Limited Commitment to Enforcement and the Evolution of Tax Compliance.

Marcello D'Amato*, Riccardo Martina**, Marisa Ratto***

Work in Progress. Very preliminary draft.

Please do not quote.

February 1, 2016

Abstract

We study a model of tax evasion dynamics with a local enforcement externality due to limited commitment by the state in the allocation of resources to fiscal juridictions. Taxpayers rationally decide their compliance behavior given they do not have precise information about the probability of apprehension, but learn it over time following a simple social learning mechanism, based on their experience.

We analyse the conditions under which the distribution of the perceived probability of apprehension and compliance behavior converge to one or multiple steady states and whether convergence features history dependence. We show that the emergence of steady state multiplicity crucially depends on the level of tax evasion compared to a measure of efficiency in the enforcement process.

The results are used to explore the long run consequences of a fiscal union in cases of equilibrium multiplicity, i.e. when historical initial conditions operate as a selection criterion. As an illustration of the long lasting consequences of initial conditions, we document persistent heterogeneity of tax compliance in Italy, suggesting an interpretation of the so called Mezzogiorno problem in terms of institutional traps, as an alternative to other interpretations based on heterogeneity in the endowment of social capital or other structural characteristics.

Keywords: tax evasion, local enforcement externality, learning, compliance dynamics.

JEL: D62, D81, H26, K41, K42.

* University of Salerno , ** University of Naples Federico II , *** University Paris-Dauphine, corresponding author: maria_luisa.ratto@dauphine.fr

"As rates rise to over 20 or 25 percent,
the income tax becomes destructive,
taxpayer compliance breaks down
and enforcement fails".
J.A. Schumpeter, Policy Essays, 1926-32.
Quoted in Musgrave R.A. (1992), p. 95.

1 Introduction

This paper investigates the dynamic implications of the enforcement externality in tax administration and compliance. The source of this externality arises at the level of a jurisdiction, where a larger level of non compliance, in the absence of a compensating investment in additional resources, increases the number of cases to deal with, thereby reducing the effectiveness of enforcement and diluting deterrence.

More generally, one of the major issue for enforcement authorities in modern states is that detection of frauds or crimes does not automatically translate into conviction of offenders. In case of tax evasion or other economic crimes this implies important revenue losses for the government, which come both from non collected tax revenues and also from lower future perceptions of being caught, which make offending more profitable on the margin. A recent report from the OECD on a comparison of tax administration regimes in OECD and non-OECD countries[1] highlights that countries which experience higher tax evasion rates also report an unusually large inventory of disputed tax debt[2].

As it is well known (Becker 1968, Ehrlich 1972, Votey and Phillips 1972) there is a fundamental complementarity, i.e. an externality, between the output of enforcement and the level of offenses, in the presence of limited resources to allocate to local enforcement. Such a complementarity may lead to multiplicity in equilibrium behavior by offenders and in the equilibrium level of deterrence. In a dynamic environment, equilibrium multiplicity can explain historical persistence and, when the externality arises at the local level, persistent geographical dispersion.

As in other branches of jurisdiction, the administration of fiscal law is enforced by local courts (see, for the case of Italy, Manestra, 2010, Alessio 1883) so the externality features a local dimension, in that the aggregate amount of evasion can affect the actual enforcement at the local level, in the absence of commitment by the state to homogeneous deterrence in the territory. So it may also be at the root of policy traps consistent with heterogenous outcomes at the geographical level. The reason why we focus on this specific source is that we think it is particularly relevant for the interpretation of the persistently ob-

[1] OECD(2015)

[2] According to the report, Italy stands out of all OECD and non-OECD countries for having the highest share of assessed taxes and penalties relative to net revenue collections and also the highest inventory of tax debt (defined as total amount of tax, including interest and penalties that is overdue for payment at the end of each fiscal period).

served geographical dispersion of compliance behavior across different regions in the fiscal history of Italy.

Although the focus of this study is on the jurisdiction, the enforcement externality may have different micro-foundations, some of them common to other realms of the public enforcement (e.g. criminal law), some more specific to the context under investigation, some of them more likely to occur at the country level, others more important to explain the local dispersion of compliance rates. This externality could in principle originate:

- in the enforcement technology that exhibits decreasing returns (in terms of apprehension, number of administrative procedures and finalized trials) with respect to the aggregate size of the evasion rate; (Ehrlich 1973, Sah 1991)
- in the size of the fiscal budget available to enforcement authorities. For a given percentage of tax revenues allocated to the enforcement agencies more evasion would entail less resources for the fiscal budget to finance the enforcement activities;
- as an important variation of the two previous sources, in the contractual arrangements provided to tax officials. Indeed, past larger evasion, by curtailing resources available to pay wages to tax officials, allows incentives for corruption to emerge and hence tax evasion may increase;
- in the local political economy: tax evasion represents a form of redistribution. Indeed, specific structural aspects affect opportunities and constraints of individual taxpayers so that preferences for a more or less aggressive enforcement policy may differ in the electorate. In a democracy this may provide incentives to politicians to compete on the actual size of resources committed to fiscal law enforcement. Similar forces can be at work at the local level where local tax enforcers are affected by local politics. Hence, initially different attitudes towards the fiscal obligation get perpetuated when the enforcement is local;[3]
- in the sociological structure of a given economy as in models based on stigma where the emergence of specific social norms, or forms of social capital involving more or less compliance, can arise as a consequence of more or less social constraints faced by local enforcers. (Traxler 2010)

Indeed, all these potential sources of externality point in the same direction: the aggregate level of enforcement in a given country or district depends on the aggregate level of compliance and they imply that a larger degree of evasion produces a lower level of enforcement and hence larger incentives to evasion.

[3] On the political economy of local fiscal administration see Giulio Alessio (1883), who already highlights the importance of the fact that when Fiscal Courts are elected at the local level (by the municipalities as in Italy at the time he wrote and, as we will document, for most of the 150 years of existence of the Italian state) they rather tend to preserve the interests of local elites, no matter if industrial capitalists or land owners. See in particular Alessio (1883) p. 352-353.

In this paper we assume that either because of budget constraints and the inability to run public debt or because of limited state capacity inherited from an unexplored political economy structure, there is limited scope for commitment to an equilibrium (or optimal) level of enforcement: hence the externality between aggregate evasion and enforcement is relevant. We characterize the conditions under which structural elements of the fiscal system can make this externality relevant in the long run and the fiscal system history dependent.

To this aim we consider a simple model where agents, given their information, rationally make the evasion choice facing the uncertain consequences of their act and they learn, over time, about the specific features of the enforcement structure. We show that the learning rule produces persistence of tax evasion. Our main result is that, when the tax rate is low enough compared to a measure of efficiency of enforcement, this persistence settles down to a low evasion equilibrium; for intermediate values of the tax rate, the high evasion equilibrium becomes part of the sets of steady states and the fiscal state will converge there, depending on initial perceptions. When the tax rate is sufficiently high, the unique equilibrium selected by history is a high evasion one, where "taxpayer compliance breaks down and enforcement fails". When arising from local externalities in the enforcement, equilibrium multiplicity can explain geographical dispersion and divergent patterns of behavior among otherwise identical taxpayers. In general this dispersion can be attributed to long lasting consequences of initial differences in some state variable that influences fiscal compliance.

Multiple equilibria make the history of fiscal compliance a relevant dimension of investigation and represent a natural framework for analyzing important aspects in the dynamics of the fiscal equilibrium of a state. In particular they allow to justify permanent differences (across countries or across local economies within the boundary of a state nation) in the level of fiscal compliance observed in politico-economic units that would be otherwise identical in any other respect.

A second important aspect of the analysis of fiscal compliance is that taxpayers (especially self-employed and entrepreneurs) seem to invest a lot of money in the acquisition of information through the hiring of experts in order to minimize the tax burden by consulting experts. In other words learning about the costs of evasion induced by the tax code (fines) and by the enforcement system (probability of apprehension and penalty administration) is a relevant dimension. [4]

We use the results in our model to study the effects on a few aspects:

i) we consider the implications of a more or less noisy signal obtained by

[4] Learning is not the unique source of dynamics one can imagine. Tax Revenues and the allocation of the budget to the enforcement system may require lags involved by tax collection procedures; delays in monitoring and judiciary activities also impose a dynamics in the revenues pertaining to several fiscal years; exogenous structural dynamics (business cycle and a changing sectoral composition) also involve dynamics in the tax compliance and hence enforcement. To keep the model at its simplest possible level we will ignore these sources of dynamics. Notice that these alternative sources of dynamics are, at the same time, among the reasons why the real probability of apprehension (i.e. the resources invested by the fiscal authority in the enforcement process) is not known to the tax payers and commitment (in any optimal mechanism governing the fiscal obligation) by the fiscal authority is prevented.

taxpayers about the actual enforcement (as induced by the amount of information provided by experts, or the noise produced by the fiscal authority itself by making non-credible announcements or reforms). The presence of experts, in a model with learning can induce the presence of multiple equilibria (driven by the fundamental externality in the individual tax evasion choice) through the influence of experts on priors belief.

ii) we provide an example of the relevance of these issues to analyze the effects of political and hence fiscal unification.

iii) we then use results and implications of our theoretical model to provide a unified interpretation of a series of historical episodes occurred in Italy after unification.

The link between aggregate crime and law enforcement has already been recognized in the crime literature. Ehrlich (1973) considers a static model of the optimal individual allocation of time between legal and illegal activities and introduces the assumption that the productivity of the resources allocated to law enforcement is lower the higher the level of criminal activity. This is because "...more offenders must then be apprehended, charged and tried in court in order to achieve a given level of P [probability of punishment]."[5] This idea is also modelled in Sah (1991) who examines how crime participation rates might evolve over time in a framework where the actual probability of punishment decreases with the aggregate crime rate and individuals hold perceptions of the probability of being punished based on their past experience. Our model is closely related to Sah (1991): we assume that the aggregate level of tax evasion creates a negative externality on the returns from audits and taxpayers are not fully informed about the probability of being detected and fined, but they learn it over time by following a simple learning mechanism. We analyze the conditions for the existence and convergence of multiple equilibria, an issue that is set aside in Sah (1991), who focuses the analysis on a stable interior equilibrium. As in Sah (1991), we show that the learning mechanism and the negative externality of aggregate tax evasion on actual enforcement provide a possible explanation for the persistence of local heterogeneity of tax evasion. Different geographical locations may end up with different levels of tax evasion, even in the presence of similar fundamentals, simply because past values of tax evasion were higher (and in our model this could be due to higher tax rates) in the past and this led to lower perceptions of being fined, which in turn affected future behaviour towards non-compliance.

Galbiati and Zanella (2012) also consider a model with tax enforcement congestion. In particular, they assume that the individual probability of detection decreases in individual reported income and that the local budget constraint of the enforcement agency cannot be relaxed promptly. In their model perceptions correspond to actual probabilities of punishment, hence there is no role for learning. Moreover, the authors consider a static model with a unique equilibrium, due to the assumption of linear probabilities and linear-in-means individual reaction functions. Persistence is explained in terms of fixed auditing resources. The importance of individual perceptions (which differ from

[5] See Ehrlich (1973), p. 540.

the actual probability of punishment) for crime participation has been shown by Lochner (2007). In line with this, we choose to follow Sah (1991) and distinguish between perceived probability of detection and actual probability of detection. While Sah (1991) adopts a very general formulation for the learning mechanism, we restrict our attention to a simple learning (myopic) mechanism in order to simplify the analysis of the dynamics of perceptions and of tax compliance and of equilibrium selection, which is not developed in Sah (1991).

The paper is organized as follows: in Section 2 we explain our assumption about the enforcement externality and provide some empirical evidence from Italy to motivate it. Section 3 presents the model and the results. Section 4 interprets the north/south divide of fiscal compliance in Italy since its unification in the light of our model implications. Section 5 illustrates the implications for a unification process. Section 6 concludes.

2 Empirical Motivation of the assumption about the enforcement externality

There are two main tasks of the enforcement activity: audits that lead to detection and the actual collection of unpaid taxes. This second task is made more difficult for the enforcement authority when there is no agreement between taxpayers and the tax authority and a tax litigation process is opened. In some OECD countries, like Chile, Iceland, Sweden and Italy, the enforced tax debt collection is undertaken outside the revenue body[6]. Moreover, the judiciary branch of the tax authority, i.e. the specialized tax courts which decide on the tax litigation cases, is in most countries organized at local level. We see two possible ways of modelling the enforcement externality: either on the detection process or on the judiciary activity dealing with tax disputes. We choose the second option and assume that the externality works on the judiciary branch of the enforcement authority. We now provide some evidence from data obtained by the Italian Revenue Agency, which we interpret as an empirical foundation of our assumption.

We gathered information on the *weighted worked hours*[7] and on *equivalent output*[8] for two specific processes (audits and litigations) and for the overall activity (all processes) of the agency, at the level of provincial directorates (111) over the period 2006-2010. We also have data on the net returns from audits and litigations and an overall rate of compliance by Italian provincial directorates, over the same period. The indexes of the returns from audits and litigations are defined as IRN (index of net returns to audits) and Invicto (index of successful finalized trials) respectively. IRN is the ratio of additional taxes,

[6] See OECD (2015), p. 63.

[7] These are the number of hours worked on a given process (e.g. audits) weighted by a parameter that considers the hourly remuneration of the staff involved in that process. This is to take into account the quality of the work and not only the total number of hours worked.

[8] Equivalent output is the actual product of a given process multiplied by the average time requested to produce one unit of that item. Hence production is measured in terms of working time and is then comparable across different production processes, like audits, litigations and other administrative work. See Alborino et al. (2008) for further details.

interests and sanctions which are expected to be collected[9] by the agency, net of audit costs, to additional taxes, interests and sanctions actually assessed by audits. Invicto is the ratio between the amount of money which has been ruled out in favour of the Revenue Agency from finalized trials and the total amount of money in litigation from finalized trials. The compliance rate (ITC) is calculated by dividing the amount of spontaneously paid taxes by the sum of spontaneously paid taxes and the tax gap[10]. We first consider the distribution of tax compliance across regions. Table 1 shows the mean, the median and the standard deviation (across provinces and time) of the index of tax compliance for each Italian Regional Directorate of the Agency over the period 2006-2010.

ITC	mean	median	Std Dev
Piemonte	0,7368	0,7157	0,0795
Valle d'Aosta	0,7897	0,7916	0,0202
Liguria	0,7629	0,7788	0,0803
Lombardia	0,8357	0,8391	0,0291
Bolzano	0,8441	0,8457	0,0055
Trento	0,8339	0,8332	0,0039
Veneto	0,7622	0,7778	0,0502
Friuli	0,8047	0,8048	0,0487
Emilia Romagna	0,8084	0,8123	0,0590
Marche	0,6991	0,6786	0,0530
Toscana	0,7358	0,7391	0,0746
Umbria	0,7033	0,7025	0,0181
Lazio	0,7605	0,6979	0,1133
Campania	0,6446	0,6121	0,0710
Abruzzo	0,7625	0,7941	0,0657
Molise	0,6022	0,6026	0,0416
Puglia	0,6526	0,6702	0,0753
Basilicata	0,5381	0,5357	0,0467
Calabria	0,5506	0,5466	0,0792
Sicilia	0,6044	0,6410	0,0998
Sardegna	0,6783	0,6661	0,0775

Table 1: Mean, median and standard deviation of the Index of Tax Compliance (ITC) by Italian Regional Directorates in the period 2006-2010.

[9] These are calculated by summing up the amount of unpaid taxes that are collected in agreement with the taxpayers and the estimated amount of unpaid taxes which will be collected after tax litigations. For this latter, the estimates are based on past years results from litigations and inside knowledge of the tax authority. Informal conversations with the tax officials revealed that on average it takes three years to collect undisputed unpaid taxes, whereas it takes ten years to collect unpaid taxes that go through tax litigations.

[10] This latter is calculated using the top-down approach, based on the comparison between tax data and National Accounts figures. These include an estimate of the underground economy, and hence provide an indicator of the "potential " tax base. From this potential base an estimate of the potential collection is then derived, through which it is possible to calculate the tax gap, defined as the missing portion of the tax potential. See Pisani (2014) for more details.

There is a large variation across regions, and regions in the south of Italy (Campania, Molise, Puglia, Basilicata, Calabria, Sicilia and Sardegna) tend to show much lower compliance rates.The average compliance rate for Italy in the period (not shown in the table) is 72.77%, ranging from 53.81% for the regional directorate of Basilicata to 84.44% for the regional directorate of Bolzano.

It is interesting to see if such variation is also found for the use of enforcement resources. In Table 2 we consider the weighted worked hours and the equivalent output for audits. In order to control for the difference in size of the provincial directorates, we consider weighted hours and equivalent output of the audit process relative to the weighted hours and equivalent output of the overall activity of a given provincial directorate and then calculate the mean, average and standard deviation at regional level. We define these variables as RelInput_Audits and RelOutput_Audits respectively.Table 2 shows the results.

Rel Input Audit	mean	median	Std Dev	Rel Output Audit	mean	median	Std Dev
Piemonte	0,4775	0,4785	0,0150	Piemonte	0,4876	0,4822	0,0386
Valle d'Aosta	0,4255	0,4192	0,0296	Valle d'Aosta	0,3693	0,3636	0,0281
Liguria	0,4794	0,4781	0,0244	Liguria	0,3934	0,4027	0,0343
Lombardia	0,4577	0,4562	0,0281	Lombardia	0,4341	0,4269	0,0341
Bolzano	0,4043	0,4072	0,0213	Bolzano	0,3600	0,3676	0,0308
Trento	0,4610	0,4651	0,0175	Trento	0,4286	0,4216	0,0224
Veneto	0,4651	0,4681	0,0201	Veneto	0,4432	0,4452	0,0267
Friuli	0,4697	0,4752	0,0287	Friuli	0,4654	0,4615	0,0330
Emilia Romagna	0,4787	0,4826	0,0189	Emilia Romagna	0,4377	0,4427	0,0321
Marche	0,4674	0,4622	0,0273	Marche	0,4334	0,4328	0,0228
Toscana	0,4840	0,4832	0,0158	Toscana	0,4717	0,4702	0,0289
Umbria	0,4708	0,4743	0,0264	Umbria	0,4558	0,4437	0,0308
Lazio	0,4614	0,4648	0,0230	Lazio	0,4496	0,4503	0,0250
Campania	0,4392	0,4435	0,0287	Campania	0,4577	0,4635	0,0452
Abruzzo	0,4522	0,4494	0,0233	Abruzzo	0,5315	0,5393	0,0474
Molise	0,4405	0,4488	0,0278	Molise	0,5249	0,5213	0,0377
Puglia	0,4355	0,4370	0,0329	Puglia	0,4394	0,4407	0,0665
Basilicata	0,4320	0,4453	0,0480	Basilicata	0,4545	0,4567	0,0220
Calabria	0,4406	0,4428	0,0143	Calabria	0,4646	0,4580	0,0515
Sicilia	0,3933	0,3975	0,0321	Sicilia	0,4805	0,4822	0,0542
Sardegna	0,4266	0,4306	0,0204	Sardegna	0,4589	0,4671	0,0280

Table 2: Mean, median and standard deviation of the weighted hours worked on audits relative to total weighted hours worked and of equivalent output from audits relative to total equivalent output by Italian Regional Directorates in the period 2006-2010

We do not find such a striking difference between southern regions and the rest of the country as for the compliance rate, although the weighted hours worked on audits relative to the overall activity are lower in the Southern regions, with Sicilia showing the lowest values. For the relative output, we do not find a clear divide between southern regions and rest of the country and actually some of the northern regions like Liguria and Lombardia have a lower percentage of normalized output relative to the total activity than southern regions, like Sicilia or Calabria. Hence for the allocation of inputs and outputs to the audit process we do not find the same geographical variation we find for the compliance index and we get a more homogeneous picture across regions.

We now consider the effectiveness of audits and litigations. Table 3 shows the mean, median and standard deviation of the two indexes. For the index of net returns from audits (IRN), the mean for Italy is 5% (not shown in the table), hence a pretty low average net return of audits. Southern regions show lower values for the whole distribution. For example, on average around 13.5% of additional taxes and fines, net of audit costs is expected to be collected in Valle d'Aosta and Trento, whereas this value is very close to zero for Sicilia. The index of success of litigations (Invicto) is the highest for Valle d'Aosta. Sicilia shows the lowest mean and median of the distribution, respectively 52% and 53%. There is a considerable variation across regions and the southern regions exhibit lower values of the index of success of litigations, although the difference between southern regions and rest of the country is not so striking as for the index of net returns of audits.

IRN	mean	median	Std Dev	Invicto	mean	median	Std Dev
Piemonte	0,0826	0,0715	0,0422	Piemonte	0,7417	0,7542	0,1642
Valle d'Aosta	0,1349	0,1296	0,0614	Valle d'Aosta	0,8226	0,8724	0,1532
Liguria	0,0574	0,0540	0,0261	Liguria	0,6388	0,6815	0,1768
Lombardia	0,0521	0,0463	0,0373	Lombardia	0,6419	0,6882	0,1796
Bolzano	0,0863	0,0766	0,0262	Bolzano	0,6613	0,6844	0,1513
Trento	0,1358	0,1485	0,0037	Trento	0,7533	0,8096	0,1637
Veneto	0,0694	0,0700	0,0323	Veneto	0,7101	0,7298	0,1638
Friuli	0,0597	0,0611	0,0289	Friuli	0,6279	0,6266	0,2130
Emilia Romagna	0,0650	0,0584	0,0324	Emilia Romagna	0,7778	0,7937	0,1732
Marche	0,0502	0,0434	0,0272	Marche	0,7537	0,7459	0,1327
Toscana	0,0987	0,0935	0,0592	Toscana	0,6948	0,7205	0,1866
Umbria	0,0624	0,0671	0,0172	Umbria	0,7972	0,8053	0,1054
Lazio	0,0203	0,0185	0,0192	Lazio	0,6094	0,6332	0,1517
Campania	0,0165	0,0149	0,0179	Campania	0,5900	0,6034	0,1281
Abruzzo	0,0432	0,0419	0,0191	Abruzzo	0,6776	0,7054	0,1860
Molise	0,0107	0,0083	0,0149	Molise	0,5770	0,5444	0,1909
Puglia	0,0209	0,0215	0,0182	Puglia	0,5496	0,5948	0,2007
Basilicata	0,0297	0,0344	0,0259	Basilicata	0,6735	0,6771	0,1371
Calabria	0,0171	0,0165	0,0209	Calabria	0,6996	0,7195	0,1820
Sicilia	0,0079	0,0131	0,0195	Sicilia	0,5224	0,5256	0,2004
Sardegna	0,0257	0,0226	0,0385	Sardegna	0,6835	0,7019	0,1859

Table 3: Mean, median and standard deviation of the Index of Net Returns of audits (IRN) and of the index of successful finalized trials (Invicto) by Italian Regional Directorates in the period 2006-2010.

Next we check if those regions exhibiting lower compliance rates than the rest of the country also exhibit lower effectiveness of enforcement. In order to do this, we regress the three indicators (ITC, IRN and Invicto) against a constant term and a dummy variable for the region[11]. We run one regression for each region. In each regression the coefficient of the constant term represents the mean of the dependent variable for the regions other than the one considered by the dummy variable. The sign of the coefficient of the dummy variable tells us if the provincial directorates in the region under consideration exhibit a higher or a lower mean value of the dependent variable than the provincial directorates in the other regions. The unit of observation is the provincial directorate over the period 2006-2010. Hence we have 555 observations. Table 4 shows only the

[11] This method allows all observations to have a different variance by calculating robust standard errors. An alternative method is to use a test for equality of means, but this restricts the variance of each group of observations to be the same. See Cameron and Trivedi (2010), p.80.

results which were statistically significant for those regions with a mean value significantly lower than the mean of the rest of the regions. It is clear that most of the regions that exhibit a lower compliance than the mean of the other regions are also those (apart from Umbria and Marche) which exhibit a lower net return from audits. We get less significant results for the index of successes of litigation, but, again, the regions which have a statistically significant lower mean index of success of litigations are those having a statistically significant lower net return from audits and a statistically significant lower mean of the compliance rate (except for Lazio).

Regions below the mean of the rest of the country											
	ITC				IRN				Invicto		
	dummy coeff	constant	obs		dummy coeff	constant	obs		dummy coeff	constant	obs
Umbria	-0,0249***	0,7282***	555	Lazio	-0,0318***	0,052***	555	Lazio	-0,0597**	0,6691***	555
	(0,0071)	(0,0046)			(0,0038)	(0,002)			(0,0267)	(0,0083)	
Marche	-0,0299***	0,7291***	555	Molise	-0,04***	0,0508***	555	Campania	-0,0797***	0,6697***	555
	(0,0071)	(0,0047)			(0,0049)	(0,0019)			(0,0245)	(0,0083)	
Molise	-0,1279***	0,73***	555	Campania	-0,0355***	0,05199***	555	Puglia	-0,1224***	0,672***	555
	(0,0133)	(0,0045)			(0,0038)	(0,002)			(0,03699)	(0,0081)	
Campania	-0,0879***	0,7324***	555	Puglia	-0,0308***	0,05174***	555	Sicilia	-0,1556***	0,678***	555
	(0,0136)	(0,0046)			(0,0038)	(0,002)			(0,0307)	(0,0080)	
Puglia	-0,0794***	0,732***	555	Basilicata	-0,0208***	0,0505***	555				
	(0,0143)	(0,0046)			(0,0080)	(0,0019)					
Basilicata	-0,1931***	0,7312***	555	Calabria	-0,0346***	0,05163***	555				
	(0,0147)	(0,0044)			(0,0045)	(0,0019)					
Calabria	-0,1855***	0,736***	555	Sicilia	-0,04585***	0,0538***	555				
	(0,0161)	(0,0043)			(0,0034)	(0,002)					
Sicilia	-0,0944***	0,7354***	555	Sardegna	-0,0253***	0,05098***	555				
	(0,0154)	(0,0046)			(0,0086)	(0,0019)					
Sardegna	-0,0512**	0,7265***	555								
	(0,024)	(0,0046)									
Robust standard errors in paranthesis			***p<0,01, **p<0,05, *p<0;1								

Table 4: OLS regressions testing the equality of means of ITC, IRN and Invicto across regions.

In conclusion, we find large differences across regional directorates in the compliance rate and in the returns from audits and litigations. These differences are not fully matched by differences in the allocation of worked hours and of the volume of normalized production for the audit process[12]. Moreover, most of regional directorates having a (statistically significant) lower compliance rate also have lower net returns from audits and lower returns to litigations. These results are not in contradiction to the assumption of our theoretical model that a greater (local) level of non compliance is linked to lower returns from the enforcement process.

The five years time interval of data we gathered (it was not possible to get data before 2006, as these indexes are available only starting in 2006), do not allow us to have an idea of the persistence of the phenomenon. However, for the index of tax compliance we can compare our results to previous findings for Italy. Pisani and Polito (2008) consider data on IRAP[13] evasion in the period 1998-2002 and find a large variation of tax compliance across regions, with a clear divide between southern regions and the rest of the country: the lowest

[12] Similar results were found for the litigation process. Results are available from the authors upon request.

[13] The Italian regional tax levied on productive activities.

evasion rate is found for Lombardia (13%) and the highest for Calabria (94%). The authors find evidence of a large variation also across provinces in the same region: in Lombardia, the lowest evasion rate is for Milan (6%) and the highest for Lodi (34%). While in Calabria the lowest evasion rate is for Reggio Calabria (53%) and the highest for Vibo Valentia (184%!). We find a very similar picture. In table 5 we show the average compliance rate for the provincial directorates of Lombardia and Calabria. In our case Lombardia ranks the second best compliant while Calabria ranks as the second-last. We also find that Milan has the highest compliance rate within Lombardia (83%), while Lodi scores the second lowest position (79.93%), while, in Calabria, Reggio Calabria scores the second highest compliance rate (62.16%) and Vibo Valentia scores the lowest (45.6%). Galbiati and Zanella (2012) consider compliance rates for the self-employed at regional level in 1987 and find large variation across regions and a tendency for lower compliance in southern regions. Given the different time periods considered in these studies, this confirms that the heterogeneity of tax compliance at local level in Italy is a persistent phenomenon.

Index of tax compliance (ITC) in the period 2006-2010				
	mean:			mean:
Lombardia	83.57%		Calabria	55.06%
Provincial directorates				
Varese	83%		Cosenza	54.21%
Como	85.33%		Catanzaro	64.47%
Sondrio	85.81%		Reggio Calabria	62.16%
Milano I	83.79%		Crotone	48.77%
Bergamo	86.28%		Vibo Valentia	45.67%
Brescia	85%			
Pavia	78.18%			
Cremona	83%			
Mantova	82.8%			
Lecco	85.09%			
Lodi	79.93%			
Monza	84.57%			
Milano II	83.59%			

Table 5: Mean index of tax compliance for Lombardia and Calabria and their provincial directorates

3 The Model

We consider a simple model with exogenous income Y_i reported by each tax-payer based on her perception of the consequences of her evasion choice. The fiscal system (state) is made of a tax rate (τ), a probability of an audit (a), a probability of conviction for audited taxpayers (θ), an actual probability of apprehension $r = a\theta$, a fine for evasion (ϕ), an initial distribution of perceptions among the taxpayers $\mathbb{P}_0(\widehat{p}_i)$, where $\widehat{p}_i$ is the perception about r formed by taxpayer i using a learning rule l_i. The population of taxpayers i is normalized to 1. Exogenous variables are collectively denoted as $\Gamma = \{Y_i, \tau, a, \phi, \mathbb{P}_0(\widehat{p}_i), l_i\}$. The model will determine equilibrium level of enforcement r_t and its steady state level r^*, the evolution of the distribution of perceptions $\mathbb{P}_t$ and its limit distribution $\mathbb{P}^*$, the evolution of the aggregate level of evasion in the economy e_t and its steady state level e^*. The triple $\{\mathbb{P}_t(\widehat{p}_i), e_t, r_t\}$ is denoted a fiscal system with its structural (exogenous) components Γ.

Tax evasion is decided depending on the size of the tax, the perceived probability of punishment, the fine and a moral benefit from abiding to the fiscal obligation. We consider risk neutral agents endowed with a linear utility function, which also includes an additional non-monetary benefit from compliance[14].

Hence taxpayer i, perceiving a probability $\widehat{p}_i$ of apprehension, will solve the following problem:

$$\underset{\delta_i}{Max} U(\delta_i) = (1-\widehat{p}_i)Y_i(1-\tau\delta_i) + \widehat{p}_i(Y_i - \tau\delta_i Y_i - \phi\tau Y_i(1-\delta_i)) + \varepsilon_i(\delta_i) \quad (1)$$

Where $\delta_i \in [0,1]$ denotes the percentage of income reported by taxpayer i, Y_i is her income, $\widehat{p}_i$ is her perceived probability of apprehension in case of evasion, τ is the tax rate, ϕ is the fine in case tax evasion is discovered by the audit and finalized in court. We assume $\varepsilon_i(\delta_i) = 0$ for $\delta_i < 1$ and $\varepsilon_i(\delta_i) = \varepsilon_i$ for $\delta_i = 1$; moreover ε_i is uniformly distributed on $[0, Y_m]$[15]. Finally, to simplify the analysis, we assume $\phi\tau = 1$ (maximal fine).

Lemma 1 *Individual compliance is as follows:* $\delta_i = 0$ *if* $\varepsilon_i \leq \widetilde{\varepsilon}_i = (\tau - \widehat{p}_i)Y_i$ *and* $\delta_i = 1$ *otherwise.*

Proof. Risk neutrality of the taxpayer implies that the solution is at the boundary of the feasible set, i.e. $\delta_i \in \{0,1\}$. Therefore, for any given Y_i, there will exist a threshold value of $\varepsilon_i = \widetilde{\varepsilon}_i$ such that taxpayer i is indifferent between evasion and full compliance. Since $U(\delta_i = 0) = (1-\widehat{p}_i)Y_i$ and $U(\delta_i = 1) = (1-\tau)Y_i + \varepsilon_i$, it is immediate to see that $\delta_i = 0$ obtains if $\varepsilon_i \leq (\tau - \widehat{p}_i)Y_i$ and $\delta_i = 1$ otherwise ■

Hence, given a perception $\widehat{p}_i$ and income Y_i, an individual with low enough (moral) benefit from compliance will evade the whole income.

Denote the distribution of perceptions, at a given moment in time, by $\mathbb{P}_t(\widehat{p}_i)$ with average $\widehat{p}_{m,t}$ (remember that $\mathbb{P}_t$ is endogenous and must be derived at equilibrium), then the average non compliance rate in the given economy is described as follows:

[14] In the appendix we show that no major qualitative results regarding the occurrence of equilibrium multiplicity are affected by this choice.

[15] This choice for the support represents normalization as it can be seen from the following analysis.

Lemma 2 *For any given distribution of perceptions $\mathbb{P}_t(\widehat{p}_i)$ the aggregate evasion rate is*

$$e_t = \tau - \widehat{p}_{m,t}. \tag{2}$$

where $\widehat{p}_{m,t} = \int_{\widehat{p}_i=0}^{1} \widehat{p}_i d\mathbb{P}_t(\widehat{p}_i)$.

Proof. From the previous Lemma we know that the tax evasion decision is $\delta_i = 0$ if and only if $\varepsilon_i \leq (\tau - \widehat{p}_i)Y_i$. So for any mass of agents with perception $\widehat{p}_i$, the measure of evaders will be $\Pr[\varepsilon_i \leq (\tau - \widehat{p}_i)Y_i]$, where ε_i is uniformly distributed on $[0, Y_m]$. It follows that the measure of evaders with perceptions $\widehat{p}_i$ is $\tau - \widehat{p}_i$, and that the individual evasion rate is $e_{i,t} = 1 - \delta_{i,t}$.The measure of all non compliant taxpayers is therefore given by

$$e_t = \int_{\widehat{p}_i=0}^{1} \int_{\underline{y}}^{\overline{y}} \Pr[\varepsilon_i \leq (\tau - \widehat{p}_i)Y_i] dG(y_i) d\mathbb{P}_t(\widehat{p}_i)$$

Using the assumption on the distribution of ε_i it holds $\Pr[\varepsilon_i \leq (\tau - \widehat{p}_i)Y_i] = \frac{(\tau - \widehat{p}_i)Y_i}{Y_m}$. Hence the above equation can be written as

$$e_t = \int_{\widehat{p}_i=0}^{1} \int_{\underline{y}}^{\overline{y}} \frac{(\tau - \widehat{p}_i)Y_i}{Y_m} dG(y_i) d\mathbb{P}_t(\widehat{p}_i)$$

By using the fact that y_i and $\mathbb{P}_t$ are independently distributed, by integrating with respect to $dG(y_i)$ and remembering that $Y_m = \int_{\underline{y}}^{\overline{y}} Y_i dG(y_i)$ we obtain

$$e_t = \int_{\widehat{p}_i=0}^{1} (\tau - \widehat{p}_i) d\mathbb{P}_t(\widehat{p}_i)$$

or

$$e_t = \tau - \int_{\widehat{p}_i=0}^{1} \widehat{p}_i d\mathbb{P}_t(\widehat{p}_i) = \tau - \widehat{p}_{m,t}.$$

This function describes the equilibrium aggregate evasion rate corresponding to an average perception computed for given $\mathbb{P}_t$. Notice that for $\widehat{p}_{m,t} \to 0$, $e_t \to \tau$ due to the presence of moral benefits from compliance, and $\widehat{p}_{m,t} \to \tau$, $e_t \to 0$, so that the perception of the probability of apprehension is large enough to induce all taxpayers to full compliance. Later, when the endogenous dynamics of $\mathbb{P}_t$ will be characterized, conditions on the parameters such that, $\widehat{p}_{m,t} \in (0, \tau)$ and hence $e_t \in [0, 1]$ will be provided. ∎

Hence, aggregate evasion increases (linearly) in the level of the tax rate and it decreases with the perceived level of enforcement measured by the average perceived probability of apprehension.

Notice that Lemma 1 describes the individual evasion choice as a function of individual perception, Lemma 2 describes the average evasion rate as a function of the average perception for a given probability of perceptions $\mathbb{P}_t$. Individual perceptions (and hence averages) may depend on the average enforcement, which in turn, due to the enforcement externality, will depend on the average past evasion. So the evolution of $\mathbb{P}_t$ is endogenous in the model and has to be studied along with the equilibrium dynamics of the fiscal system.

In other words, $\mathbb{P}_t$ is endogenously determined in our model starting from a learning rule by the taxpayers and the externality in the enforcement. This latter links the actual probability of apprehension to the average level of tax evasion. The learning rule relates the taxpayers' perception to their past experience and to the real probability of apprehension. In order to study the evolution of $\mathbb{P}_t$ and the associated evolution of e_t we move now to provide a description of the tax enforcement system and of the learning rule.

3.1 The tax enforcement system

In this section we briefly describe our specification for the enforcement technology which is meant to emphasize the externality in the enforcement system relating the actual level of punishment to the aggregate level of tax evasion. We assume that the enforcement system comprises two different processes: the audit activity, the actual collection of unpaid taxes and fines. The former is performed by the tax agency, the latter is performed by the judiciary branch of the enforcement agency.

Denote $a_i(.)$ the (actual) probability that the taxpayer i is audited as a function of her observable characteristics. In general, this can arise as a function of the reported income and other observable characteristics (e.g. as part of an optimal mechanism design or a signaling game). To focus on the dynamic implications of the presence of the externality, we assume that a random auditing process is in place, and the tax authority audits any report by the taxpayers with a constant probability, a.

The total number of non-compliant cases reported by the tax auditors to the judiciary branch will be given by ae_t (the joint probability that tax was actually evaded and also audited, normalized on a population of taxpayers equal to one)[16].

Moreover we assume that the judiciary branch of the enforcement agency produces a certain level of sanctions, for any audited taxpayer having been caught in committing evasion, with a probability that depends on the number of cases to be decided by the jurisdiction, $\theta(ae_t)$[17].

To keep the analysis simple suppose that the judiciary will deliver a high level of sentences if the number of cases is low: $\theta_t = \overline{\theta}$ for $ae_t \leq a\widetilde{e}$ and a low number of cases is large enough, i.e. $\theta_t = \underline{\theta}$ for $ae_t > a\widetilde{e}$. In other words the fraction of finalized trials, given an audit produced evidence of evasion, is

[16] We are making the assumption that detected unpaid taxes are regularly disputed and the actual collection of due taxes depends on the outcome of finilized trials. If a_i depends on reported income a correlation between $\mathbb{P}$ and the income distribution would arise making the analysis more cumbersome.

[17] We assume that the jurisdiction decides all cases in the current period. A lag in the jurisdiction, for example, as when cases pertaining audits in the previous periods ae_{t-1} are processed at time t would add additional source of dynamics. Once again no further insights on equilibrium multiplicity and their implications would obtain by considering these lags.

equal to $\overline{\theta}$ whenever the number of administrative trials $a\theta_t$ is below a certain threshold $\widetilde{e}$ and is equal to $\underline{\theta}$ above that threshold. So the real probability that fines are enforced in the fiscal system is:

$$r_t = a\theta_t = \begin{cases} a\overline{\theta} & \text{for } e_t \leq \widetilde{e} \\ a\underline{\theta} & \text{for } e_t > \widetilde{e} \end{cases} \tag{3}$$

To study the dynamics and the steady state of the fiscal system (which includes the distribution of perceptions $\mathbb{P}_t$) we need to describe how taxpayers set expectations about the probability r_t of being apprehended and fined for evading taxes.

3.2 Information set and the learning rule

Given the simple institutional setting described above, the learning process by agent i on the consequences of non compliance can be formulated, in general terms, as follows:

$$\begin{aligned} \widehat{p}_{it} &= E[r_{i,t}|I_{it}, L_{it}] \\ &= E[a\theta|I_{it}, L_{it}] = l_i(I_{i,t}) \end{aligned} \tag{4}$$

where I is the information set about current aggregate variables and L is the learning rule, based on the history of individual past perceptions and experienced punishments.

Notice that learning is about an aggregate r_t (due to the simplification of random auditing and enforcement in the administrative trial), but past individual experience is considered in $I_{i,t}$. This general representation (see Sah, 1991) is specialized in the following for the case of myopic learning. Suppose that in each period t taxpayer i observes a private signal s_{it} about r_t. The noisy signal depends on the level of enforcement actually performed in the previous period by the tax administration as follows:

$$s_{i,t} = r_{t-1} + \eta_{i,t}$$

where $\eta_i \in [-\overline{\eta}, \overline{\eta}]$ is a i.i.d (across time) noise component with distribution function $H(\eta)$, such that $E(\eta_i) = 0$ and $Var(\eta_i) = \sigma^2_\eta$.

As for the learning rule we assume that agent i myopically forms her perceptions as follows:

$$\widehat{p}_{i,t} = \alpha_{i,t}\widehat{p}_{i,t-1} + (1 - \alpha_{i,t})s_{i,t}$$

Where α_{it} is the individual specific weight on past experience and $(1 - \alpha_{i,t})$ is the weight of new information about the real probability of apprehension[18].

We further restrict our analysis to a simple myopic learning scheme and we assume $\alpha_{i,t} = \alpha$. Given the properties of r_t from (3) the learning rule becomes:

$$\widehat{p}_{i,t} = \alpha\widehat{p}_{i,t-1} + (1 - \alpha)(r_{t-1} + \eta_{i,t})$$

[18] Notice that this formulation is not inconsistent with Bayesian learning rule, (see Sah, 1991) where the accumulation of experience modifies weights following the Bayes rule, see also the literature on OLS learning (Honkapoja 2010).

Notice that the transition of individual perceptions are driven by the aggregate level, r_t as induced by the resources collected in the past. By using equation (3) lagged one period, $\widehat{p}_{i,t}$ can be written as

$$\widehat{p}_{i,t} = \begin{cases} \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\overline{\theta} + \eta_{i,t}) \text{ for } e_{t-1} \leq \widetilde{e} \\ \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\underline{\theta} + \eta_{i,t}) \text{ for } e_{t-1} > \widetilde{e} \end{cases}$$

From (2) $e_{t-1} = \int_{\widehat{p}_i=0}^{1}(\tau - \widehat{p}_i)d\mathbb{P}_{t-1}(\widehat{p}_i) = \tau - \widehat{p}_{m,t-1}$ and hence the evolution of individual perceptions is given by:

$$\widehat{p}_{i,t} = \begin{cases} \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\overline{\theta} + \eta_{i,t}) \text{ for } \widehat{p}_{m,t-1} > \tau - \widetilde{e} \\ \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\underline{\theta} + \eta_{i,t}) \text{ for } \widehat{p}_{m,t-1} \leq \tau - \widetilde{e} \end{cases}$$

As for the average, by taking expectations over i, the aggregate perceived probability of apprehension (consider it as synthetic measure of the credibility of the enforcement system) is given by

$$\widehat{p}_{m,t} = \alpha\widehat{p}_{m,t-1} + (1-\alpha)E_i s_{i,t}$$

or

$$\widehat{p}_{m,t} = \alpha\widehat{p}_{m,t-1} + (1-\alpha)r_{t-1}$$

$$\widehat{p}_{m,t} = \begin{cases} \alpha\widehat{p}_{m,t-1} + (1-\alpha)a\overline{\theta}_t \text{ for } \widehat{p}_{m,t-1} > \tau - \widetilde{e} \\ \alpha\widehat{p}_{m,t-1} + (1-\alpha)a\underline{\theta}_t \text{ for } \widehat{p}_{m,t-1} \leq \tau - \widetilde{e} \end{cases} \tag{5}$$

Where the average perception is defined by $\widehat{p}_{m,t} = \int_{\widehat{p}_i=0}^{1}\widehat{p}_i d\mathbb{P}_t(\widehat{p}_i)$ and $\widehat{p}_{m,t-1} = \int_{\widehat{p}_i=0}^{1}\widehat{p}_i d\mathbb{P}_{t-1}(\widehat{p}_i)$ and $\mathbb{P}_t$ is defined starting from $\mathbb{P}_{t-1}$, for any t.

Hence the dynamics of the whole distribution of perceptions among taxpayers has to be characterized in order to study the dynamics of aggregate tax evasion and the steady states.

A salient feature of the dynamics of the fiscal system is that individual transitions depend on the aggregate distribution through the learning rule and they include r_{t-1} which, in turn, depends on past distribution of perceptions (cfr. interactive models of wealth distributions). More specifically, the evolution of individual perceptions $\mathbb{P}_t$ depends on the individual histories (as dictated by shocks and by the learning rule) and the average of the distribution $\mathbb{P}_{t-1}$ (inducing e_{t-1}).

Hence the past distribution of perceptions $\mathbb{P}_{t-1}(\widehat{p}_i)$ (through its average only) induces the current distribution of perceptions, $\mathbb{P}_t(\widehat{p}_i)$, along with the structural parameters Γ of the tax system.

The dynamics of individual perceptions is the source of persistence in time of tax evasion, and enforcement levels, i.e. of the tax system in general.

So, in order to obtain $\widehat{p}_{m,t}$ we need to study the evolution of the whole sequence of $\mathbb{P}_t(\widehat{p}_i)$ starting from $\mathbb{P}_0$, to show its existence and convergence to a steady state $\mathbb{P}^*$. This analysis is performed in the following section.

3.3 Convergence of the distribution of perceptions

The object of analysis in this section is the evolution of a random process $\mathbb{P}_t(\widehat{p}_i)$ describing the distribution of perceptions at each period t induced by $\mathbb{P}_{t-1}(\widehat{p}_i)$, the past distribution of perceptions, given the exogenous learning rule and the signal, $s_{i,t}$, obtained by each agent i. For the model defined in the previous subsections we provide a complete characterization of the possible evolution of $\mathbb{P}_t(\widehat{p}_i)$ and the stationary sets at steady state, starting from an arbitrary initial distribution $\mathbb{P}_0(\widehat{p}_i)$.

For each taxpayer the evolution of individual perceptions $\widehat{p}_{i,t}$ follows:

$$\Phi_i(\widehat{p}_{i,t-1};\eta_{i,t},r_{t-1}) = \alpha\widehat{p}_{i,t-1} + (1-\alpha)(r_{t-1}+\eta_{i,t}) \tag{6}$$

Notice that individual perceptions evolve depending on the aggregate level of enforcement, r_{t-1}.

By using the definition of r_{t-1}, equation (6) can be rewritten as

$$\Phi_i(\widehat{p}_{i,t-1};\eta_{i,t},r_{t-1}) = \begin{cases} \Phi_{i,h} \equiv \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\overline{\theta}+\eta_{i,t}) \text{ for } \widehat{p}_{m,t-1} > \tau - \widetilde{e} \\ \Phi_{i,l} \equiv \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\underline{\theta}+\eta_{i,t}) \text{ for } \widehat{p}_{m,t-1} \leq \tau - \widetilde{e} \end{cases}$$

where Φ_i is a correspondence $\widehat{p}_{i,t-1} \rightarrow \Phi_i(\widehat{p}_{i,t-1},.)$ mapping the set of equilibrium perceptions achievable by agent i at time t who started with a perception $\widehat{p}_{i,t-1}$, for different values of $\eta_{i,t}$ and given the aggregate state (r_{t-1},e_{t-1}) of the fiscal system. For future reference we denote $\overline{\Phi}_{i,h} \equiv \sup_\eta \Phi_h(.)$, $\underline{\Phi}_{i,h} \equiv \inf_\eta \Phi_h(.)$, $\overline{\Phi}_{i,l} \equiv \sup_\eta \Phi_{i,l}(.)$ and $\underline{\Phi}_{i,h} \equiv \inf_\eta \Phi_{i,h}(.)$. Suppose, for the moment, that $\overline{\eta}$ is small and it satisfies the following condition $\overline{\Phi}_l(t-\widetilde{e}) \leq \underline{\Phi}_h(t-\widetilde{e})$ or $\alpha(\overline{\theta}-\underline{\theta}) \geq \overline{\eta} - \underline{\eta} = 2\overline{\eta}$.

It is easily shown that the evolution of perceptions driven by Φ follows a linear Markov process with transition probability Q satisfying the usual definition.

Definition 1 *For all $x', x > 0$ the transition function*[19]

$$\begin{aligned} Q(x',x;r_{t-1}) &= \Pr\{\widehat{p}_{i,t} \leq x' | \widehat{p}_{i,t-1} = x\} \\ &= \Pr\left\{\eta_{i,t} \leq \frac{x'-\alpha x}{1-\alpha} - r_{t-1}\right\} \\ &= H\left(\frac{x'-\alpha x}{1-\alpha} - r_{t-1}\right) \end{aligned} \tag{7}$$

describes the cumulative distribution function of $\widehat{p}_{i,t} = x'$ given $\widehat{p}_{i,t-1} = x$.

Notice that the transition function depends on r_{t-1}. It follows that the sequence of distribution functions $\mathbb{P}_t$ satisfies the following recursive relation

[19] A transition probability on the state space $\widehat{p}_i \in Z$ is a function $Q : Z \times \mathcal{Z} \rightarrow [0,1]$ such that: 1. $Q(\widehat{p}_i,.)$ is a probability measure and 2. $Q(.,S)$ is a $\mathcal{Z}$ measurable function on $\mathbb{R}_+$, where S denotes a Borel set. It is easily verified that Q satisfies both conditions in our case.

$$\mathbb{P}_t(x') = \int Q(x', x; r_{t-1}) d\mathbb{P}_{t-1}(x) \tag{8}$$

For any given $\mathbb{P}_{t-1}$ we define:

$$\Phi^e_{i,h} \equiv \mathbb{E}_\eta[\Phi_{i,h}(.)] = \alpha\widehat{p}_{i,t-1} + (1-\alpha)a\overline{\theta}$$

and

$$\Phi^e_{i,l} \equiv \mathbb{E}_\eta[\Phi_{i,l}(.)] = \alpha\widehat{p}_{i,t-1} + (1-\alpha)a\underline{\theta}$$

By evaluating the state of perceptions at $\widehat{p}_{i,t-1} = \widehat{p}_{m,t-1}$ equation 5 can be rewritten as

$$\widehat{p}_{m,t} = \begin{cases} \Phi^e_{m,h} \text{ for } \widehat{p}_{m,t-1} > \tau - \widetilde{e} \\ \Phi^e_{m,l} \text{ for } \widehat{p}_{m,t-1} \leq \tau - \widetilde{e} \end{cases} \tag{9}$$

An equilibrium distribution of perceptions as of time t is a probability measure $\mathbb{P}_{t+1} = T^*\mathbb{P}_t$ and T^* is the self adjoint Markov operator associated to the transition $Q(x', x)$. $\mathbb{P}_t$ may converge to an invariant distribution, possibly depending on $\mathbb{P}_0$.

Definition 2 *A steady state of the fiscal system (invariant distribution of perceptions and induced evasion behavior) is a measure $\mathbb{P}$ satisfying $\mathbb{P}^* = T^*\mathbb{P}^*$.*

Endowed with these definitions we will now provide some results about the existence of $\mathbb{P}^*$, its convergence and the dependence of the dynamics on $\mathbb{P}^0$.

Our exposition will be as follows: first we discuss the steady states for the aggregate variables of our fiscal system $(\widehat{p}_m, r, e)$ assuming that $\mathbb{P}^*$ exists, then we will prove and characterize convergence of the distribution of perceptions $\mathbb{P}_t \to \mathbb{P}^*$.

3.4 Steady States of the fiscal system

Suppose, for the moment, that there exists (not necessarily unique, for any initial distribution $\mathbb{P}_0$) a distribution of perceptions such that $\mathbb{P}^* = T^*\mathbb{P}^*$.

Let $\widehat{p}^*_m = \int_{\widehat{p}_i=0}^1 \widehat{p}_i d\mathbb{P}^*(\widehat{p}_i)$ be the average perception. From the agents optimal evasion choice at equation (2) it holds:

$$e^* = \tau - \widehat{p}^*_m$$

which, by definition of $\widehat{p}^*_m$ must satisfy:

$$e^* = \begin{cases} (1-\alpha)\tau + \alpha e^* - (1-\alpha)a\overline{\theta} & \text{for } e^* \leq \widetilde{e} \\ (1-\alpha)\tau + \alpha e^* - (1-\alpha)a\underline{\theta} & \text{for } e^* > \widetilde{e} \end{cases}$$

or

$$e^* = \begin{cases} e^*_l = \tau - a\overline{\theta} \text{ and } e^* \leq \widetilde{e} \\ e^*_h = \tau - a\underline{\theta} \text{ and } e^* > \widetilde{e} \end{cases} \tag{10}$$

The steady state level of the average perceptions is therefore:

$$\widehat{p}^* = \begin{cases} \alpha\widehat{p}^* + (1-\alpha)a\overline{\theta} \text{ for } \widehat{p}^* > \tau - \widetilde{e} \\ \alpha\widehat{p}^* + (1-\alpha)a\underline{\theta} \text{ for } \widehat{p}^* \leq \tau - \widetilde{e} \end{cases}$$

Moreover the steady state probability of apprehension is as follows:

$$r^* = a\theta = \begin{cases} a\overline{\theta} \;\; \text{for } e^* \leq \widetilde{e} \\ a\underline{\theta} \;\; \text{for } \;\; e^* > \widetilde{e} \end{cases} \tag{11}$$

The results can be summarized in the following

Proposition 1 *Suppose there exists an invariant distribution of perceptions $\mathbb{P}^*$. Then there exist at most two steady states levels of aggregate perceptions $\overline{p}_m^* = a\overline{\theta}$ and $\underline{p}_m^* = a\underline{\theta}$ and at most two aggregate evasion rates e^* associated to $\mathbb{P}^*$. Depending on the tax rate (τ) and the enforcement system $(\widetilde{e}, a, \theta)$ there are three possible regimes:*

i) low tax level: if $\tau \in [0, \widetilde{e} + a\underline{\theta}]$ then there exists a unique steady state $\widehat{p}_m^* = \overline{p}_m^* = \overline{r} = a\overline{\theta}$ and $e^* = e_l^* \leq \widetilde{e}$;

ii) intermediate tax level: if $\tau \in [\widetilde{e} + a\underline{\theta}, \widetilde{e} + a\overline{\theta}]$ then there exists two steady states $\widehat{p}_m^* = \overline{p}_m^* = \overline{r} = a\overline{\theta}$ and $e_l^* \leq \widetilde{e}$ or $e^* = e_h^* > \widetilde{e}$ and or $\widehat{p}_m^* = \underline{p}_m^* = \underline{r} = a\underline{\theta}$

iii) high tax level: $\tau \in [\widetilde{e} + a\overline{\theta}, 1]$ then there exists a unique steady state $\widehat{p}_m^* = \underline{p}_m^* = \underline{r} = a\underline{\theta}$ and $e^* = e_h^* > \widetilde{e}$.

Proof. Suppose there exists $\mathbb{P}^* := S^* \to [0,1]$, where $S^* \subseteq [0,1]$ is the support of $\mathbb{P}^*$. Then it holds $\widehat{p}_m^* = \int_{\widehat{p}_i=0}^{1} \widehat{p}_i d\mathbb{P}^*$ and $e^* = \tau - \widehat{p}_m^*$. That every possible steady state coincides with a rational expectation equilibrium and it holds: $\widehat{p}_m^* = r$ follows as a straightforward consequence of the definition of $\widehat{p}_m^*$. That at most two steady states aggregate evasion rates exist clearly follows from (10). Depending on Γ, there are three possible regimes under which the fiscal system operates in the steady state. These are obtained from the study of the average transition function

$$\widehat{p}^* = f(\widehat{p}) \begin{cases} \alpha\widehat{p}^* + (1-\alpha)a\overline{\theta} \text{ for } \widehat{p}^* > \tau - \widetilde{e} \\ \alpha\widehat{p}^* + (1-\alpha)a\underline{\theta} \text{ for } \widehat{p}^* \leq \tau - \widetilde{e} \end{cases}$$

Where it is easy to see that for $\tau \in [0, \widetilde{e} + a\underline{\theta}]$ only $\overline{p}_m^* = a\overline{\theta}$ satisfies $f(a\overline{\theta})$ and hence $e = e_l^* = \tau - a\overline{\theta}$. For $\tau \in [\widetilde{e} + a\overline{\theta}, 1]$ only $\underline{p}_m^* = a\underline{\theta}$ satisfies $f(a\underline{\theta})$ and $e = e_h^* = \tau - a\underline{\theta}$. Finally, for $\tau \in [\widetilde{e} + a\underline{\theta}, \widetilde{e} + a\overline{\theta}]$ both $\underline{p}_m^*$ and $\overline{p}_m^*$ can satisfy f. For any $\widehat{p}_m^*$ the associated e^* can be easily recovered by using (10) in each of the two steady states. Clearly, whenever multiple steady states for aggregate perceptions and aggregate evasion are obtained, they must be supported by different limit distribution $\mathbb{P}^*$. ■

In words, if the tax rate τ is high compared to a measure of efficiency of the enforcement system $(\widetilde{e}, a)$ only the high evasion-low enforcement equilibrium can be a steady state; if the tax rate is low compared to the efficiency of the enforcement system only the low aggregate evasion rate can be consistent with steady state, in all the intermediate cases two equilibria will be consistent with steady state; which one of the two will prevail may depend on historical initial conditions described by $\mathbb{P}_0$.

These results capture the standard intuition associated to the enforcement externality, although in a dynamic setting with learning. When the economic profitability of tax evasion is large (i.e. taxation is large) then there will be enough agents (given the distribution of the moral cost ε) willing to evade, the judiciary will be less effective in sentencing the fines and the high evasion equilibrium becomes self-sustaining due to the enforcement externality. The opposite happens if the tax rate is low enough. For intermediate levels of taxation both equilibria are possible.

Notice that in all cases the limit probability of perceptions are correct on average (i.e. after integrating with respect to η_i) and the steady states coincide with rational expectation equilibria where $\widehat{p}_m^* = r^*$, i.e. the average individual perception coincides with the true probability of apprehension.

We should also notice that, for the steady states of aggregate evasion, the larger the congestion effect on the enforcement system, i.e. the lower the value of $\widetilde{e}$, the lower must be the tax rate in order to enforce the low evasion equilibrium at steady state. Moreover, the greater the difference between $\underline{\theta}$ and $\overline{\theta}$, i.e. the greater the consequences of the externality on the returns from audits, the larger the interval for the existence of multiple steady-states.

3.5 Dynamics of the Fiscal State

Our next task is to prove that, indeed, the learning model with the enforcement externality converges time to the rational expectations equilibrium over time. In the following we prove that in all of the three regimes (defined by the values of the parameters of the fiscal system) the distribution of perceptions will converge to $\mathbb{P}^*$. In particular, we will prove that in case i) and iii) in proposition (1) convergence to a unique $\mathbb{P}^*$ is obtained independently of the initial distribution $\mathbb{P}^0$; whereas in case ii) the Markov process will not be ergodic and which distribution of perceptions the fiscal system will induce will depend on the initial distribution of perceptions $\mathbb{P}_0$.

We highlight that the study of the dynamics in the case when multiple equilibria can emerge is important in that we will consider the initial conditions of the system, i.e. the initial distribution of perceptions by taxpayers, as the selection device for the learning process leading to a steady state equilibrium.

Proposition 2 *Starting from an arbitrary distribution of perceptions $\mathbb{P}_0(\widehat{p}_i)$, the Markov process defined by Φ will converge to a unique $\mathbb{P}^*$, possibly depending on $\mathbb{P}_0$.*

Proof. We proceed by proving the result in a few steps. Start by noticing that which real apprehension probability an individual taxpayer faces only depends on aggregate evasion and hence on the average perception in the past period according to equation (3). This means that which regime, $\Phi_{i,h}$ or $\Phi_{i,l}$, governs the individual transition probability depends both on the past individual state and on the aggregate state and hence on the past average perception, through the signal $s_{i,t}$ observed by each taxpayer as described by equation (6). Also recall that the current average perception is completely determined by the state of the past average perception according to equation (5). The fact that the dynamics only depends on the first moment of the current distribution greatly simplifies the analysis of our interactive model of the dynamic evolution since it allows to apply standard results. Guided by proposition 1 the analysis of the dynamics has to distinguish among different regimes. *Case i) (low tax level).* Suppose $\tau \in [0, \widetilde{e}+a\underline{\theta}]$. From equation (9) it follows that $\Phi^e_{m,l}(\tau-\widetilde{e}) > \tau-\widetilde{e}$ and $\Phi^e_{m,h}(\tau-\widetilde{e}) < \tau-\widetilde{e}$. (Graphically modify Φ^e_l evaluated at $(\tau-\widetilde{e})$ accordingly in fig. 1) $\Phi^e_{m,l}$ cannot cross the 45° line whereas $\Phi^e_{m,h}(\tau-\widetilde{e})$ has to cross it). Now suppose that $\mathbb{P}_0$ is such $\widehat{p}_{m,0} < \tau-\widetilde{e}$, hence individual perceptions are driven by $\Phi_{i,l}$ (see equation 6) and $r_t = a\underline{\theta}$. Then there must exist a finite $n > 0$ such that for all $n' > n$ it holds $\widehat{p}_{m,n'} > \tau-\widetilde{e}$ and hence individual transitions $\widehat{p}_{i,t}$ are governed by $\Phi_{i,h}$ (see equation 6) in all subsequent periods $t \geq n'$ and it also holds $r_t = a\overline{\theta}$, for $t > n'$. From the properties of $\Phi_{i,h}$ (see equation 6) it will exist n" such that individual perceptions $\widehat{p}_{i,t} \in [c,d]$, with $c \geq \tau-\widetilde{e}$ and $d \leq 1$. It also holds that any individual state in $[c,d]$ is reachable starting from any other state in the same interval. This mixing condition Condition M in Stockey and Lucas (SLp) p. 348 guarantees that $\mathbb{P}_t$ converges to a unique limit distribution on $[c,d]$. Suppose, instead that $\mathbb{P}_0$ is such $\widehat{p}_{m,0} > \tau-\widetilde{e}$ then individual transitions are governed by $\Phi_{i,h}$ (see equation 6) it holds $n' = n = 0$ and similar arguments for the application of the mixing conditionapply. Therefore for any $\mathbb{P}_0$ the limit distribution converges to $\mathbb{P}^*$ such that $\widehat{p}_{m,\infty} \to a\overline{\theta}$. *Case iii) (high tax level)* Suppose $\tau \in [\widetilde{e}+a\overline{\theta}, 1]$. A similar argument as in the previous case establishes convergence of $\mathbb{P}_0$ to a unique distribution $\mathbb{P}_*$ supported on $[a,b]$ with $b \leq \tau-\widetilde{e}$ and $\widehat{p}_{m,\infty} \to a\underline{\theta}$. Once again, the key observation is that in this case there necessarily exists n such that $\widehat{p}_{m,t} < \tau-\widetilde{e}$ for $t \geq n$ so that individual perceptions are necessarily governed by $\Phi_{i,l}$ from a point in time onwards. *Case ii) (intermediate tax level)* Suppose $\tau \in [\widetilde{e}+a\underline{\theta}, \widetilde{e}+a\overline{\theta}]$. From equation (9) it follows that $\Phi^e_{m,l}(\tau-\widetilde{e}) < \tau-\widetilde{e}$ and $\Phi^e_{m,h}(\tau-\widetilde{e}) > \tau-\widetilde{e}$. (Graphically, see fig. 1, both $\Phi^e_{m,l}$ and $\Phi^e_{m,h}(\tau-\widetilde{e})$ must cross the 45° line). Suppose $\mathbb{P}_0$ is such that $\widehat{p}_{m,0} < \tau-\widetilde{e}$ then $r_t = a\underline{\theta}$ (from equation 3) and $\widehat{p}_{m,t} < \tau-\widetilde{e}$ for all $t > 0$. Hence, individual perceptions are driven by $\Phi_{i,l}$ (see equation 6) for all $t > 0$. Then there exists n and $a, b \in [0, \tau-\widetilde{e}]$ such that $\widehat{p}_{i,n} \in [a,b]$ for all $t > n$. Hence, starting from n, the process satisfies Condition M in Stockey and Lucas (SLp) p. 348. The process must converge to $\mathbb{P}_*$ such that $\widehat{p}_{m,n} = a\underline{\theta}$ and $e^* = e_h$. Suppose, instead that $\mathbb{P}_0$ is such that $\widehat{p}_{m,0} > \tau-\widetilde{e}$. Then $r_t = a\underline{\theta}$ (from equation 3) and $\widehat{p}_{m,t} > \tau-\widetilde{e}$ for all $t > 0$. Hence, individual perceptions are driven by $\Phi_{i,h}$ (see equation 6) for all $t > 0$. By the properties of $\Phi_{i,h}$ there must exist n and $c, d \in [\tau-\widetilde{e}, 1]$ such that $\widehat{p}_{i,n} \in [c,d]$ for all $t > n$. In this case the process must converge to $\mathbb{P}^*$ and $\widehat{p}_{m,\infty} \to a\overline{\theta}$, $e^* \to e_l$. We conclude that in case i) and iii) the dynamics of the fiscal state is ergodic and depends on the fundamental Γ. In case ii) which fiscal state emerges in the limit depends on the initial distribution of perceptions.

In other words, when the parameters of the fiscal system are as in case ii) in

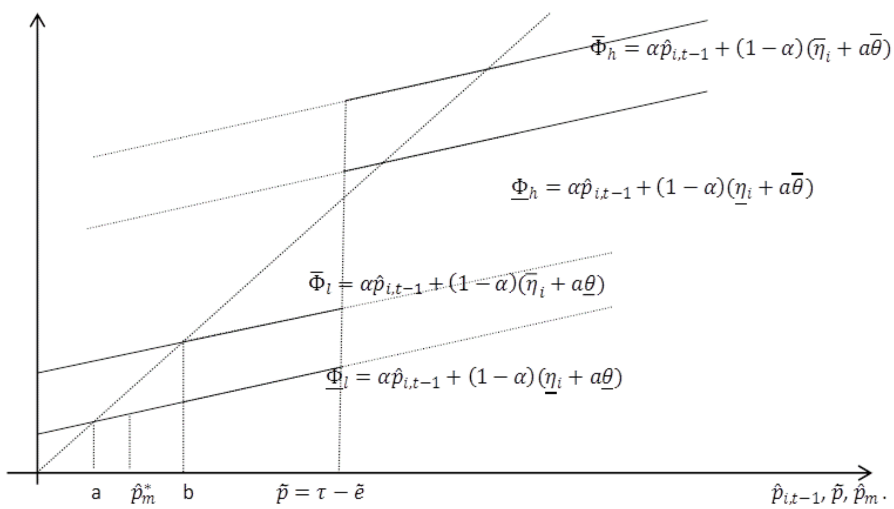

Figure 1: Convergence of aggregate perceptions

Proposition 1, the evolution of the fiscal system will feature history dependency, the speed of adjustment is dictated by α, the dispersion of individual behavior will be dictated by σ_η. As a consequence the fiscal system will feature a high evasion or low evasion equilibrium depending on the initial perceptions $\mathbb{P}_0$ of taxpayers. ■

3.6 Convergence when the support of η_i is large

All the results in the previous section were derived under the assumption that $\overline{\eta}$ is low enough. In particular $\overline{\eta}$ was such that $\overline{\Phi}_l(t-\widetilde{e}) \leq \underline{\Phi}_h(t-\widetilde{e})$ or $\alpha(\bar{\theta}-\underline{\theta}) \geq \overline{\eta}-\underline{\eta} = 2\overline{\eta}$. It is a standard result of the Markov process that if the random component of the dynamic transition has a large enough variance, then ergodicity is obtained. In the context of our model it is easy to see the consequences for the convergence of $\mathbb{P}_t$ in the complementary case when $\alpha(\bar{\theta}-\underline{\theta}) < 2\overline{\eta}$, that is when the noisy signal in the learning process is more precise around its average.

This situation can occur either when the information technology to the taxpayers features lower precision (think about the role of experts) or when there is an intrinsically large noise component in the announcement that the fiscal authority makes about the enforcement system (once again due to absence of commitment and state capacity). This is interesting to emphasize since it highlights the role of experts in the reinforcement of the fiscal externality.

Indeed it is easy to see that if the support of s_i is large i.e. whenever $\overline{\eta} > \frac{\alpha(\bar{\theta}-\underline{\theta})}{2}$ is large enough, the model will exhibit different dynamics and global convergence is obtained. The result is summarized in the following

Corollary 1 *If $\overline{\eta} > \frac{\alpha(\bar{\theta}-\underline{\theta})}{2}$ then the fiscal system converges to a unique distribution $\mathbb{P}^*$.*

Proof. If $\overline{\eta} > \frac{\alpha(\bar{\theta}-\underline{\theta})}{2}$ then $\overline{\Phi}_l(t-\widetilde{e}) > \underline{\Phi}_h(t-\widetilde{e})$. Define $\widehat{p}_{i,\min}$ as the fixed point of $\widehat{p}_{i,\min} = \underline{\Phi}_l(\widehat{p}_{i,\min})$ and $\widehat{p}_{i,\max} = \underline{\Phi}_l(\widehat{p}_{i,\max})$. Clearly, they both exist and $\widehat{p}_{i,\min} < \widehat{p}_{i,\max}$. Moreover, there exists a finite n such that any state $\widehat{p}_i \in [\widehat{p}_{i,\min}, \widehat{p}_{i,\max}]$ must be reached starting from any $\widehat{p}_i \in [0,1]$. By definition any $\widehat{p}_i \in [\widehat{p}_{i,\min}, \widehat{p}_{i,\max}]$ is an invariant set for $t > n$ and any state in it is reachable starting any other state in this interval. Hence, from the mixing condition M, there exists a unique limit distribution for the sequence (8). Consider case II in Proposition (1) then for $\overline{\eta} \leq \frac{\alpha(\bar{\theta}-\underline{\theta})}{2}$ the limit distribution does not depend on initial condition. In case i) and iii) in Proposition (1) the model was already proved to be ergodic. ■

Even if it is a standard conclusion from general principles in Markov processes the result above highlights the role of experts. The more efficient the market for expert is, the larger the precision of the signal the taxpayers get, the stronger will be the enforcement externality and the emergence of history dependence, if the fundamentals are consistent with their presence.

3.7 Fiscal unifications and the Mezzogiorno problem (to be completed)

The model can also be a guide for interpreting the long run consequences of a fiscal unification. Suppose that two economies (North and South) start from two different steady states. Suppose North features a high tax rate τ_N (and good infrastructures) a given level of monitoring a_N and is in the high enforcement $\bar{\theta}$ and low evasion steady state. Suppose that South features low tax rates τ_S (and bad infrastructures) and a given level of monitoring a_S and it also is in a high enforcement $\bar{\theta}$ and low evasion steady state. Suppose all other elements of Γ (fiscal morality, income distribution, learning process, threshold level of evasion, signal precision etc.) are the same.

Suppose that in the newly unified economy tax rates is raised to τ_N in the two regions, the audit technology is unified $a_S = a_N$, but the judiciary branch of the fiscal administration operates in a decentralized way, i.e. the efficiency of the fiscal courts can depend on the local level of aggregate tax evasion. Suppose all other elements of Γ (fiscal morality, income distribution, etc.) are unchanged. What are the consequences on the long run evolution of the newly unified fiscal system? Suppose that the parameters Γ are such that regime ii in proposition (1) prevails. Then the answer depends on the initial distribution of perceptions in the two regions at the time of unification and whether the unification process affects it.

We need to distinguish two cases. Suppose that the unification process is such that every tax-payer in each region N and S is endowed with the same learning process and they sample from the same population. Then the enforcement externality operates at the national level and the whole economy will converge to the same fiscal equilibrium. In particular, if the initial average perception of the unified economy (notice they are different due to different tax rates) is below the threshold $\widetilde{p} = t_N - \widetilde{e}$ then both regions will converge to the same high evasion low enforcement steady state. If the initial perception is above both regions will converge to the high enforcement-low evasion fiscal equilibrium.

Consider instead the case in which each taxpayer samples the signal s_i from the local population in N and S respectively. Then the enforcement externality becomes a local externality and the two regions can converge to two different fiscal equilibria. Notice that for those taxpayers leaving in N there has been no change in the considered example since all exogenous parameters are the same. By assumption $\widehat{p}^N_{m,} > \tau_N - \widetilde{e}$ so $\widehat{p}_{m,N} = r_N = \bar{\theta} a_N$ will not change and the local fiscal equilibrium will not change after unification. For those taxpayers leaving in S the tax rate after unification is larger and the monitoring probability is different. Suppose $a_S < a_N$ before unification so the initial perception of (and not necessarily evasion) apprehension is lower $\widehat{p}_{m,s} < \tau_N - \widetilde{e}$, at the moment of unification. Then, by proposition (1) $\widehat{p}_{m,s}$ will converge to $\widehat{p}_{m,S} = r_S = \underline{\theta} a_N$, and S will be trapped in a low enforcement- high evasion equilibrium.

In other words if unification entails a raise in the tax rate and a decentralized fiscal jurisdiction in which the efficiency of the judiciary is influenced by the local level of tax evasion the model above is consistent with a fiscal trap in the annexed region, the so called Mezzogiorno problem. In the next section we will provide some evidence in support of the idea that both ingredients (an increase in tax rates in the annexed region and decentralized enforcement) were present in the post unification in Italy after 1870.

4 A suggested interpretation for the north/south divide of fiscal compliance in Italy (Preliminary)

In this section we summarize some of the main aspects of the history of the enforcement of fiscal law in Italy since its unification. We will document how the local externality in the judiciary branch in the design of the enforcement process have been particularly relevant in the case of Italy so that our focus on the divergence of compliance behavior by taxpayers at regional level. This will provide historical evidence that motivates our aim of investigating a model where the joint dynamics of compliance and enforcement is induced by a local enforcement externality.

Our analysis has shown that different initial perceptions, large tax rates and local responsibility for the enforcement are the key ingredients for the emergence of a situation where multiple equilibria may emerge. When one aspect of the enforcement has an important local component it the model predicts divergence in compliance behavior in different areas of the same country.

Of course it is not our aim to understress the structural differences, in sectorial composition, civil progress and social capital that can also be at the root of different fiscal compliance, already investigated in the existing literature on the North-South divide. [20]

Our aim is instead to list a variety of reasons why some of the ingredients that produce multiple equilibria and make history matter in the design of the fiscal system. There are we believe three ingredients that are relevant for our view:

1. Most of the enforcement was implemented at the local level;[21]

2. The tax rate in the new nation was quite larger compared to other states in Europe and moreover, tax revenues were raised in Southern Italy, often in a drastic way;

3. Initial beliefs about the enforcement system were very different in different areas of the country.

[20] quote recent references- e.g. Romeo, Gershenkron, Fenoaltea-Ciccarelli, Malanima.

[21] In Appendix D we provide a simple model where lower perception about enforcement lead to larger tax evasion and hence the local political equilibrium favors more lenient attitudes in the jurisdiction of tax evasion

4.1 An overview of the fiscal system in Italy after unification

Italy was unified in 1861. This was the final outcome of a social, political and military campaigning deeply rooted in the geopolitics of Europe on the XIXth Century. The process was led by the Savoy monarchy based in the north west (Piedmont) and successively interested regions in the Centre and in the South (Kingdom of Naples and the Papal State). The socio economic structure of these political units was still based, to a significantly different degree, on the inherited hubris of feudal socioeconomic institutions governed by monarchies restored by the Congress of Vienna (1815) all over Europe at the end of the Napoleonic wars.

We will not overview the evolution of the fiscal state in Italy. Moreover in 1861, fiscal systems in the different (pre-union) Italian states were characterized by significant differences, with the fiscal pressure being substantially lower in the Southern regions with respect to the other areas and, especially, to Lombardy and Veneto. Not surprisingly, as stressed by a large empirical evidence, a "low " fiscal pressure went along with a "low" level of infrastructures which would have constituted the essential "prerequisites" for the take-off of the economic systems. This was not the case for regions such as Piedmont, where a quite "high" fiscal pressure guaranteed a higher level of "public" investment[22].

From the point of view of the enforcement of fiscal law the main problem faced the government was that of transforming heterogeneous populations into loyal taxpayers[23]. This was especially true for Southern Italy where there had been almost no social base and political support for the unification ad where the process although not in the form of a proper annexation war had been due the collapse of the Bourbon political order and its institutions accelerated by small scale military operations (Spedizione dei Mille).

It is widely believed among historians that the Italian divide between the Centre-North and the South of the country is a long run phenomenon with complex and multifaceted aspects that still need to be carefully investigated. In 1861, when the country was unified, economic as well as social conditions in Italian regions were deeply different. It comes as no surprise that, in designing a general fiscal system, differences and local specificities had to be taken properly into account (Dominici-Marongiu p.12, 2005).

In this framework, the decision to extend to the newly born country the fiscal regime of Piedmont was seen as a necessary step toward economic integration for a number of reasons: on one side, it was felt that a modern State

[22] The evolution of the fiscal system since unification is outside the scope of our discussion here, we only mention that Italy slowly evolved from a mix of property tax and indirect tax to a system based on income tax (Imposta di Ricchezza Mobile, IGE and then Irpef) and value added tax (IVA).

[23] So even if structural characteristics had been the same (and we know they were not), northern and southern italy had two different attitudes and beliefs vis à vis the enforcement system. Southern Italy has been on the verge of a fiscal riot (brigantaggio, performed by farmers but actually financed by landowners and nostalgic supporters of the Borbone Kingdom, in many instances): of course there must be quite different perceptions about the credibility of the enforcement system if riot episodes only arose in southern regions.

had to rely on a sufficiently larger amount of resources in order to be able to finance external debt (this was the case of Piedmont but not, however, of the Reign of Sicily), while, on the other, it was thought that, even if the regional differences were clearly recognizable, the homogeneity of fiscal duties will have reinforced the faith in the newly born State.

One important characteristic of the fiscal regime which was introduced in 1864, based on a tax on "Ricchezza Mobile", was that, for the first two years, the revenues had been defined at the central level (Sistema del contingente) rather than computed after the compliance of taxpayers, as it happened since 1866 onwards. The reason for this was that the Government wanted to make sure that tax evasion would not determine a significant reduction of the revenues that the Government expected to receive. In the "contingente" regime, as Quintino Sella put it in 1862 " every taxpayer, in each place, will know that a cent concealed by a member of the community will determine a heavier tax burden for himself; thus, you will find groups of men inspecting each other in order to satisfy the community needs. . . "

In his analysis of the Italian Fiscal system Giulio Alessio (1883) has an interesting discussion of tax evasion. Figures reported in p.350 show that, in 1877 the total income reported by private workers in all sectors (industria, commercio, professioni, arti e mestieri) amount to 495 millions Liras whereas total public expenditures amount to 495 millions Liras of which 266 are wages to public employees. By adding wages to public employees in local administrations the toal amount recorded for the wages in the public sector is 319 millions. Compared to the composition of the labor force at the time this figure is considered as unbelievable. The causes of the tax evasion are identified to the large tax rate equal to 13,20 for the Imposta sui redditi della ricchezza mobile (as a benchmark the income tax rate was 2,05-2,46 in England in 1869) and on the limitations of the "Commissioni di Accertamento", the tax agency.

One important limitation of the Commissioni di accertamento is the appointment system, which in our view, is the real originator of the local enforcement externality on which the present investigation is focused. It is worth to report the whole passage:

"Aggiungasi che la composizione delle Commissioni, tratte dai consigli comunali, non è sufficiente guarentigia della loro energia nel sostenere gli interessi dello Stato, mentre il Comune non ha alcuna partecipazione al prodotto, e d'altra parte i ceti agiati e potenti, che vi predominano nelle assemblee, assai difficilmente mandano delegati intesi a colpire realmente le maggiori fortune; di solito la possidenza e l'alto commercio hanno cura di farvi entrare, e prevalentemente i propri rappresentanti: anzi il manteimento delle cariche nelle stesse persone per lunghi anni, senza divivieti di rieleggibilità , rende possbile il formarsi di gruppi interessati a vantaggio di dati organismi industriali e di alcune forme di produzione".

Basically the idea is that the level of the enforcement is determined by local interests, since the tax agent and the Commissione di Accertamento (the original form of a modern tax court, where however both income estimates and jurisdictional functions were performed).

As it appears clear, tax evasion was certainly an expected outcome of the post-union fiscal reform and it was perceived to be concentrated in similar sector as in modern times.

A quick overview of the history of this local dimension in the design of the enforcement of fiscal law confirms that this original tract has been almost a constant after unification (see Galeotti, 1967 and Palelologo, ed. 2005).

The Commissioni di accertamento were composed of five members for each municipality, the chairman was appointed by the Prefect (the representative of the central government in the Province) others were elected by the city council. Originally they had both a role of inspection and initiative in the estimation of individual income. Against their decision taxpayers could appeal at a Commissione Provinciale also composed by 5 members, 2 of them were appointed by the province council (the local parliament), two of them were appointed by the Chamber of Commerce (the local organization of the entrepreneurs) and one by the Prefect. A first reform of the design of the enforcement of the fiscal law was passed in 1866, as a consequence the Commissione di Accertamento was transformed into a proper fiscal court so that the auditing activity was the responsibility of the tax agents. This arrangement established a system that as for the local dimension of the enforcement of the fiscal law survived almost untouched until 1972. In the period in between there was a large debate among specialists about the nature of the nature of the fiscal court and their role as special judges with particular attention to the role of the Commission Tributaria Centrale (the High court in the fiscal process). In the fascist period a new reform (riforma degli ordinamenti tributari) was passed in 1936 and a royal decree on the composition of the Commisison Tributarie was passed a year later (1937, n.1516). Mostly consisting in clarifying the nature of the jurisdiction by fiscal courts and not much impact on the composition of the Commissioni Tributarie. Members of the Commissioni distrettuali (first degree) were appointed by the intendente di finanza (the local representative of the finance minister) on a list composed by Unioni sindacali e del consiglio provinciale presideuto dal prefetto (Lignani 2005, p.61).

In the republican era the rules were formally changed again: majors of municipalities composed the list of candidates to Commissioni distrettuali and the prefect the lists for the Commissioni Provinciali. These lists included three times the number of members to be appointed, the selection was made by "intendente di finanza" and by the Ministry. The opinion of Lignani (2005, p.63) is that although the new requirement was that of independence from

The first important reform in the Republican era was that of 1972 where the discipline of the system of appeals (ordinary law caourts and fiscal law courts) was disciplined. No major changes occurred in the composition and in the appointment od the judges in the Commussione Tibutaria and in the Commisisone Povinciale (see Trotta, p.37).

****

The third important dimension is the local enforcement.

However, historians tend to stress that large efforts were made in order to guarantee that the distribution of the tax burden among provinces and local communities were fair enough, leaving at decentralised structures (Commissioni) the task to determine the personal incomes to be taxed. These Commissions were considered to be fundamental pillars of a democratic fiscal system where voters-taxpayers were asked to participate (at least indirectly) to the formation of these administrative bodies. In these commissions the majority was attributed to the elected components and, moreover, the expertise of the elected had to be proved. In his critical revision of the procedure adopted, the Ministry of finance, Quintino Sella noted that this system had worked in a largely satisfactory way and that it had proved to be extremely effective in the peculiar context of the newly born country where a system based on the interaction between local authorities (as expressed by the Commissioni) and taxpayers had been felt as more palatable than any other more centralised scheme.

However, on the other side, the reliance on local resources in the design of the structure of the new tax system determined a further element of asymmetry among tax districts with the likely effects of reinforcing existing differences that a unified framework should have aimed at diluting, in order to favour convergence in economic and social conditions.

5 Some consequences for the formation of a fiscal union (very preliminary).

A fiscal union between two different countries can occur in the form of conquest, voluntary annexation or in the form of a federation. The model above allows us to consider different cases.

Consider case ii) in proposition 1. Suppose there are two economies with the same structure Γ of the fiscal system, one starting from steady state $\mathbb{P}_l^*$ the other $\mathbb{P}_h^*$ nothing will happen if the local dimension of the enforcement system is left untouched. If instead the same enforcement system is extended to both the unified regions, which steady state will prevail depends on the size of the two populations in the joining countries. In the next version we will provide a formalization of this result. In any case even if $\mathbb{P}_h^*$ will be induced in the limit, it will take time for the low compliance country starting from $\mathbb{P}_l^*$ to achieve high standards of compliance.

6 Conclusion

Our analysis shows that different initial perceptions, large enough tax rates and local responsibility for tax enforcement are the key ingredients for the existence of multiple equilibria. In the presence of multiple equilibria, the equilibrium selection mechanism relies on history and initial perceptions. Persistence of tax evasion is driven by the learning mechanism and the externality of the enforcement system.

These results can help to interpret geographical patterns of tax evasion in Italy. Different initial perceptions of the efficiency of the (local) enforcement system and different initial historical conditions, e.g. brigantaggio in the south of Italy in the post-unification period, can explain the evidence of higher tax evasion and lower enforcement which persists in the South of Italy since the eighteenth century. Our explanation for this situation is based fundamentally on a bad initial institutional design, rather than in structural differences or differences in social capital between North and South of the country. We do not exclude the presence of these differences, but we find it more difficult to interpret the long lasting situation in the South of Italy simply on the basis of differences in endowments (of wealth or social capital). There are important policy implications that can be drawn from our analysis: if history matters (as for example with multiple equilibria and learning about detection rates), it may take some time to see a change in compliance after a change in the enforcement policy, as perceptions are based on what happened in the past and hence they do not immediately adjust to the true probability. Moreover, as already suggested by Sah (1991), the change in enforcement policy, like more resources devoted to audits, might need to be persistent before observing any change in compliance. So a reallocation of audit resources may not immediately translate into greater compliance.

7 References

Alborino N., Dongiovanni, S. and Spingola , A., 2008 , Proposta di indicatori di efficienza per l'Agenzia delle Entrate: metodologia. *Documenti di Lavoro dell'Ufficio Studi 2008/5, Agenzia Entrate.*

Alessio, G., 1883, Saggio sul Sistema Tributario in Italia e i suoi effetti economici e sociali, Parte Prima, Vol.1, Fratelli Bocca Librai.

Becker, G.S., 1968, Crime and Punishment: An Economic Approach, *Journal of Political Economy,* 76, pp.169-217.

Blackburn, Keith, Bose, Niloy and Emranul Haque, M., (2006). The incidence and persistence of corruption in economic development, *Journal of Economic Dynamics and Control*, vol. 30(12), pages 2447-2467, December.

Cameron, Colin, A. and Trivedi Pravin, K. (2010) Microeconometrics using Stata, *Stata Press.*

Carbone, E. and Spingola, A., (2015), Analisi delle direzioni provinciali attraverso il DB GEO, *Argomenti di Discussione 01/15, Agenzia delle Entrate.*

Ehrlich, Isaac (1972), The Deterrent Effect of Criminal Law Enforcement, *The Journal of Legal Studies*, 1(2), pp.259-276..

Ehrlich, Isaac (1973), Participation in Illegitimate Activities: A Theoretical and Empirical Investigation, *Journal of Political Economy*, 81, pp.521-565.

Ferrer R. (2010), Breaking the law when others do: a model of law enforcement with neighborhood externalities, *European Economic Review*, 54, pp. 163-180.

Galbiati, R. and G. Zanella (2012), The tax evasion social multiplier: evidence from Italy, *Journal of Public Economics*, 96, pp. 485-494.

Gale D. (1996), What have we learned from social learning?, *European Economic Review*, 40, pp.617-628

Galeotti, (1967), Le Origini dell'Imposta sui redditi della ricchezza mobile in Italia, *Giuffrè editore, Milano.*

Lochner, Lance (2007) Individual Perceptions of the Criminal Justice System, *American Economic Review*, 97(1), pp. 444-460.

Manestra Stefano, (2010), Per una storia della tax compliance in Italia, *Questioni di economia e finanza, Volume 81, Occasional Papers, Banca d'Italia,* Roma.

Musgrave R.A. (1992), Schumpeter's crisis of the tax state: an essay in fiscal sociology, *Journal of Evolutionary Economics*, vol. 2, pp. 89-113.

OECD (2015), Tax Administration 2015: Comparative Information on OECD and Other Advanced and Emerging Economies, *OECD Publishing*, Paris. http://dx.doi.org/10.1787/tax_admin-2015-en

Paleologo G. Ed. (2005), La giustizia Tributaria Italiana e la sua commissione centrale, studi per il quarantennio, *Giuffrè editore*, Milano

Pisani, Stefano and Polito Stefano, (2006), Analisi dell'evasione fondata sui dati IRAP, Anni 1998- 2002, *Documenti di Lavoro dell'Ufficio Studi, Agenzia delle Entrate.*

Pisani, Stefano (2014), An approach to assess how the activity of the Italian Revenue Agency affects compliance. *Discussion Topics 01/2014, Agenzia Entrate.*

Sah, Raaj K. 1991. Social Osmosis and Patterns of Crime, *Journal of Political Economy*, 99, 1272-1295.

Christian Traxler, (2010), Social norms and conditional cooperative taxpayers, *European Journal of Political Economy*, vol. 26,(1), pp. 89-103

Votey, H.L.Jr and Phillips L., (1972), Police Effectiveness and the Production Function for Law Enforcement, *The Journal of Legal Studies*, 1, pp. 423-436