is valid starting on December 31, 2005

	2002	2003	2004	2005	2006	2007	2008	2009
VR	23.668	23.422	23.429	23.371	23.782	23.920	23.362	21.837
NVR	17.475	17.320	17.503	17.517	17.641	17.815	17.478	16.742
AUT	21.257	21.185	21.246	21.325	21.629	21.923	21.617	20.612

Table 4 – GDP per capita for virtuous, non virtuous and autonomous Regions

Region	Gini concentration index
Piedmont	0,291
Aosta Valley	0,310
Lombardy	0,295
<b>Liguria</b>	<b>0,290</b>
<i>Autonomous Province of Bolzano</i>	<i>0,298</i>
<i>Autonomous Province of Trento</i>	<i>0,280</i>
Veneto	0,266
Friuli-Venezia Giulia	0,265
Emilia-Romagna	0,297
Tuscany	0,283
Umbria	0,280
Marche	0,289
<b>Lazio</b>	<b>0,324</b>
<b>Abruzzo</b>	<b>0,263</b>
<b>Molise</b>	<b>0,319</b>
<b>Campania</b>	<b>0,327</b>
Apulia	0,310
Basilicata	0,289
<b>Calabria</b>	<b>0,314</b>
<b>Sicily</b>	<b>0,335</b>
Sardinia	0,292
Italia	0,314

Table 5 – Gini concentration index per Region year 2008. Source ISTAT

---

---

	Years
Piedmont	2000 2005 2010
Aosta valley	2003 2008
Lombardy	2000 2005 2010
Autonomous Province of Bolzano	2003 2008
Autonomous Province of Trento	2003 2008
Veneto	2000 2005 2010
Friuli	2003 2008
Liguria	2000 2005 2010
E. Romagna	2000 2005 2010
Tuscany	2000 2005 2010
Umbria	2000 2005 2010
Marche	2000 2005 2010
Lazio	2000 2005 2010
Abruzzo	2000 2005 2010
Molise	2000 2005 2010
Campania	2000 2005 2010
Apulia	2000 2005 2010
Basilicata	2000 2005 2010
Calabria	2000 2005 2010
Sicily	2001 2006 2008
Sardinia	2004 2009

---

---

Table 6 – years in which the Regions had Regional election

<b>Virtuous</b>	<b>Non-virtuous</b>	<b>Autonomous</b>
Piedmont	Liguria	Aosta Valley
Lombardy	Lazio	A. P. Bolzano
Veneto	Abruzzo	A. P. Trento
Emilia Romagna	Molise	F.V.Giulia
Tuscany	Campania	Sardinia
Umbria	Calabria	
Marche	Sicily	
Apulia		
Basilicata		

Table 7 – Classification of the Region in Virtuous, non.virtuous and autonomous Regions

	2002	2003	2004	2005	2006	2007	2008	2009	2010
VR	7.765	8.232	9.108	9.686	10.398	10.909	11.882	12.414	12.833
<i>var %</i>		6,02%	10,65%	6,35%	7,35%	4,92%	8,91%	4,48%	3,37%
NVR	5.666	6.012	6.719	7.300	7.672	7.707	8.159	8.108	8.317
<i>var %</i>		6,12%	11,76%	8,65%	5,10%	0,45%	5,86%	-0,62%	2,57%
AUT	605	629	704	745	779	829	913	915	936
<i>var %</i>		4,04%	11,84%	5,89%	4,50%	6,36%	10,16%	0,24%	2,34%
Italy	14.035	14.873	16.532	17.732	18.850	19.445	20.953	21.438	22.086
<i>var %</i>		5,97%	11,15%	7,26%	6,30%	3,16%	7,76%	2,31%	3,03%

Table 8 – Expenditure for health services bought from the private sector (in millions of euro)

	2002	2003	2004	2005	2006	2007	2008	2009
VR	0,10%	0,10%	0,11%	0,11%	0,11%	0,12%	0,11%	0,12%
NVR	0,05%	0,05%	0,06%	0,07%	0,07%	0,08%	0,07%	0,08%
AUT	0,09%	0,10%	0,10%	0,10%	0,10%	0,11%	0,11%	0,12%
Italy	0,08%	0,09%	0,09%	0,09%	0,10%	0,11%	0,10%	0,11%

Table 9 – percentage of copayments on health expenditure

---

	2002	2003	2004	2005	2006	2007	2008	2009
<i>VR</i>	0,08%	0,08%	0,09%	0,09%	0,09%	0,10%	0,10%	0,10%
<i>NVR</i>	0,02%	0,02%	0,03%	0,03%	0,03%	0,04%	0,04%	0,04%
<i>AUT</i>	0,01%	0,01%	0,01%	0,01%	0,01%	0,01%	0,01%	0,01%
<i>Italy</i>	0,12%	0,12%	0,12%	0,13%	0,14%	0,15%	0,14%	0,15%

---

Table 10 – percentage of copayment on Regional GDP

## Bibliografia

- Ambrosiano, M. F., & Bordignon, M. (2007). *Internal Stability Pacts: The European Experience*. EEGM papers.
- Ambrosiano, M. F., & Bordignon, M. (2007). *Internal Stability Pacts: The European Experience*. EEGM papers.
- Attanasio, O. P., Banks, J., & Tanner, S. (2002). Asset holding and consumption volatility. *Journal of political economy* .
- Barer, M. L., & Lomas, J. (1996). Re-minding our Ps and Qs: medical cost controls in Canada. *Health Affairs* .
- Bodenheimer, T. (2005). High and Rising Health Care Costs. *Annals of internal medicine* .
- Bodenheimer, T. (2005). High and Rising Health Care Costs. Part 2: Technologic Innovation. *Annals of internal medicine* .
- Bordignon, M., & Turati, G. (2003). Bailing out expectation and health expenditure in Italy. *CESifo working paper No. 1026* .
- Chernew, M. E., Hirth, R. A., S., S. S., Ermann, R., & Fendrick, A. M. (1998). Managed care, medical technology, and health care cost growth: a review of the evidence. *Medical Care Research and Review* .
- Downs, A. (1967). *Inside bureaucracy*. Rand Corp.
- Enthoven, A. C. (2003). Employment-based health insurance is failing: now what? *Health Affairs* .
- European Observatory on Health System and Policies. (2009). Italy Health system review. *Health Systems in Transition* , 11 (6).
- Fuchs, V. R., & Sox, H. C. (2001). Physicians' views of the relative importance of thirty medical innovations. *Health Affairs* .
- Garston, N. (1993). *Bureaucracy: three paradigms*. Springer.
- Gaynor, M., & Town, R. J. (2011). Competition in health care marke. *Handbook of Health Economics* .
- Gelijns, A., & Rosenberg, N. (1994). The dynamics of technological change in medicine. *Health affairs* .
- Hall, R. E., & Jones, C. I. (2007). The Value of Life and the Rise in Health Spending. *Quarterly Journal of Economics* .
- Keeler, E. B., Melnick, G., & Zwanziger, J. (1999). The changing effects of competition on non-profit and for-profit hospital pricing behavior. *Journal of Health Economics* .
- Liebowitz, S. (1994). *Why Health Care Costs Too Much?* Policy Analysis.
- Newhouse, J. P. (1993). An iconoclastic view of health cost containment. *Health Affairs* .

Niskanen, W. J., & Niskanen, W. A. (2007). *Bureaucracy & representative government*. Transaction Publishers.

Reinhardt, U. E. (2003). Does the aging of the population really drive the demand for health care? *Health Affairs* .

Robinson, J. C. (2004). Consolidation And The Transformation Of Competition In Health Insurance. *Health Affairs* .

Soroka, S. N., & Wlezien, C. (2005). Opinion–Policy Dynamics: Public Preferences and Public Expenditure in the United Kingdom. *British Journal of Political Science* .

Yip, W. C. (1998). Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. *Journal of Health Economics* .

# Heterogeneous expectations and stock market participation

Hermes Morgavi

## **Abstract**

Stock market non-participation is a contradiction of the classic financial economic models. This was mostly studied with models based on the entry costs approach. Few models use instead the heterogeneous expectation approach, which has been mostly used for asset pricing models.

In this paper, I present some facts collected in the literature on stock market participation and I compare the two approaches above on the basis of the ability of the models to explain these facts. This problem is more frequently studied using the entry costs approach. Nevertheless, this approach is not able to explain some of the facts presented in the literature.

For this model I instead propose a model based on the heterogeneous expectations approach. The model is very flexible and can be developed in many directions. Using very simple assumptions, this model can explain in a very easy way the participation choice in a general equilibrium framework and is coherent with most of the facts on stock market participation.

## **Introduction**

Though stock market non-participation is by itself a contradiction of the standard financial economic models, the stock market participation choice had been for long time neglected as a relevant economic question. The standard analysis of this phenomenon is the presence of entry costs. This approach is intuitively and has important features that made it very popular.

In this paper, instead, I propose a different approach to explain the participation choice based on different beliefs among individuals. This approach has been much more used in a partial equilibrium environment to explain differences in the portfolio compositions among individuals. Only in very few cases, these models include an explanation of the non-participation choice.

This approach can easily model the stock market participation choice and is coherent with most of the observed facts on stock market participation.

The paper is organized as following. In Section 1 is presented the problem of stock-market participation as a contradiction of the standard financial economics models. In Section 2 instead the two main approaches are presented: the entry costs approach and the heterogeneous expectations approach. In Section 3 the proposed model is shown. In Section 4 and Section 5, respectively, the predictions and the caveats of the model are presented. In Section 6 the two approaches are compared with respect to the ability to explain the main facts on stock market participation. Finally, in Section 7 conclusions are drawn.

### **Section 1. The contradiction of stock market non-participants**

Economic models had excluded for long time stock market participation from the relevant arguments of financial economics. Nevertheless, stock market non-participant are a serious

---

contradiction of the standard Consumption based CAPM model. In fact, in the standard C-CAPM the no-arbitrage condition, assure that in a competitive market all the individuals have rational expectations, face the same state-prices, the same stochastic discount factor and therefore, in equilibrium each individual is indifferent between participating or not to the stock market. The basic assumption of the model can be summarized in:

- A.1 Each individual has a income sequence  $e_t$ ;
- A.2 In the real market units of consumptions are sold and each individual can freely buy any amount of consumption units  $c_{i,t}$  at the current market price;
- A.3 In the stock market are traded a set of assets, which at each time  $t$  guarantee some payoff  $x_{j,t}$  with some known distribution  $P(\cdot)$ ;
- A.4 Each individual can freely buy and sell any quantity of any asset  $q_{j,t}$  at the current market price  $p_{j,t}$ ;
- A.5 Each individual has the same utility function  $U(c_{i,t})$ ;
- A.6 Each individual has the same discount factor  $\beta$ ;
- A.7 Each individual maximizes his expected utility function, given his sequential budget constraint;
- A.8 In equilibrium the sum of the amount of stock detained buy all individuals must be equal to one;
- A.9 In equilibrium the sum of the amount of bonds detained buy all individuals must be equal to zero.

In this framework, there is no way to explain why some individuals participate to the stock market and some others do not. Each individual in fact face the same Euler equations

$$1 = E_t[M_{i,t,t+s}R_{j,t+s}] \quad (3)$$

where

$$M_{i,t,t+s} = \beta^s \frac{U'(c_{i,t+s})}{U'(c_{i,t})} \quad (4)$$

and

$$R_{j,t+s} = \frac{x_{j,t+s}}{p_{j,t}} \quad (5)$$

If, for simplicity we assume no initial wealth, and only two periods, from the first order condition, it can be written:

$$U'(y_0 - q_f p_f - q_1 p_1 - \dots - q_J p_J) = \beta E_0 [R_j U'(y_1 + q_f R_f - q_1 R_1 - \dots - q_J R_J)], j = f, 1, \dots, J \quad (6)$$

where  $f$  indicates the risk-free asset.

If zero stockholding were an interior solution, (6) would hold also for  $q_1 = q_2 = \dots = q_J = 0$ . Then it could be written:

$$(E_0 [R_j] - R_f) E_0 [U'(y_1 + q_f R_f)], \quad j = 1, \dots, J \quad (7)$$

This, in presence of equity premium, is a contradiction. Each individual therefore participates to the stock market with different portfolio, according with the different sequences of endowments.

The study of different stock-market participation behaviour began having more relevance when (Mankiw & Zeldes, 1991) first understood that the differences between stockholders and non-stockholders could help explain the equity premium puzzle. In their paper, in fact, they argue that, analyzing the consumption of stockholders and non-stockholders separately, it is possible to explain the equity premium with a coefficient of relative risk aversion of only six. These results were confirmed by following studies on different data sets. (Attanasio, Banks, & Tanner, 2002), for example, show that the differences between shareholders and non-shareholders in UK are consistent with the implication of the Consumption CAPM. (Vissing-Jørgensen, 2002), instead, estimates two different values of the elasticity of intertemporal substitution for stockholders and non-stockholders

in the UK. These estimates are coherent with values of the coefficient of relative risk aversion considered admissible by (Mehra & Prescott, 1985). The same result is achieved by (Vissing-Jørgensen & Attanasio, 2003). While these studies have been somewhat successful in reconciling the C-CAPM with the empirical evidence, they take limited participation as given and do not attempt to rationalize it.

## **Section 2. Two models for stock market participation: Entry costs and heterogeneous beliefs**

### **The entry costs approach**

Most of the authors explain the choice of not to participate to the stock market as due to some entry cost or fixed cost linked to the participation to the stock market. This idea was first introduced by (Haliassos & Bertaut, 1995) and was soon followed by many other authors. Costs that deter entry into the stock market may take several forms. (Vissing-Jørgensen, 2002) categorizes participation costs as fixed entry costs, fixed and variable transaction costs and per period trading costs. The first category comprehends the fixed costs of brokerage and the time spent implementing the trade, as well as the time and money spent to acquire all the relevant information to build up the optimal trading strategy. The second category, instead, comprehends the variable costs of brokerage and the bid and asks spread. The last category comprehends the time spent to follow the stock market and take information on future returns and taxes on returns.

The quantification of the threshold levels of costs that discourage households from participating to the stock market followed two approaches. The first approach is based on simulations of an opportunely calibrated life-cycle model and solved numerically to find the levels of costs that rationalize the non-participation choice; the second instead is based on direct estimation of the entry costs.

(Haliassos & Bertaut, 1995), followed the first approach. (Alan, 2006), instead, quantified that a one-time entry cost of around 2% of the permanent component of the labour income is sufficient to discourage households from investing in the stock market.

The second approach was followed first by (Luttmer, 1999) who estimated a fixed cost of at least 5.75% of monthly consumption to rationalize the non-participation choice. Other authors followed the same approach. Among the others, (Vissing-Jørgensen, 2002) analyzing the data on the financial portfolios of US portfolios derived that a very small amount of per period trading costs could explain between 50% and 75% of the cases in which households decided not to participate to the stock market. (Paiella, 2007), instead measures the forgone gains of non-participants. According to her quantification, they range from 0.7% to 3.3% of their spending in nondurable goods and services. Thus, the entry cost level must be greater or equal than this amount for the non-participants.

### **The heterogeneous expectation approach**

The first to use the heterogeneous expectations approach was (William, 1977) who studied the effect of heterogeneous beliefs on asset prices. Not surprisingly, he tried to overcome some of the problems generated by the CAPM, which predicts that all individuals hold portfolios essentially identical in composition. After this seminal paper, the idea of heterogeneous beliefs was used, among the others, by (Abel, 1990) and more recently by (Basak, 2005) and (Carceles-Poveda & Giannitsarou, 2008) to study the dynamics of asset prices. Some models include limited stock market participation, as (Basak & Cuoco, 1998) or (Gallmeyer & Hollifield, 2008), but do not investigate the non-participation choice.

The first to pose the question of heterogeneous beliefs and stock market participation were (Haliassos & Bertaut, 1995) using a probit regression approach to test whether, among other factors, differences in investor beliefs can account for the stock market participation decision. (Cao, Wang,

& Zhang, 2005), instead, endogenizes the non-participation choice and studies the impact of non-participation to the stock on asset prices in a partial equilibrium environment, modelling the differences in investors beliefs in terms of model uncertainty.

Surprisingly, many authors find the heterogeneous beliefs approach very useful to model asset prices dynamics but very few of them use this approach to model the stock market participation choice.

The main objective of this paper is to introduce a simple model to account for the stock market participation choice in a general equilibrium model. The main feature of the model is the simplicity of the rule to individuate the stock market participants and the numerous available directions of improvement due to the extreme flexibility of the model.

### **Section 3. The model**

Let us assume there are  $N$  individuals with finite time horizon  $T$ . This means that in each period there are  $T$  generations present in the economy. At age zero, individuals do not receive any initial endowment. They just receive an annual labour income  $W_t$ , which is equal for each worker. Each individual chooses whether to consume or to hold assets. In the market there are only two types of assets: a risk free asset of price  $P_t^B$ , which returns one unit of consumption the following year, and a risky asset, of price  $P_t^S$ , which returns the following year dividends equal to  $D_{t+1}$ .

Therefore each individual, of age  $a$  at time  $t$ , solves the following utility maximization problem:

$$\max_{C_{i,a,t}, \theta_{i,a,t}, B_{i,a,t}} E_{i,t} \left[ \sum_{s=0}^{T-a} \beta^s U(C_{i,a+s,t+s}) \right] \quad (8)$$

s.t.

$$\begin{aligned} C_{i,a+s,t+s} + P_{t+s}^S \theta_{i,a+s,t+s} + P_{t+s}^B B_{i,a+s,t+s} \\ = (P_{t+s}^S + D_{t+s}) \theta_{i,a+s-1,t+s-1} + B_{i,a+s-1,t+s-1} + W_{t+s} \end{aligned} \quad (9)$$

where

$$U(C) = \begin{cases} \frac{C^{1-\gamma}}{1-\gamma}, & \text{if } \gamma > 1 \\ \ln C, & \text{if } \gamma = 1 \end{cases} \quad (10)$$

The parameters  $\gamma > 1$  and  $\beta \in (0,1)$ , respectively, represent the household risk aversion and time discount factor. The variables  $\theta_t$  and  $B_t$  are the amount of equity shares and risk-free one period bonds held, respectively. The supply of equity is assumed to be constant and normalized to one, and bonds are assumed to be in zero net supply. I assume that individuals can buy only positive amounts of shares.

From the first-order condition, it can be written:

$$P_t^S = E_t [M_{t,t+1} (P_{t+1}^S + D_{t+1})] \quad (11)$$

$$P_t^B = E_t [M_{t,t+1}] \quad (12)$$

where

$$M_{t,t+j} = \beta^j \frac{C_{t+j}^{-\gamma}}{C_t^{-\gamma}} \quad (13)$$

Alternatively, we can rewrite the equations in terms of the gross asset returns as

$$1 = E_t[M_{t,t+1}R_{t+1}^S] \quad (14)$$

where

$$R_{t+1}^S = \frac{D_{t+1} + P_{t+1}^S}{P_t^S} \quad (15)$$

$$1 = E_t[M_{t,t+1}R_{i,t+1}^f] \quad (16)$$

where

$$R_{i,t+1}^f = \frac{1}{P_{i,t}^B} \quad (17)$$

On the production side, let us consider a representative firm that produces a single consumption good  $Y_t$  combining the aggregate capital stock  $K_{t-1}$  with labor according to the following constant returns to scale technology.

$$Y_t = Z_t K_{t-1}^\alpha N_t^{1-\alpha} \quad (18)$$

where  $Z_t$  is a random productivity shock following the stationary process:

$$z_t = \log Z_t = \zeta + \rho \log Z_{t-1} + \varepsilon_t \quad (19)$$

with  $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$  and  $\rho \in (0,1)$ .

The firm maximizes the current value of the firm, equal to the discounted value of the dividends paid out to the stock-holders. These are equal to the gross profits minus the cost of self financing the investment,  $I_t$ . Thus,  $D_t = Y_t - W_t N_t - I_t$ . Furthermore, capital fully depreciates at the end of each period.

The representative firm therefore solves:

$$\max_{Y_t, I_t} E_t \left[ \sum_{t=0}^{\infty} M_{0,t} D_t \right] \quad (20)$$

s.t.:

$$D_t = Y_t - W_t N_t - K_t \quad (21)$$

$$Y_t = Z_t K_{t-1}^{\alpha} N_t^{1-\alpha} \quad (22)$$

$$z_t = \log Z_t = \zeta + \rho \log Z_{t-1} + \varepsilon_t \quad (23)$$

From the first-order conditions, it can be written:

$$W_t = (1 - \alpha) \frac{Y_t}{N_t} \quad (24)$$

$$\alpha M_{0,t+1} Z_{t+1} K_t^{\alpha-1} N_{t+1}^{1-\alpha} = M_{0,t} \quad (25)$$

Finally, market clearing implies that

$$Y_t = N_t C_t + K_t \quad (26)$$

$$\sum_{i \in I_t} B_{i,t} = 0, \quad \sum_{i \in I_t} \Theta_{i,t} = 1 \quad (27)$$

From (9) summing up for all individuals it can be written

$$\sum_i C_{i,t} + P_t^S \sum_i \Theta_{i,t} + P_t^B \sum_i B_{i,t} = (P_t^S + D_t) \sum_i \Theta_{i,t-1} + \sum_i B_{i,t-1} + W_t N_t \quad (28)$$

Combining (28) with (27)

$$\sum_i C_{i,t} = D_t + W_t N_t \quad (29)$$

Therefore, it can be written

$$\left. \begin{array}{l} Y_t = \sum_i C_{i,t} + K_t \\ \sum_i C_{i,t} = D_t + W_t N_t \end{array} \right\} \Rightarrow \alpha Y_t = D_t + K_t \quad (30)$$

$$\left. \begin{array}{l} \alpha Y_t = D_t + K_t \\ M_{0,t+1}^S \alpha \frac{Y_{t+1}}{K_t} - M_{0,t}^S = 0 \end{array} \right\} \Rightarrow K_t = P_t^S \quad (31)$$

To model the stock market participation, I assume that each individual,  $i$ , has some *a priori*,  $\zeta_i$ , on the parameter  $\zeta$ , with  $\zeta_i > 0$ . At each realization of  $z_t$  each individual learns about the real value of  $\zeta$  and updates his estimated value of  $\zeta_i$ . The learning process can be expressed as a Kalman filter problem. In this framework, the state equation describes the evolution of the vector of state variables that is unobserved by the consumer

$$\underbrace{\begin{bmatrix} \zeta \\ z_{t+1} \end{bmatrix}}_{\xi_{t+1}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & \rho \end{bmatrix}}_F \underbrace{\begin{bmatrix} \zeta \\ z_t \end{bmatrix}}_{\xi_t} + \underbrace{\begin{bmatrix} 0 \\ \varepsilon_{t+1} \end{bmatrix}}_{v_{t+1}} \quad (32)$$

The observational equation instead expresses the observable variables  $Y_t$  as a linear function of the underlying hidden state:

$$y_t = \ln Y_t = [0 \quad 1] \begin{bmatrix} \zeta \\ z_t \end{bmatrix} + \alpha k_{t-1} + (1 - \alpha)n_t = \mathbf{H}' \xi_t + \alpha k_{t-1} + (1 - \alpha)n_t \quad (33)$$

Individual's best forecasting of the state  $\xi_{t+1}$ , given the observation of the state at time  $t$  is:

$$\xi_{t+1}^i = E_{i,t}[\xi_{t+1}] = F \xi_{t|t}^i \quad (34)$$

I also assumed that each individual of the new generation has at each time  $t$  an *a priori* value of  $\zeta_{i,1}^t$  which is randomly distributed with mean equal to the current mean at the moment they were born.

For example:

$$\zeta_{i,1}^t \sim \log \mathcal{N}(\bar{\zeta}_{i,>1}^t, \sigma_\zeta) \quad (35)$$

where  $\bar{\zeta}_{i,>1}^t$  is the mean of  $\zeta_i$  in the moment they were born and  $\sigma_\zeta$  is the variance of the distribution of the *a priori* values of  $\zeta$ .

Notice that each individual has his own expectations on future returns of stocks. In fact, it can be written:

$$E_{i,t} \left[ \frac{P_{t+1}^S + D_{t+1}}{P_t^S} \right] = \alpha E_{i,t}[Z_{t+1}] K_t^{\alpha-1} N_{t+1}^{1-\alpha} = \alpha e^{\zeta_i} Z_t^\rho E_t[e^{\varepsilon_{t+1}}] K_t^{\alpha-1} N_{t+1}^{1-\alpha} \quad (36)$$

Therefore

$$P_t^B = \beta E_t \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} \right] = \beta E_t \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} \right] \left. \vphantom{P_t^B} \right\} \Rightarrow \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] > \beta E_{j,t} \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (37)$$

Let us suppose now that

$$P_t^S = \beta E_{j,t} \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (38)$$

This implies that

$$P_t^S < \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (39)$$

This means that individual  $i$  sees arbitrage opportunities and tries to buy as many stocks as he can.

This makes the stock price increase and therefore now

$$P_t^S > \beta E_{j,t} \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (40)$$

making individual  $j$  sell all his stocks and buy bonds. This process stops when

$$P_t^S = \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (41)$$

Where  $\frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}}$  is the stochastic discount factor for the most optimistic individual. That is

$$\hat{i} = \{j: \zeta_j = \max \zeta_i\} \quad (42)$$

Thus, only the most optimistic individuals participate to the stock market.

The model introduces three sources of variability in individuals' behaviour. Individuals are in fact different along these three dimensions:

- **Endowment.** Each individual has different endowment and therefore face different budget constraints;
- **Age.** Individuals of different age have different time horizon and therefore different consumption and investment profile;
- **Estimated value of  $\zeta$ .** Each individual has in general different estimate of  $\zeta$  and each individual estimate change over the time according with experienced prices over individual's lifetime. It must be noticed that all stock market participants share the same estimate value of  $\zeta$ ; non participants instead have in general different estimates which are all less than participants one.

## Section 4. Predictions of the model

Recalling the first order conditions for the consumers

$$1 = \beta E_t \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} R_{t+1}^S \right] \quad (14)$$

$$1 = \beta E_t \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} R_{i,t+1}^f \right] \quad (16)$$

It can be noticed that stocks are priced only by stockholders' intertemporal marginal rate of substitution; while bonds are determined by the intertemporal marginal rate of substitution of all agents. This result is similar to the one in (William, 1977). In his model, in fact the usual security market line is modified by the replacement of true mean returns on securities with the consensus in the market about mean returns.

It is worth noticed that, according to the model, the interest rate depends *coeteris paribus* from the percentage of stock market participants and the distribution of wealth. The greater, in fact, is the percentage of stock-market participants, the less is the cost of the capital born by each stockholder, the less each stockholder needs to get into debt and therefore the lower is the interest rate. On the contrary, the less is the percentage of stockholders the greater is the cost born by each stockholder, the more each stockholder needs to get into debt and therefore the higher is the interest rate.

For the same motive, wealth distribution is relevant. The richer are the stockholders, in fact, the less they need to get into debt to buy the shares and therefore the lower is the interest rate. On the contrary, the less wealthy are the stockholders the more they need to get into debt to buy the share and therefore the higher is the interest rate.

As said, I assumed that each period  $N$  new individual are introduced in the economy and each of them have an *a priori* value  $\zeta_i$  that is random distributed with mean equal to the mean of the current *a priori* values present in the economy at that time. This implies that the greater is the number of new individuals, the higher is the price of capital. The greater in fact is the number of individuals the higher is the probability of high values of  $\zeta_i$  to be extracted.

The new individuals amplify the price effect of a series of positive or negative idiosyncratic shocks. A series of positive shocks, in fact, let all individuals update upward their estimates of  $\zeta$ , with the only exception of some too much optimistic individuals<sup>22</sup>. Thus the average value of  $\zeta_i$  rises and the new individuals have on average greater values of  $\zeta_i$ . Hence, it is more likely to have higher values of  $\zeta_i$  and therefore higher stock prices. A series of negative shock generates the opposite dynamic.

---

<sup>22</sup> Those are the ones that have such an higher *a priori* of  $\zeta_i$  that even a series of positive shocks cannot be rationalized with their value of  $\zeta_i$ .

Coherently with the results of (Chen, 1986), assets' risk consists of three components: the market consensus of volatility risk, the market consensus of the risk induced by changes in the investment opportunity set, and risk associated with uncertain shifts in investors' subjective expectations

I have assumed  $N$  to be constant over the time. However, assuming  $N$  to vary, this implies that *coeteris paribus* an higher increase in capital is to be observed in the years in which an higher increase in the number of individuals is registered. This can be more clearly seen if one adds another simplification assumption. Assuming joint conditional lognormality and homoskedasticity of asset returns and consumption, from the Euler equation it can be written:

$$E_t[\ln R_{t+1}^S] + E_t[\ln M_{t+1}] + \frac{1}{2}(var_t R_{t+1}^S + var_t M_{t+1} + 2cov_t[R_{t+1}^S, M_{t+1}]) \quad (43)$$

$$E_t[\ln R_{t+1}^S] = -\ln \beta + \gamma E_t[\Delta c_{t+1}] - \frac{1}{2}(var_t R_{t+1}^S + \gamma^2 var_t c_{t+1} - 2\gamma cov_t[\ln R_{t+1}^S, c_{t+1}]) \quad (44)$$

While from the Euler equation of the risk-free asset it can be written

$$\ln R_{t+1}^f = -\ln \beta + \gamma E_t[\Delta c_{t+1}] - \frac{\gamma^2}{2} var_t c_{t+1} \quad (45)$$

Combining (44) and (45):

$$\begin{aligned} E_{i,t}[\ln R_{t+1}^S] &= \ln R_{t+1}^f - \frac{1}{2}(var_t \ln R_{t+1}^S - 2\gamma cov_t[\ln R_{t+1}^S, c_{t+1}]) \\ &= \ln R_{t+1}^f - \frac{1}{2}(var_t \ln R_{t+1}^S - 2\gamma var_t \ln R_{t+1}^S var_t c_{t+1} corr_t[\ln R_{t+1}^S, c_{t+1}]) \end{aligned} \quad (46)$$

Notice that

$$\ln R_{t+1}^S = \ln \alpha + z_{t+1} - (1 - \alpha)(k_t - n_t) = \ln \alpha + \zeta + \rho z_t - (1 - \alpha)(k_t - n_t) + \varepsilon_{t+1} \quad (47)$$

Therefore

$$\ln \alpha + \zeta_{i,t} + \rho z_t - (1 - \alpha)(k_t - n_t) = \ln R_{t+1}^f - \frac{1}{2}(\sigma_\varepsilon^2 - 2\gamma\sigma_\varepsilon^2 \text{var}_t c_{t+1} \text{corr}_t[\ln R_{t+1}^S, c_{t+1}]) \quad (48)$$

If one assumes that  $\sigma_c^2 = \text{var}_t c_{t+1}$  and  $\rho_{r^S, c} = \text{corr}_t[\ln R_{t+1}^S, c_{t+1}]$  are constant over the time

$$\zeta_{i,t} = \ln R_{t+1}^f - \ln \alpha - \rho z_t + (1 - \alpha)(k_t - n_{t+1}) - \frac{1}{2}(\sigma_c^2 - \rho_{r^S, c}) \quad (49)$$

Hence

$$\zeta_{i,t} - \zeta_{i,t-1} = \ln R_{t+1}^f - \ln R_t^f - \rho(z_t - z_{t-1}) + (1 - \alpha)(k_t - k_{t-1} - n_{t+1} + n_t) \quad (50)$$

Therefore, an increase of  $n$  implies, *coeteris paribus*, an increase in  $\ln R^f$  and  $k$ .

This last result gives an extra feature of the model. Using the data of the risk-free rate, production and capital it is possible to measure the optimism of the stock market investors<sup>23</sup>. To do this, I have used the data on output of major industry group<sup>24</sup>, on all employees from the private sector<sup>25</sup>, on the Current-Cost Net Stock of Private Fixed Assets for the private sector<sup>26</sup> and on three months US

<sup>23</sup> In this case the term optimism is more appropriate than mood or sentiment, which are the most used terms in the literature. This index, in fact, captures only the mood of the most optimistic individuals instead of the average mood of all individuals

<sup>24</sup> Series G17/IP\_MAJOR\_INDUSTRY\_GROUPS/IP.B50001.S from the Board of Governors of the Federal Reserve System website

<sup>25</sup> Series CES0500000001 from the Bureau of Labor Statistics website

<sup>26</sup> Table 3.1ES from the Bureau of Economic Analysis website

Treasury bills on the secondary market<sup>27</sup>. I estimated the production function in (22) and (23) and using the relation in (50) I estimated the variation in the optimism index of the stock-market investors. It is worth to notice that the estimates of  $\zeta_{i,t}$  may refer to different individuals. Moreover only the variation of  $\zeta_{i,t}$  can be estimated using the available data. Therefore, in the following figures it was assumed that the absolute value at the first period is equal to zero. The results are shown in Figure 20- Figure 33

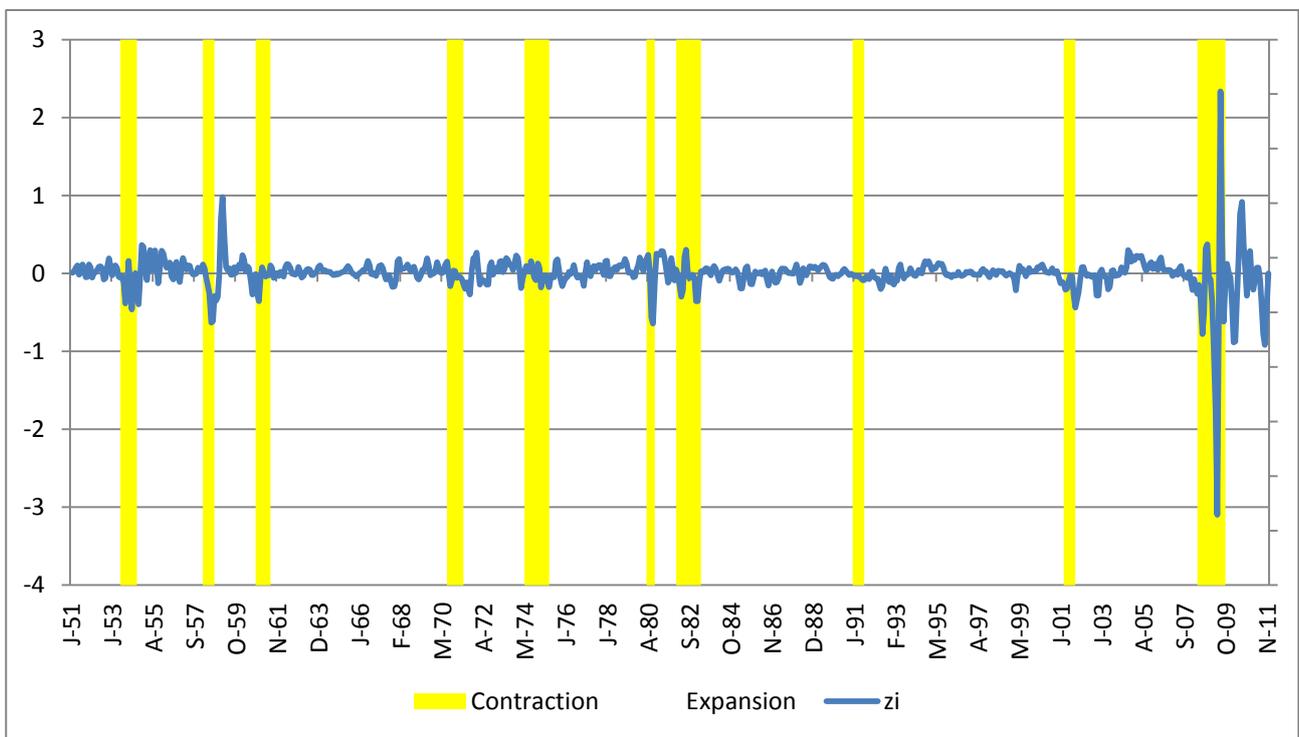


Figure 20 – Optimism index. Years 1951-2011. In yellow the official NBER contraction periods

<sup>27</sup> Series H15/H15/RIFSGFSM03\_N,M from the Board of Governors of the Federal Reserve System website

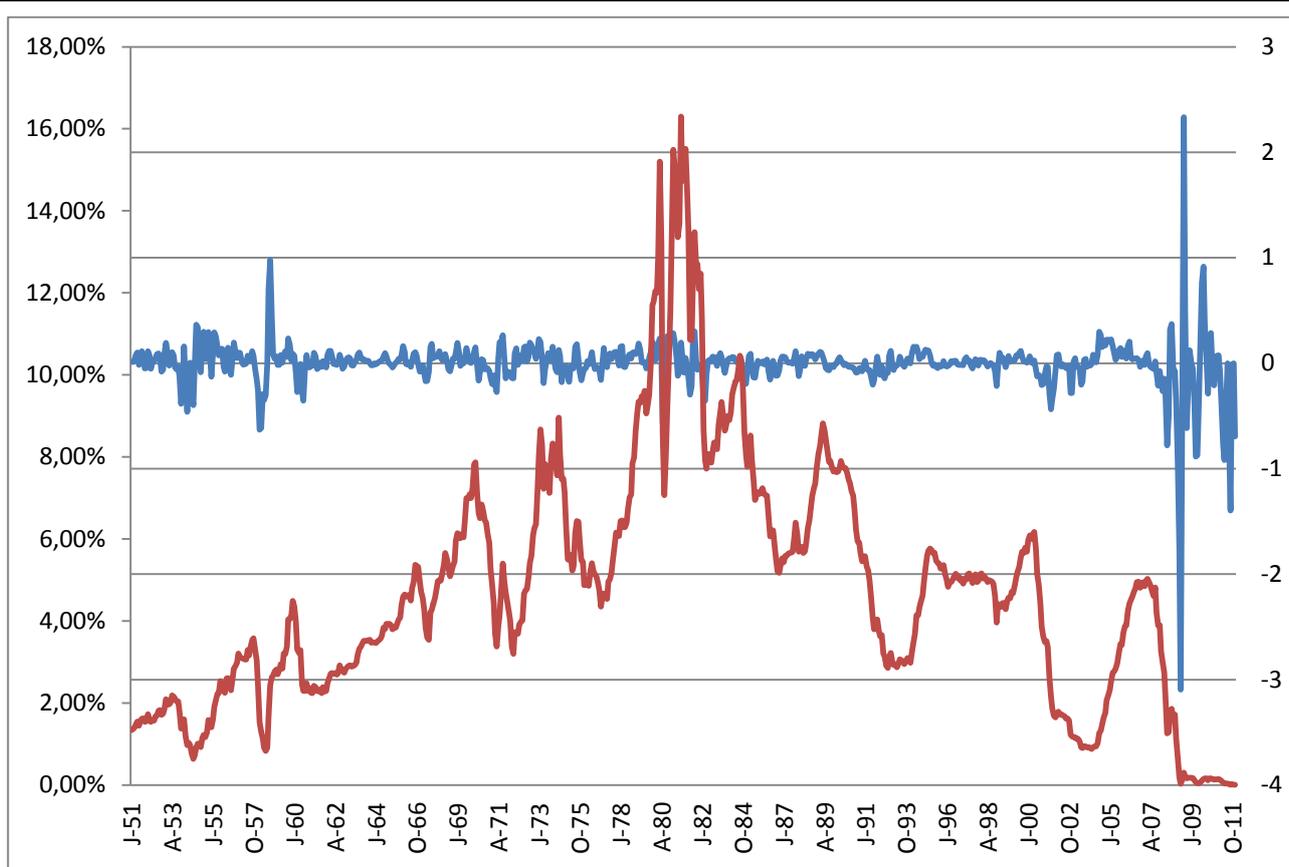


Figure 21 – Optimism index (in blue) and risk-free rate (in red). Years 1951-2011

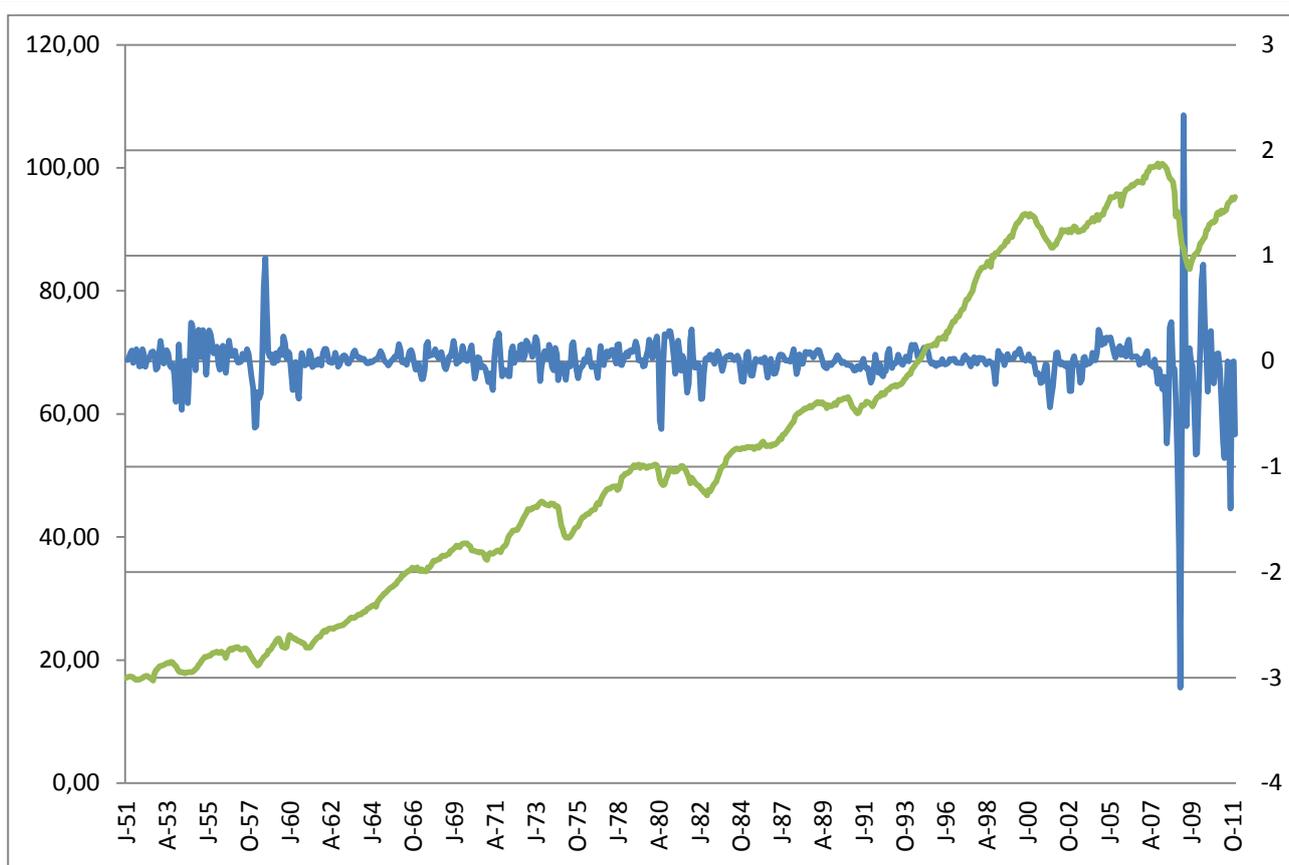


Figure 22 – Optimism index (in blue) and industrial production (in green). Years 1951-2011

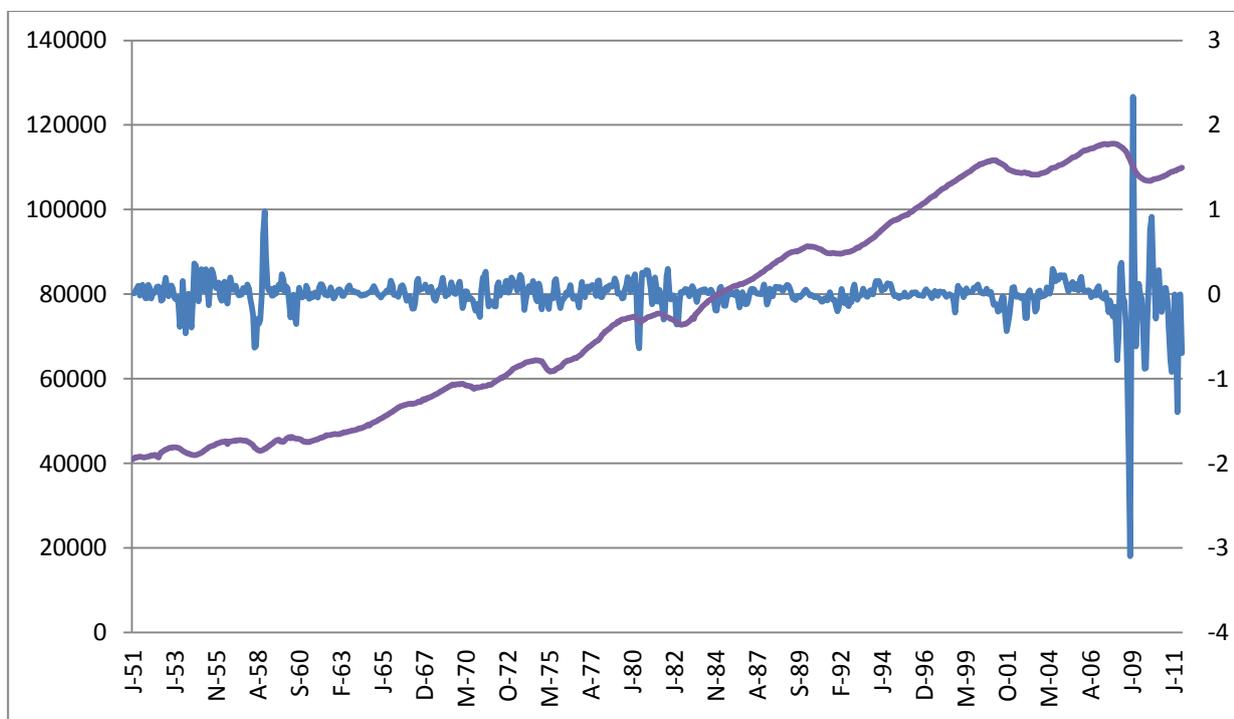


Figure 23 – Optimism index (in blue) and number of employees in the private sector (in purple). Years 1951-2011

The relevance of the optimism index is even more evident if one compares its dynamic with the variation of the relevant macro variables.

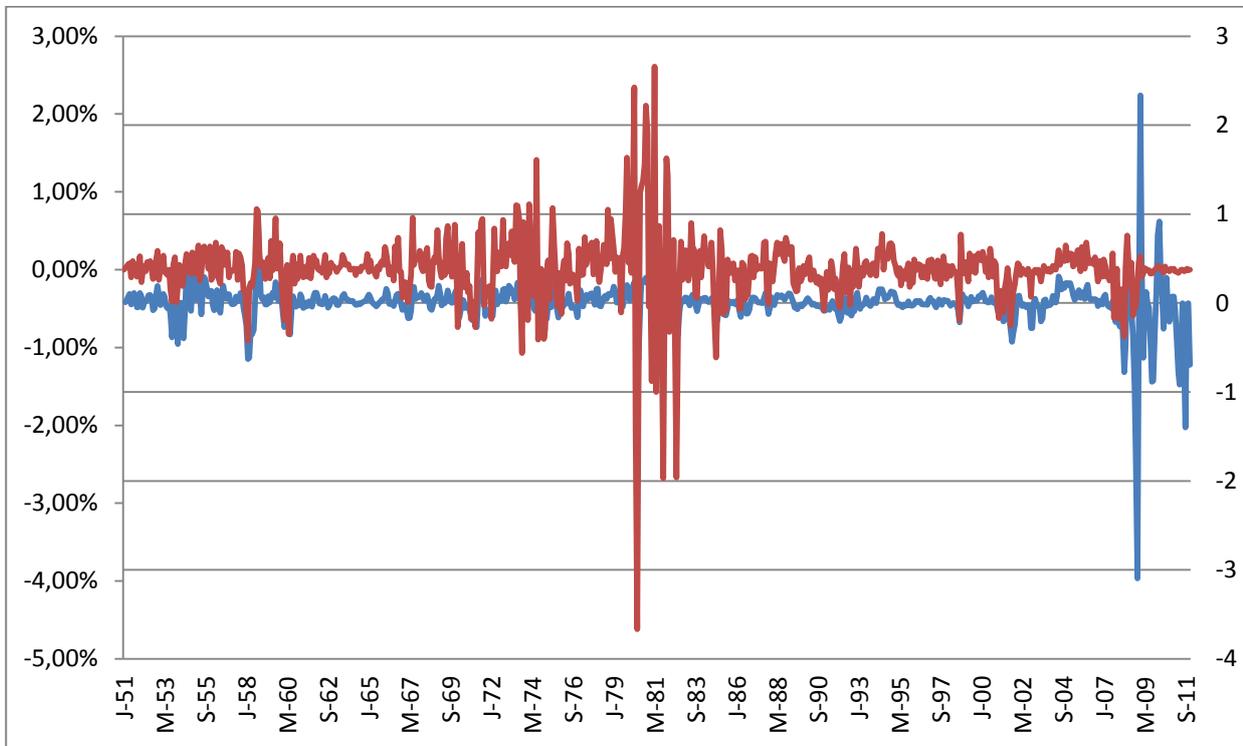


Figure 24 – Optimism index (in blue) and variation of the risk free rate (in red). Years 1951-2011

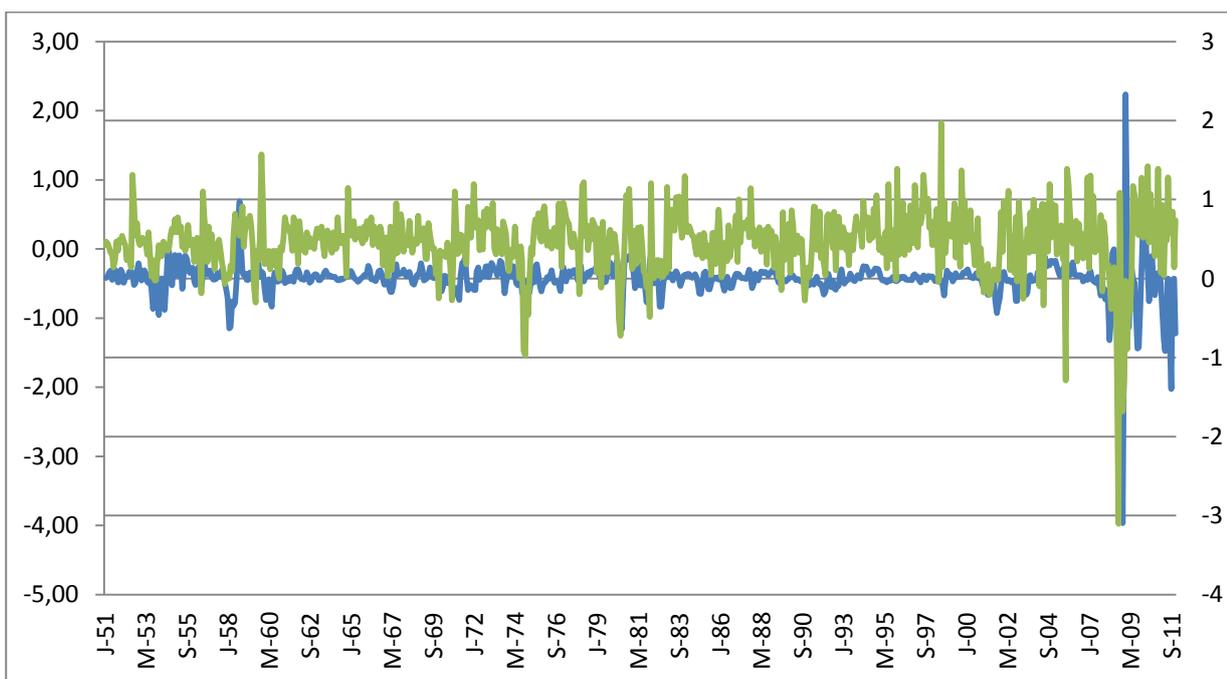


Figure 25 – Optimism index (in blue) and variation of the industrial production (in green). Years 1951-2011

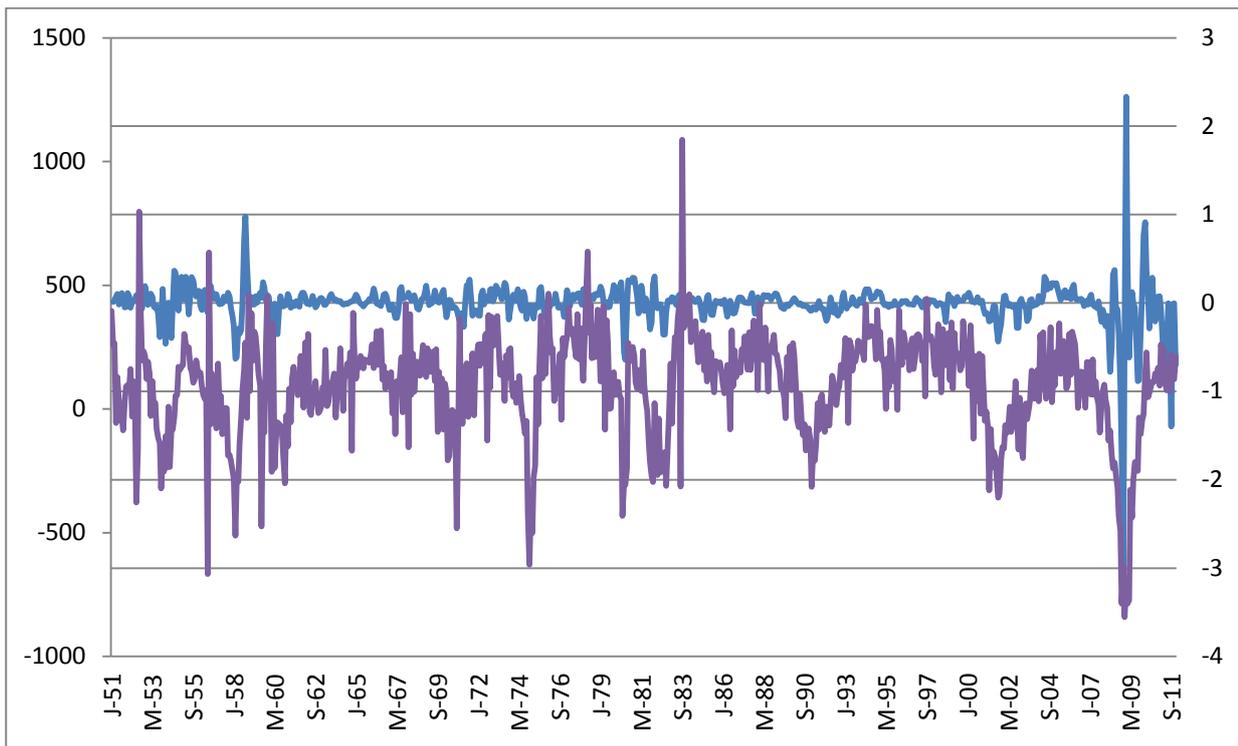


Figure 26 – Optimism index (in red) and variation of the employees in the private sector (in purple). Years 1951-2011

If one focuses on the last five years, it is possible to compare the optimism index with the most relevant macro variables during the last financial crisis. Years 1951-2011.

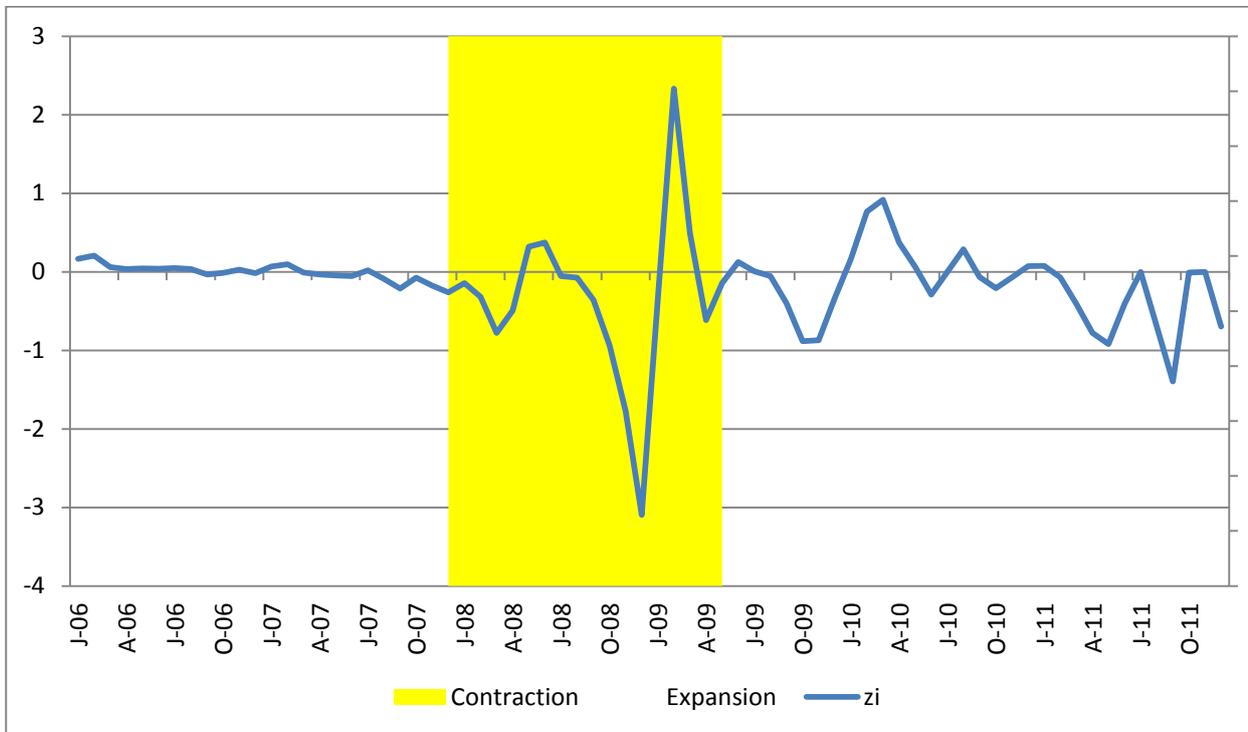


Figure 27 – Optimism index. Years 2006-2011. In yellow the official NBER expansion contraction period

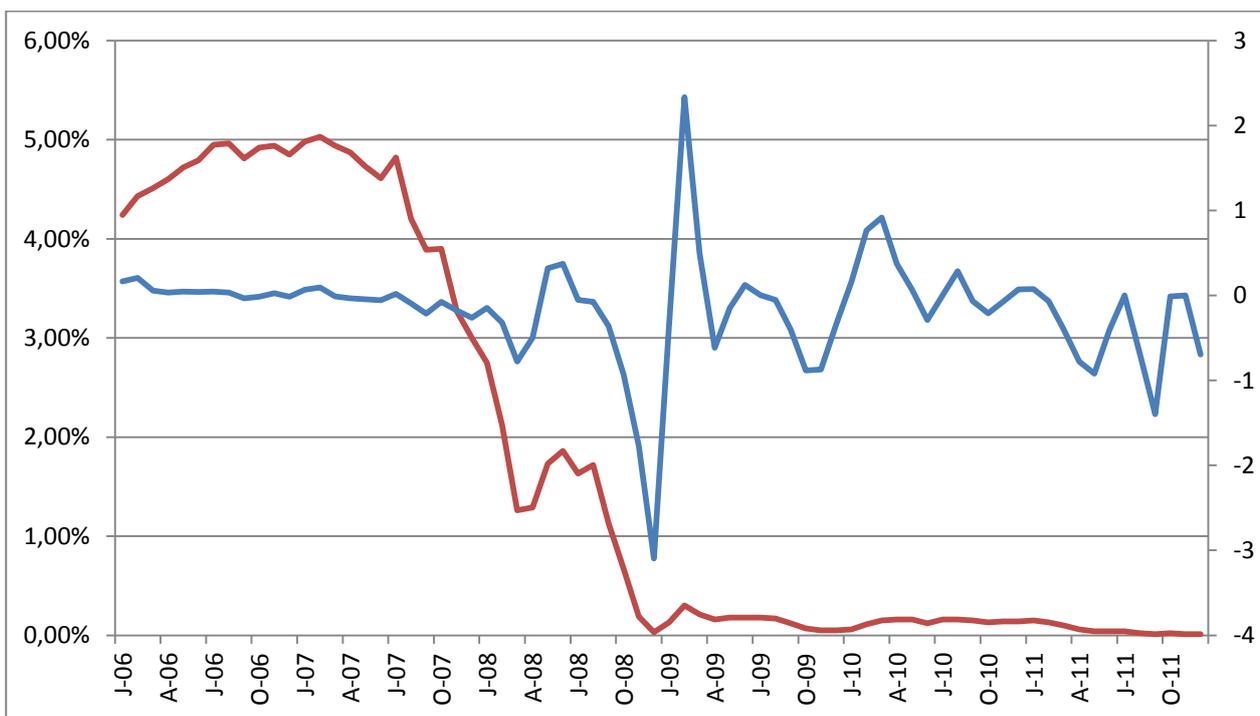


Figure 28– Optimism index (in blue) and risk free rate (in red). Years 2006-2011



Figure 29 – Optimism index (in blue) and industrial production (in green). Years 2006-2011

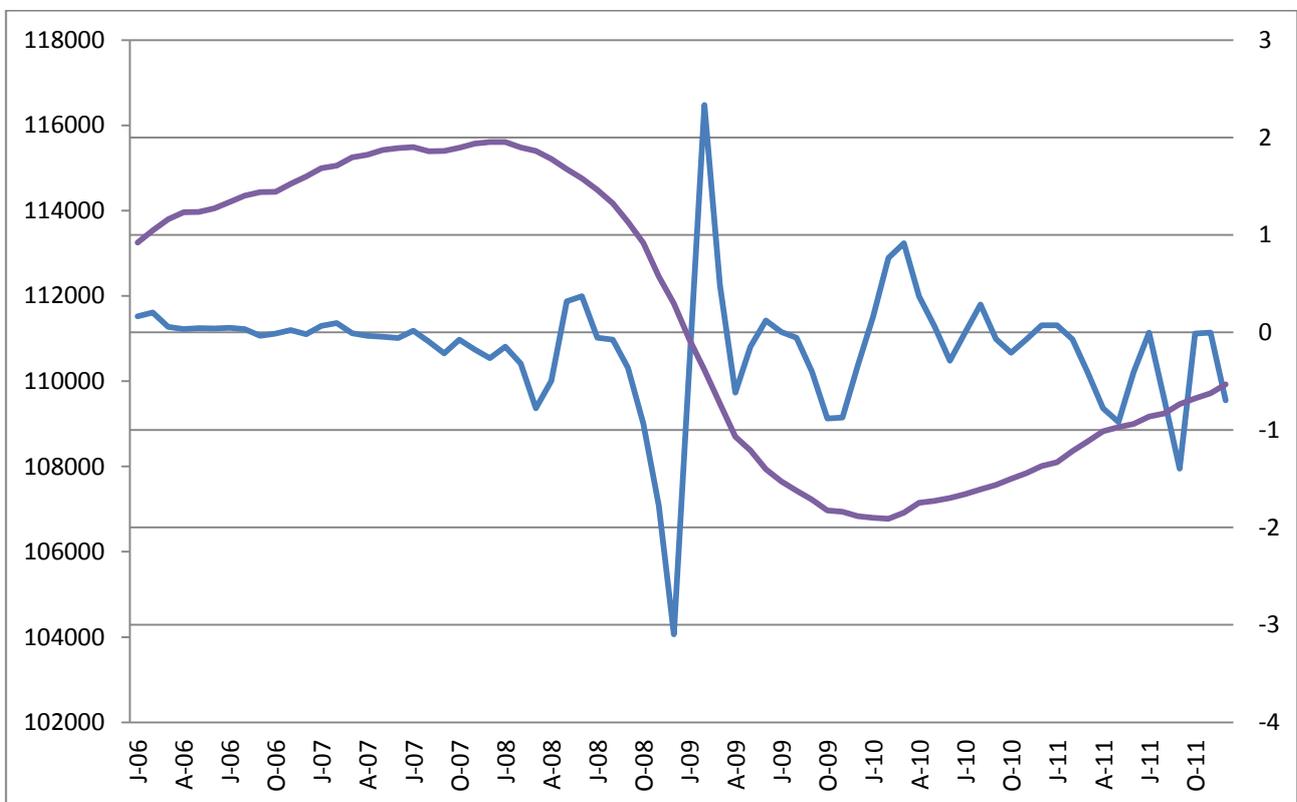


Figure 30 – Optimism index (in blue) and number of employees in the private sector (in purple). Years 2006-2011

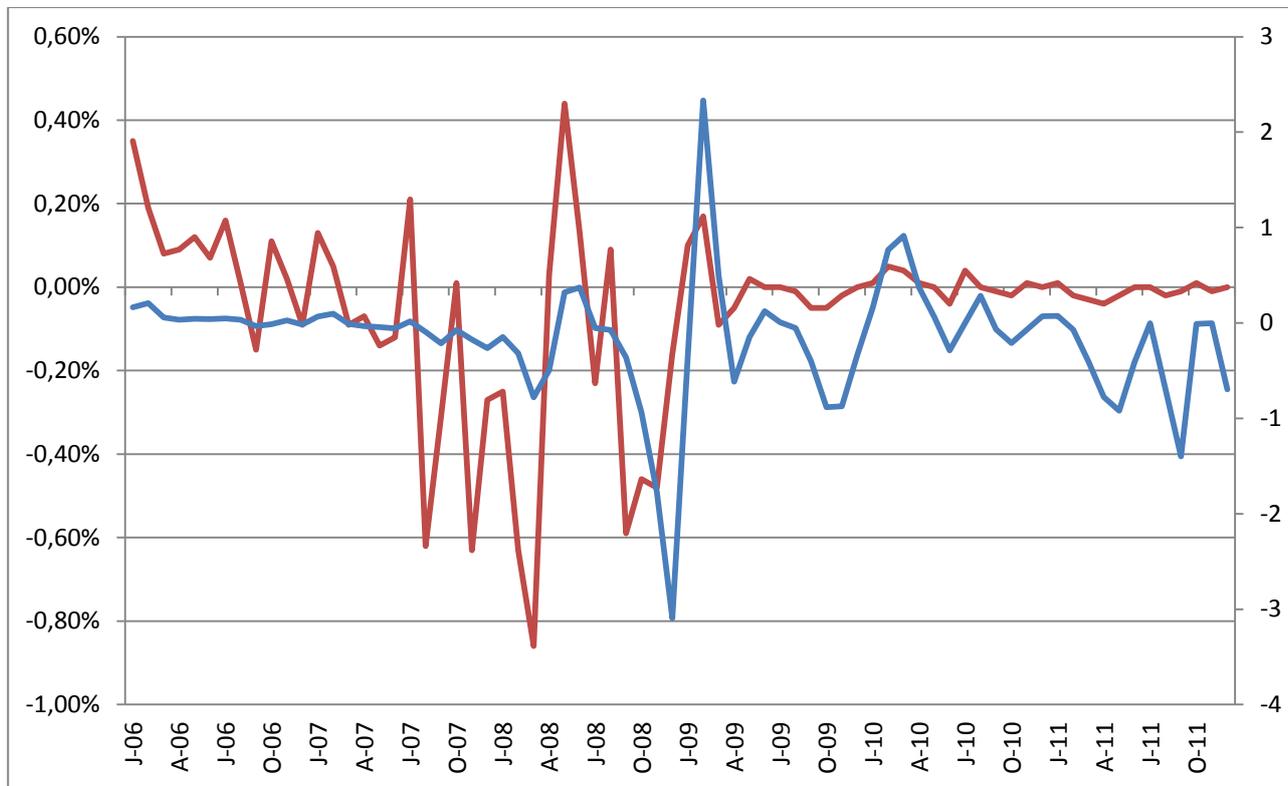


Figure 31 – Optimism index (in blue) and variation of the risk free rate (in red). Years 2006-2011



Figure 32 – Optimism index (in blue) and variation of the industrial production (in green). Years 2006-2011



Figure 33 – Optimism index (in red) and variation of the employees in the private sector (in purple). Years 2006-2011

Industrial production and occupation seem to be more reactive to the optimism index than the risk-free rate. Moreover, it is worth to notice that the fall in the optimism index during all the year 2008 is anticipated by a period of negative outcomes of the industrial production and, in the same way, the following peak, is anticipated by the peak in the industrial production. The variability in the risk-free rate and in the industrial production in the following period is almost mirrored by the variability of the optimism index. The occupation instead seems to be less reactive and with a couple of month of retard.

According to this analysis, the financial crisis had its first effects on the production side. The significant effect on the production side negatively affected investors' expectations. This change in investors' expectations worsened the effect on the financial crisis. The following period is

characterized by contradictory signals and therefore investors' mood goes up and down, sometimes amplifying the observed variation on the production side, sometimes smoothing them.

It is also worth notice that the periods of contraction are usually characterized by variability in the optimism index.

## Section 5. Caveats

A first critique may be that, if one assumes that  $\zeta_i$  can take any positive value with the cardinality of the continuum, this implies that only one individual among  $N$  participates to the stock market. This problem can be overcome in two ways: the first is by introducing more than one risky asset; the second is by introducing a discretization filter.

The first approach complicates significantly the model as it is also necessary to introduce more than one firm and consequently some hypothesis on the real market structure. This can be an interesting direction to improve the model but exceeds the scope of the present paper. It is worth notice that eventually this would not change the condition of stock market participation for the individuals. In fact, the introduction of several firms would not change the validity of the reasoning previously made. Let us consider the generic firm  $q$ , the expression of the individual expectation on future returns on stocks remains unchanged:

$$E_{i,t} \left[ \frac{P_{q,t+1} + D_{q,t+1}}{P_{q,t}} \right] = \alpha E_{i,t} [Z_{q,t+1}] K_{q,t}^{\alpha-1} N_{q,t+1}^{1-\alpha} = \alpha e^{\zeta_{q,i}} Z_{q,t}^{\rho} E_t [e^{\varepsilon_{t+1}}] K_{q,t}^{\alpha-1} N_{q,t+1}^{1-\alpha} \quad (51)$$

As before, it can be written:

$$P_t^B = \beta E_t \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} \right] = \beta E_t \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} \right] \left. \vphantom{P_t^B} \right\} \Rightarrow \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{q,t+1} + D_{q,t+1}) \right] > \beta E_{j,t} \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} (P_{q,t+1} + D_{q,t+1}) \right] \quad (52)$$

In the same way, if one supposes that

$$P_{q,t} = \beta E_{j,t} \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} (P_{q,t+1} + D_{q,t+1}) \right] \quad (53)$$

This implies that

$$P_{q,t} < \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{q,t+1} + D_{q,t+1}) \right] \quad (54)$$

This, again, means that individual  $i$  sees arbitrage opportunities and tries to buy as many stocks as he can. This makes the stock price of the firm  $q$  increase and therefore now

$$P_{q,t} > \beta E_{j,t} \left[ \frac{C_{j,t+1}^{-\gamma}}{C_{j,t}^{-\gamma}} (P_{q,t+1} + D_{q,t+1}) \right] \quad (55)$$

making individual  $j$  sell all his stocks and buy bonds. This process stops when

$$P_{q,t} = \beta E_{i_q,t} \left[ \frac{C_{i_q,t+1}^{-\gamma}}{C_{i_q,t}^{-\gamma}} (P_{q,t+1} + D_{q,t+1}) \right] \quad (56)$$

Where  $\frac{C_{i_q,t+1}^{-\gamma}}{C_{i_q,t}^{-\gamma}}$  is the stochastic discount factor for the most optimistic individual on firm  $q$ . That is

$$\hat{i}_q = \{j: \zeta_{q,j} = \max \zeta_{q,i}\} \quad (57)$$

The second approach instead needs only to map the continuous value of  $\zeta_i$  into a discrete set  $Z$  and attribute to individuals, instead of  $\zeta_i$ , the corresponding discrete value  $Z_i$ . This would not change the main features of the model and obtain the desired feature of having more than one individual participating to the stock market.

The non-short-sale assumption is instead more problematic. Heterogeneous beliefs on  $\zeta$ , in fact make the equilibrium in the stock market impossible without the non-short-sale assumption.

Without this assumption, in fact, individuals with higher  $\zeta_i$  would buy indefinite quantities of stocks until

$$P_t^S < \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (58)$$

as they would see arbitrage possibility.

While individuals with lower  $\zeta_i$  would short-sell indefinite quantities of stocks until

$$P_t^S > \beta E_{i,t} \left[ \frac{C_{i,t+1}^{-\gamma}}{C_{i,t}^{-\gamma}} (P_{t+1}^S + D_{t+1}) \right] \quad (59)$$

This implies that if two individuals  $i$  and  $j$  have different values of  $\zeta_i$ , there is no equilibrium stock price that can satisfy the Euler equation of both individuals. This problem can be overcome introducing borrowing constraints but it significantly complicates the model.

## **Section 6. Facts on stock market participation vs. the two approaches**

(Vissing-Jørgensen, 2002), analyzing the data on the financial portfolios of US portfolios, evidenced four facts. First, a large fraction of households does not participate in the stock market at any given point of time. Second, there is substantial turnover in the set of stock market participants. Third, within the set of participants there is a substantial heterogeneity in the share of financial wealth invested in stocks. Fourth, for many stock market participants there are large changes in the portfolio shares for equity over time.

Entry costs models can easily explain the first, the third and the fourth fact, though they need to assume variability of entry costs among individuals, whose plausibility is not sure. It is harder to

explain the third fact as it is necessary to introduce variability in the entry cost for the same individual in different periods.

The model I presented can instead easily explain the first two facts; the last two are a bit more problematic. Although from the first order conditions is clear that more than one combination of stocks and bonds are compatible with the same flow of consumption, the model does not provide any insight that may justify any change in the individual portfolio over the time or any significant difference in the combination of stocks and bonds between two stockholders. However, it must be noticed that most of the papers on asset prices that try to explain the differences in the investors' portfolios use the heterogeneous expectations approach. Reconciling the present model with those one can be thought as a possible direction of improvement.

Fixed or variable participation or transaction costs are interesting explanations of the phenomenon of non-participant households and can explain why households do not participate when they have low financial wealth, why households tend to trade very infrequently and why the participation rate is increasing with financial wealth<sup>28</sup>. Nevertheless, they do not help explaining two facts. The first is that there are many wealthy non-participating households. Why a wealthy household should renounce to a significant increase of its consumption instead of paying a small entry cost? Or, in alternative, why for these household the entry cost is so high? This is mirrored by the fact that some non-wealthy households do participate to the stock market. In this case, the question is reversed. Why a non-wealthy household should pay a fixed cost so high to discourage wealthier households to have a lower increase in consumption? Or, alternatively, why for these households the entry cost is so low? Most of the papers that use the entry cost approach explain this phenomenon with differences in the risk aversion among the households. This is in part coherent with the estimates of (Vissing-Jørgensen, 2002), and (Vissing-Jørgensen & Attanasio, 2003) but does not explain the

---

<sup>28</sup> See (Malloy, Moskowitz, & Vissing-Jørgensen, 2005)

second fact. That is that some households participate to the stock market one in year but not the other and eventually participate again in the following years. (Vissing-Jørgensen, 2002), in fact, evidences that, comparing the shares of financial wealth held in stocks in the years 1989 and 1994, 27.6% of the households pass from participants to non-participants or vice-versa.

The present model, instead, can easily explain why many households trade very infrequently and why there may be wealthy non-shareholders households and non-wealthy shareholders participants. The frequency of trading, in fact, depends on the changes in individuals beliefs and therefore ultimately on the learning process. Until the generic individual  $i$  remains the most optimistic, he keeps on holding stocks. We would observe trading when his learning process and/or the access of new investors makes him shift from the most optimistic position to a lower optimistic rank, with respect to the remaining individuals. In the model there are not clear insights of changes in the portfolio composition but it is not incompatible with changes in the portfolio composition within the choice of stock market participation. The fact that some individuals participate one year but not the following one can be easily explained by some new consumers who in that time has a random generated *a priori* value of  $\zeta_i$  greater than the previous individual's one. The phenomenon of younger investors having more optimistic expectations was observed in (Cowgill, Wolfers, & Zitzewitz, 2008). It is much harder to explain why the first investor may participate to the stock market again the year after. This needs some unlikely combination of random generation of  $\zeta_i$ s and of idiosyncratic shocks  $\varepsilon_t$  that make some individual stay in the threshold of the participation choice. Instead the present model has no clear explanation of why the participation rate is increasing with financial wealth.

Many papers using the entry cost approach make interesting analysis on the importance of take information on the stock market and try to quantify the cost of it, though they do not introduce any differences in the set of information among individuals. It is worth notice that some authors instead

evidenced that individuals belonging to the same group of population tend to have similar trading strategies. For example, (Grinblatt, Keloharju, & Linnainmaa, 2009) show that, controlling for wealth, income, and other demographic and occupational information, there is a high correlation between IQ and stock market participation and that, *coeteris paribus*, high IQ investors are more likely to hold mutual funds and larger numbers of stocks. (van Rooij, Lusardi, & Alessie, 2007), (Cole & Shastry, 2009) and (Guiso & Jappelli, 2005), among the others, similarly find an association between financial literacy and stock market participation.

Other authors go even further, studying the importance of the information sets of the investors in defining their investing strategies. (Brown, Ivković, Smith, & Weisbenner, 2008), for example, examine empirically the influence of “community effects,” in the form of word-of-mouth communication, on the decision of whether to participate in the stock market or not. They found that there is a relation between individual and community choice of market participation. The authors analyzed this phenomenon again in the entry cost point of view. They assumed that there may be some kind of spill-over effect that reduce the individual entry costs when the community he belongs to participates to the stock market. Though this could be the case, I am more prone to think that people belonging to the same community rather share the same beliefs on future outcomes of the economic variables. I think that this gives a better explanation of the phenomenon and better explains why people of the same community, but with different wealth and income, may choose to have similar participation behaviours with no need to assume significant differences in the entry costs among individuals.

This point of view is confirmed by (Cowgill, Wolfers, & Zitzewitz, 2008) who studied the data from a corporate experiment conducted by Google with prediction markets among its employees. They found that there are strong correlations in trading among those who sit within a few feet of one another and for those who belong to the same social networks or have work relationships. The

fact that social proximity seems to be relevant not only for the participation behaviour but also for the trading strategies is in my opinion a confirmation of the flow of information approach and of the presence of different beliefs among individuals.

Another confirmation comes from (Battalio & Mendenhall, 2005) who provide empirical evidence that different types of investors base their trading decisions on different information sets. More interestingly, they demonstrate that, even if some investors have systematically biased expectations their trading choices do have significant effect that can help explaining earning-related anomalies such as stock-prices retarding in answering to earning announcements.

This mechanism and the previous exposed flow of information are fully coherent with the model I presented. Since the stock price is determined by the most optimistic individuals, it resents of this systematic distortion. The learning process helps individual to gradually improve their estimates of the real value of  $\zeta$  and thus slowly incorporate new information on future outcomes. On the other hand, new investors are a continuous source of systematically biased expectations, which can retard the market consensus to fully represent the fundamentals of the firms or in other cases can let the market overreact to a series of positive or negative shocks. This is coherent also with (Cowgill, Wolfers, & Zitzewitz, 2008) who observed that new investors tend to have systematically biased expectation, while individuals with more experience trading in the markets were better calibrated.

For simplicity, I have assumed the same learning process for each individual. Nevertheless, the model can be easily extended to incorporate the possibility of subset of individuals having access to different information sets. This makes it possible to model the proximity effect in (Cowgill, Wolfers, & Zitzewitz, 2008).

Moreover, most of the papers using the entry costs approach need to add also variability in the risk aversion and in the entry costs among the individuals. The variability of the risk aversion among individual is, in part, coherent with the estimates made by (Attanasio, Banks, & Tanner, 2002), (Vissing-Jørgensen, Limited Asset

Market Participation and the Elasticity of Intertemporal Substitution, 2002), and (Vissing-Jørgensen & Attanasio, 2003) and with the answers of the consumer to the following question in the Survey of Consumer Finances: “Which of the statements on this page comes closest to the amount of financial risk that you are willing to take when you save or make investments?”

- (1) Take substantial financial risks expected to earn substantial returns.
- (2) Take above average financial risks expected to earn above average returns.
- (3) Take average financial risks expecting to earn average returns.
- (4) Not willing to take any financial risks.

As reported by (Malloy, Moskowitz, & Vissing-Jørgensen, 2005), averaging the numbers corresponding to the answers of the stockholders, in fact, one obtains a number which is much less than averaging the numbers corresponding to the answers of the non-stockholders. Although it is not fully correct to average ordinal non-metric variables, this seems to be a clear sign of differences in the risk aversion among individuals. Nevertheless, the answers (1) and (2) do not seem incoherent with the choice of participate to the stock market in the heterogeneous beliefs environment. It is particularly interesting the fact that the question asks whether the investors take substantial risk to earn substantial returns or take above the average risk to earn above the average returns. This seems to confirm the idea that stock market investors tend to have higher returns expectations than non-stockholders and, because of that, they are willing to take more risk. In the present model, in fact, stockholders do take more financial risk than non-stockholders as they invest a good part of their wealth in stocks<sup>29</sup>. Non-stockholders instead do not bear any other risk than the idiosyncratic one.

---

<sup>29</sup> Notice that the amount of wealth stockholders invest in stocks in the model depends on the percentage of stockholders among the individuals

The hypothesis of different risk aversion among the individuals can be tested comparing the consumption data of stockholders with those of non-stockholders. From the Euler equation for bonds it can be written

$$P_t^B = \beta E_t \left[ \frac{C_{t+1}^{-\gamma}}{C_t^{-\gamma}} \right] \quad (60)$$

This equation, in order to be estimated, needs the conditional expectation of  $C_{t+1}^{-\gamma}/C_t^{-\gamma}$  to be known. Unfortunately this distribution cannot be known by the econometrician. It is instead possible to estimate the unconditional version. Therefore, assuming as in (Hansen & Singleton, 1983) and (Vissing-Jørgensen, 1999) that the Euler equation can be log-linearized it can be written

$$\ln R_{t+1}^f = -\ln \beta + \gamma E_t[\Delta c_{t+1}] - \frac{\gamma^2}{2} \text{var}_t c_{t+1} + \epsilon_{t+1} \quad (61)$$

This equation holds both for stockholders and non-stockholders. If stockholders had different risk aversion than non-stockholders the estimated values of  $\gamma$  for stockholders would be different from the non-stockholders one. To do this, I have used the data from the PSID<sup>30</sup>. Unfortunately, in the PSID only food consumption is recorded, then I must assume that the remaining consumptions follow the same dynamic. In the years before 1999 data were collected every year. Therefore, it is possible estimate the following non linear regression model

$$\ln R_{t+1}^f = \alpha_0 + \alpha_1 (\ln C_{t+1} - \ln C_t) - \frac{\gamma^2}{2} \text{var}_t [\ln C_{t+1} - \ln C_t] + \epsilon_{t+1} \quad (62)$$

<sup>30</sup> Panel Study of Income Dynamics, public use dataset. Produced and distributed by the Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI.

Where, for the rate of return of the risk-free asset, I used the three months Treasury bill. I estimated, instead, the variance of the variation of the log-consumption from the observed values. The result of the regression does not show any significant difference in the estimates of the parameter  $\gamma$  for stockholders and non-stockholders. Nevertheless, the result does not seem it can effectively be used as evidence against the hypothesis of different values of  $\gamma$  for stockholders and non-stockholders. The regression has, in fact, very low goodness of fit.

## **Section 7. Conclusions**

The problem of stock market non-participation can be considered of recent interest, though it is by itself a contradiction of the standard models of financial economics. It started being an appealing topic when (Mankiw & Zeldes, 1991) first understood that the differences between stockholders and non-stockholders could help explain the equity premium puzzle.

The most frequent explanation of this phenomenon has been the presence of entry costs to the stock-market. This is the most intuitive and immediate way to model, though it presents some problems. Entry costs in fact do not help to explain some observed facts on stock market participation. In fact it is not easy to rationalize with these models why there are many wealthy non-participating households and some non-wealthy households do participate to the stock market; why there is substantial turnover in the set of stock market participants; why some individuals participate to the stock market intermittently and the social proximity effects on participating and trading choices of individuals. Most of the model based on entry costs need to add assumptions on differences in the risk aversion among the households and differences in the entry costs among individuals to give a partial and not entirely satisfying explanation of these phenomena.

A model based on heterogeneous beliefs can give an easy explanation of different stock market participation behaviour and accounts for most of the facts observed about stock market

participation, though some facts still remain unexplained. The model in fact needs to assume non-short-sales on stocks, it does not have clear explanations of the intermittent participation of some individuals and why the participation rate is increasing with financial wealth. This approach is however very flexible and is compatible with different models of portfolio composition and gives many possibility of extensions and developments.

## Bibliography

- Abarbanell, J. S., Lanen, W. N., & Verrecchia, R. E. (1995). Analysts' forecasts as proxies for investor beliefs in empirical research. *Journal of Accounting and Economics* .
- Abel, A. B. (1990). Asset prices under heterogeneous beliefs: Implications for the equity premium. *mimeo* . Wharton School, University of Pennsylvania.
- Alan, S. (2006). Entry Costs and Stock Market Participation Over the Life. *Review of Economic Dynamics* .
- Attanasio, O. P., Banks, J., & Tanner, S. (2002). Asset holding and consumption volatility. *Journal of political economy* .
- Basak, S. (2005). Asset pricing with heterogeneous beliefs. *Journal of banking and finance* .
- Basak, S., & Cuoco, D. (1998). An Equilibrium Model with Restricted Stock Market Participation. *The Review of Financial Studies* .
- Battalio, R. H., & Mendenhall, R. R. (2005). Earnings expectations, investor trade size, and anomalous returns around earnings announcements. *Journal of Financial Economics* .
- Brown, J. R., Ivković, Z., Smith, P. A., & Weisbenner, S. (2008). Neighbors Matter: Causal Community Effects and Stock Market Participation. *The Journal of finance* .
- Cao, H. H., Wang, T., & Zhang, H. H. (2005). Model Uncertainty, Limited Market Participation, and Asset Prices. *The Review of Financial Studies* .
- Carceles-Poveda, E., & Giannitsarou, C. (2008). Asset pricing with adaptive learning. *Review of Economic Dynamics* .
- Chen, S.-N. (1986). An Intertemporal Capital Asset Pricing Model Under Heterogeneous Beliefs. *Journal of Economics and Business* .
- Cole, S., & Shastry, G. K. (2009). Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation. *Working paper* . Harvard Business School.
- Cova, J. D., & Moskowitz, T. J. (1999). Home Bias at Home: Local Equity Preference in Domestic Portfolios. *The Journal of Finance* .
- Cowgill, B., Wolfers, J., & Zitzewitz, E. (2008). Using Prediction Markets to Track Information Flows: Evidence from Google. *mimeo* .
- Favilukis, J. (2006). Wealth Inequality, Stock Market Participation, and the Equity Premium. *mimeo* . The London School of Economics and Political Science.
- French, K., & Poterba, J. (1991). Investor diversification and international equity markets. *American Economic Review* .
- Gallmeyer, M., & Hollifield, B. (2008). An Examination of Heterogeneous Beliefs with a Short-Sale Constraint in a Dynamic Economy. *Review of Finance* .
- Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2009). IQ and Stock Market Participation. *mimeo* . University of California at Los Angeles.

- Guiso, L., & Jappelli, T. (2005). Awareness and Stock Market Participation. *Review of Finance* .
- Haliassos, M., & Bertaut, C. C. (1995). Why do so Few Hold Stocks? *The Economic Journal* .
- Hansen, L. P., & Singleton, K. J. (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *The Journal of Political Economy* .
- He, H., & Pearson, N. D. (1991). Consumption and Portfolio Policies with incomplete Markets and Short-Sale Constraints: The Infinite Dimensional Case. *Journal of Economic Theory* .
- Luttmer, E. (1999). What Level of Fixed Costs Can Reconcile Consumption and Stock Returns? *Journal of Political Economy* .
- Malloy, C. J., Moskowitz, T. J., & Vissing-Jørgensen, A. (2005). Human Capital Risk, Stockholder Consumption, and Asset Returns. *Journal of Finance* .
- Mankiw, N. G., & Zeldes, S. P. (1991). The consumption of stockholders and non-stockholders. *Journal of Financial Economics* .
- Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics* .
- Paiella, M. (2007). The Forgone Gains of Incomplete Portfolios. *The Review of Financial Studies* .
- van Rooij, M., Lusardi, A., & Alessie, R. (2007). Financial Literacy and Stock Market Participation. *NBER working paper* .
- Vissing-Jørgensen, A. (2002). Limited Asset Market Participation and the Elasticity of Intertemporal Substitution. *Journal of Political Economy* .
- Vissing-Jørgensen, A. (1999). Limited Stock Market Participation and the Equity Premium Puzzle. *Working paper* . Department of Economics, University of Chicago.
- Vissing-Jørgensen, A. (2002). Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures. *mimeo* . University of Chicago.
- Vissing-Jørgensen, A., & Attanasio, O. P. (2003). Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion. *The American Economic Review* .
- William, J. T. (1977). Capital asset prices with heterogeneous beliefs. *Journal of financial economics* .