Local Enforcement Externalities and the Long Run Evolution of Tax Compliance in Italy

Antonio Acconcia* Marcello D'Amato† Riccardo Martina ‡
Marisa Ratto § Preliminary draft. Please do not circulate.
¶

May 22, 2019

Abstract

We study a model of tax evasion dynamics in the presence of an enforcement externality and social learning by taxpayers. Conditions under which the distribution of the perceived probability of apprehension and compliance behavior converge to one or multiple steady states are studied. We show that the emergence of long run history dependence crucially depends on the level of the tax rate compared to a measure of efficiency in the (local) enforcement.

The insights obtained from the model are used to interpret the high level of geographical dispersion of tax evasion and its persistence in Italy as consequences of an unanticipated and substantial tax raise and of the local dimension of tax jurisdiction since unification. We finally use a novel data set from the Italian Revenue Agency and historical data from the Ministry of Finance in 1870 to show that persistent geographical dispersion in tax evasion is consistent with an institutional trap by the unification process.

Keywords: tax evasion, local enforcement externality, learning, tax compliance dynamics.
JEL: D62, D81, H26, K41, K42.

*Università di Napoli Federico II, (DiSES), Csef, e-mail: acconcia@unina.it
†Università di Napoli Suor Orsola Benincasa, (Csef, Celpe), via Suor Orsola 10, 80121, Napoli, Italy, e-mail: damato@unisa.it.
‡Università di Napoli Federico II, (DiSES), Csef, e-mail: riccardo.martina@unina.it
§University Paris-Dauphine (LEDa), e-mail: maria_luisa.ratto@dauphine.fr
¶We thank Stefano Pisani and Andrea Spingola at the Agenzia delle Entrate (Ufficio Studi, Roma) for helping to secure the data for us.

"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

Long-run persistence of individual behavior has long attracted the attention of economists and law scholars. In this paper the case of tax compliance is investigated theoretically and empirically. The research questions we intend to address are related to the evolution of tax compliance in Italy.

In particular, we are interested to explain why tax evasion can persistently (meaning more than a century) differ across geographical areas of a unified nation and whether Italy is a case in point and why. We focus on two issues: geographic dispersion in tax compliance and its persistence, despite of a nation-wide tax code. From a positive point of view we aim to establish under which conditions geographically heterogeneous compliance levels are likely to emerge and persist in the long run. Is it about culture, social capital or social norms? Is it about genetics? Is it about political institutions? Or is it the response to the design of apparently minor administrative details, such as a problem of decentralized enforcement? These questions and the historical example seem to us quite relevant for two reasons. Convincing answers could, firstly, help us in understanding the long term consequences of the design of enforcement procedures in fiscal legislation by nation states and in implementing their reforms. Secondly, they could also set a frame for the assessment of the difficulties involved in the process of nation building or political unification arising from economic integration.

To address these questions we set up a simple dynamic model characterized by two main assumptions: an enforcement externality arising from limited amount of resources available to fiscal jurisdiction and a social learning process determining the perceptions of the probability of conviction.

Due to the enforcement externality the actual level deterrence, as measured by level of sanctions for evasion actually administered by the fiscal jurisdiction, depends on the amount of evasion in each district. Individual behavior and actual evasion rates in each district, on the other hand, depend, among other elements, on the perceptions by the tax payers about the sternness of the enforcement system they experienced in the past.

We allow individual perceptions to evolve over time according to a social learning rule that also involves the acquisition of a signal about the actual frequency of punishment that occurred at the aggregate level. In this setting, individual tax compliance is affected by current individual perceptions that incorporate information about local enforcement levels in the past.

More specifically we shall consider a simple model of individual tax compliance where private agents, given their information, make their choice based on

their perception of the probability that non compliance will be sanctioned, on their individual incentives as set by the income tax rate, by the level of the fines and by an individual cultural trait that measures attitudes towards the state, i.e. tax morale. Hence cultural traits, by shaping the aggregate behavior, contribute to the dynamics of the social learning. Clearly then, current choices are affected by current individual perceptions that incorporate information about aggregate enforcement levels in the past. So, also the current aggregate behavior affects future compliance through the enforcement externality.

In this setting the evolution of perceptions (and evasion behaviors) at local level are simultaneously determined along with the equilibrium level of deterrence locally provided by the state and both depend on the nationwide tax rate enshrined in the tax code. Indeed the mutual relationship between the level of enforcement and evasion behavior in each fiscal district is a dynamic one, so the current distribution of perceptions in a given area will depend on the past distribution of perceptions in that same area. The theoretical analysis identifies conditions under which geographically persistent patterns of tax compliance may emerge. We show that the equilibrium distribution of taxpayers perception and behavior converge to a steady state distribution, possibly depending on the initial distribution. In particular, for high enough tax rates—not so high, however, to disrupt compliance—both high and low evasion equilibrium becomes part of the set of steady states. In this case the specific outcome the economy approaches to is determined by the initial perception at local level. Under these circumstances differences in the initial perceptions and attitudes by taxpayers at local level may be perpetuated or even magnified into the future. The model shows that an increase in the level of taxation can get initial, even small, differences in the perceptions of the enforcement system, which are then magnified in the dynamic process of social learning and become persistent at local level.

The historical evidence we provide on the way the tax system was designed in Italy since unification (1870) supports the insights of our theoretical model. In particular, experts at the time pointed out that tax evasion was a main issue for the newly born country since unification and the main factors thought of being responsible for that were the high tax rates and the local design of the enforcement process. We use data on tax evasion at province level in 1870, as registered by the Ministry of Finance at that time and in 2006, as officially assessed by the Italian Tax Revenue Agency, and we find a positive correlation between the historical and current measures of tax evasion, which turn to be robust to controls that account for structural characteristics of the provinces and for the efficiency of the current enforcement system. In an attempt to explain the initial level of tax compliance we exploit the sudden and abrupt increase in taxes that characterized the aftermath of the Italian unification. Since at that time the government needed to increase the tax revenue (without increasing the public debt) as well as to homogenize the tax burden across areas of the country that fomerly were part of different states and kingdom, the tax burden was differently increased spatially, mainly through the new income tax and the reform of the land tax. Overall the burden of the taxes in the unified nation became quite heavy when compared to those in other European country,

due to the fiscal adjustment required by the large public debt accumulated in the military campaigns for unification in the previous decades. Arguably, an index capturing such differential spatial increments provides a good instrument for the historical measure of tax evasion. We find that this is indeed the case: the historical measure of tax evasion is strongly positively correlated with the index of tax change realized by the first government of the new-born Kingdom of Italy. When the index is used as an instrumental variable for historical evasion then the long-run persistence of tax evasion is fully confirmed. We conclude that the empirical evidence gives support to the intrepretation that persistence of geographical dispersion in tax evasion is consistent with an institutional trap originated by relative large tax rates and by a decentralized system of fiscal jurisdiction.

Several contributions have underscored the importance of the enforcement externality (see, among others, Becker 1968, Ehrlich 1973, Votey and Phillips 1972).[1] In particular, Ehrlich (1973) introduces the assumption that the productivity of the resources allocated to law enforcement is lower the higher the level of criminal activity. The evolution of perceptions in the presence of an enforcement externality has been studied in Sah (1991), who focuses on a model with a unique stable equilibrium. We stress, instead, the convergence process of individual perceptions when the institutional setting allows for an important role of history in selecting the relevant dynamic path and on long-term persistence of perceptions and behavior.[2] The specific focus of our study is about the local dimension of the enforcement externality and its ability to persistently support heterogeneous patterns of non compliance even in the presence of a nationwide tax code.[3]

Both the model and the empirical analysis is related to the literature on the long run persistence of behavior, attitudes, culture and expectations as surveyed by Acemoglu et al. (2005). Compared to other contributions (e.g. Guiso et al. 2016) rather than local culture, we emphasize the long run consequences of apparently small details in the legislation.[4] It is possible- and perfectly consistent with our model- that initial cultural or even ideological attitudes towards the Piemontese monarchy who unified the state got perpetuated and affected fiscal compliance behavior at the local level. We agree that there are good reasons to believe that mechanisms of cultural transmission (within the households or other social mechanisms in place for the reproduction of trust and social capital) are important. However we also believe that, in the attempt

[1] Shavell (2014) provides a survey on mechanisms of deterrence.

[2] The importance of individual perceptions for crime participation has been shown by Lochner (2007).

[3] Acemoglu and Jackson (2017) consider the interplay between social norms and private cooperation with law enforcement provided by the members of a society. However the mechanism that sustains dynamic persistence in that paper is quite different, since it is based on the matching by individuals of different generations. Some of the mechanisms on which these authors focus could be applied to tax evasion and could also be at work in the historical example we consider. Here, we focus on the learning by agents since it seems more appropriate for the case in which individuals decide about tax compliance compared to law breaking in the case of criminal law or offenses of administrative law.

[4] In particular the composition of local fiscal courts and the decentralized trait of the fiscal jurisdiction and tax collection in Italy.

to provide an explanation for the persistence of social behavior regulated by law and assisted by a formal sanction, the institutional design that shape the deterrence value of the sanctions in a context with learning is relevant, both from the theoretical and the empirical point of view. In this respect our investigation and results are closer to the view that long run persistence of tax compliance is more related to what has been called "institutions". In the case of fiscal compliance it turns out that apparently minor details of the organization of the fiscal jurisdiction and tax collection - the mode of their decentralization in particular - can play an important role.

The paper is organized as follows: Section 2 presents the model and the results. In Section 3 we provide an interpretation for the high level of tax evasion and its persistence in Italy since unification (1870) in the light of our model implications. In Section 4 we provide an empirical analysis using novel data from the Italian Revenue Agency and some historical data in the period 1868-1870. Section 5 concludes.

2 Model

To study the consequence of decentralized enforcement on the persistent geographical dispersion of tax compliance we consider a country whose territory is divided in n districts denoted by d, each district being populated by a unit measure of individuals denoted by i with exogenous income Y_i^d who have to report a fraction of their income $\delta_i^d \in [0,1]$.

A nationwide tax code is in place, whose essential features are the income tax rate denoted by τ and the fines, denoted by ϕ, to be administered in case evasion by a non compliant taxpayer is proved in court.

In our model detection of tax evasion, which is the result of an audit policy, does not necessarily induce a sanction on detected tax evaders, as in the standard model of tax evasion. A fraction of tax reports $a \in [0,1]$ is audited and, for the sake of simplicity, it is assumed to be the same in all districts and so it does not depend on d. In the event that evasion has been found by the auditor, the case goes to a fiscal trial, which is instead decentralized at the district level. Only a fraction θ^d of the trials is successfully finalized by the local fiscal court and the taxpayer is sentenced to pay fines in proportion of the unreported income tax $\phi\tau(1-\delta_i)Y_i$.

There exists an enforcement externality in fiscal jurisdiction such that the fraction of successfully finalized fiscal trials (and hence the probability of punishment governing compliance behavior) depends negatively on the aggregate level of cases of evasion in that district denoted e_t^d, so $\theta_t^d = \theta(e_t^d)$.

Therefore the actual probability that fines will be enforced, r_t^d faced by the taxpayer in district d is given by the joint probability that the taxpayer is audited and that a sentence is inflicted. Formally, in each period t, it holds:

$$r_t^d = a\theta(e_t^d) \tag{1}$$

where $\theta(x)$ is a function such that $x'' \geq x'$ implies $\theta_t^d(x'') \leq \theta_t^d(x')$. This property is assumed to arise from a congestion externality that can originate

from decreasing returns to the activity of jurisdiction but also as a consequence of the local political economy in a system where the fiscal court is appointed by local governments and therefore reflects attitudes by local taxpayers towards the fiscal state: the larger the fraction of non compliant agents in a given district, the larger the possibility that a more lenient attitude in the local fiscal court will emerge. To keep the analysis simple we further restrict the function $\theta(e_t^d)$ at eq. (1) and we suppose that the congestion effect is triggered when the evasion is above a fixed threshold $\tilde{e}_d$. So the real probability that fines are enforced in the fiscal system is

$$r_t^d = a\theta(e_t^d) = \begin{cases} a\overline{\theta}^d & \text{for } e_t^d \leq \widetilde{e}^d \\ a\underline{\theta}^d & \text{for } e_t^d > \widetilde{e}^d, \end{cases} \tag{2}$$

where $\overline{\theta}^d > \underline{\theta}^d > 0$ and $\widetilde{e}^d > 0$ are fixed exogenously.

For the ensuing analysis it is relevant to notice that the key simplifying aspect of our assumption about the enforcement externality built into the analytic properties of eq. (2) is not in the specific functional form of the threshold externality. It is rather the fact that the actual level of deterrence (r_t) will only depend on the average non compliance [5] prevailing in each district.

The real probability that a fine for evasion is inflicted, however, is not observed by the taxpayer, who takes her decision on the basis of her perception $\widehat{p}_{i,t}^d$ about r_t^d. This perception evolves over time according to an adaptive learning process defined as follows: $\widehat{p}_{i,t}^d = \mathbb{E}[r_t^d \mid I_{i,t}^d] \equiv L_i^d(p_{i,t-1}^d; s_{i,t}^d)$.

To simplify the analysis we assume that L_i^d to be as follows:

$$\widehat{p}_{i,t}^d = L_i^d(\widehat{p}_{i,t-1}^d, s_{i,t}) = \alpha\widehat{p}_{i,t-1}^d + (1-\alpha)s_{i,t}, \tag{3}$$

where α is weight on past experience[6] and $(1-\alpha)$ is the weight on new information coming from a non distorted signal $s_{i,t}$ of the actual frequency of apprehension prevailing in the district of residence in the previous period. In particular the signal $s_{i,t}$ depends on the level of deterrence actually enforced in the previous period as follows:

$$s_{i,t} = r_{t-1} + \eta_{i,t} \tag{4}$$

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 $Var(\eta_i) = \sigma_\eta^2$. Notice that the individual signal is undistorted in the sense that the signal on the average frequency of punishment in the past period is correct on average among the taxpayers.

The dynamics of individual perceptions can therefore be written as

$$\widehat{p}_{i,t} = \alpha\widehat{p}_{i,t-1} + (1-\alpha)(r_{t-1} + \eta_{i,t}) \tag{5}$$

[5] Hence it will depend on the first moment of the distribution of perceptions that taxpayers hold about r_t^d.

[6] Notice that this formulation is not inconsistent with Bayesian learning rule (see Sah, 1991), where weights evolve according to the accumulation of experience following the Bayes rule. Our choice of time invariant α greatly simplifies the dynamic analysis. As an alternative specification of adaptive mechanisms see the literature on OLS learning (Honkapoja 2010).

This learning process formalizes the idea that individual i uses past experience and she acquires information (at a negligeable fixed cost by direct sampling the experience among peers in period $t-1$ or from informed consultants) about the average enforcement level prevailing in district d in the past. Notice that this representation formalizes the notion of an externality since individual perceptions (and hence compliance behavior) evolve based on the aggregate past level of enforcement.

The mappings $\theta(e_t^d)$ and $L_i^d(p_{i,t-1}^d, s_{i,t}^d)$ illustrate the mutual relationship between social learning and the enforcement externality. Any positive shock to aggregate evasion in any period will affect signals that all individual agents will receive in the next period: at any period t the current distribution of perceptions (and induced compliance behavior) will depend on the past distribution of perceptions due to the learning rule.

This distribution of perception $\widehat{p}_{i,t}^d$ prevailing among the taxpayers in district d at time t is denoted by $\mathbb{P}_t^d$. It will evolve according to the evolution of individual perceptions defined by the learning rule and the enforcement externality, starting from an initial distribution $\mathbb{P}_0^d$.

A dynamic equilibrium of the model consists, therefore, of a sequence of actual enforcement levels r_t^d and an associated distribution of perceptions $\mathbb{P}_t^d$ starting from an initial condition r_0^d and $\mathbb{P}_0^d$, with a steady state being denoted by r_*^d and $\mathbb{P}_*^d$.

It is important to notice that in the proposed formulation, by complete decentralization of the fiscal jurisdiction, the level of enforcement in each district d only depends on the distribution of perceptions and behavior by taxpayers in that district. On the other hand, given that only local jurisdiction matters, taxpayers will only have incentives to learn about the probability of a sanction being inflicted by a local court. In other words due to complete decentralization of jurisdiction, the evolution of tax compliance in each of the districts does not interact with each other and it only depends on the initial conditions and local features of fiscal jurisdiction: if history matters or not may depend on aggregate parameters of the fiscal system like the tax rate or the fines, but, if it does, only local history matters. For this reason we drop the superscript d henceforth from all the variables of the model and only reintroduce it when needed for the discussion of the empirical implications of the model.

Summarizing the main features of the fiscal system, we denote the exogenous elements of the model by $\Gamma = \{Y_i, \tau, a_i, \theta, L_i, r_0, \mathbb{P}_0(\widehat{p}_i)\}$, the analysis will focus on endogenous variables defined by the triple $\{\mathbb{P}_t(\widehat{p}_i), e_t, r_t\}$ describing the evolution of perception compliance and enforcement levels, i.e. the elements of a fiscal system, at the district level.

In the following we describe the tax evasion choice by each agent i in a given district for a given distribution of perceptions $\mathbb{P}_t(\widehat{p}_i)$, then we illustrate the details of the learning rule and the externality in the fiscal jurisdiction. Then the dynamics of individual perceptions will be characterized along with the convergence of the distribution of perceptions and provide the results on the convergence of $\mathbb{P}_t$ to its limit distribution $\mathbb{P}_*$. Additional specific parametric assumptions will be made to keep the model tractable. These parametric assumptions will pertain to the utility function used by agents to decide

compliance. Subscript t will only be considered when relevant, i.e. in the characterization of the dynamics.

2.1 Tax evasion

Given the institutional framework described above, tax evasion is decided by each taxpayer i depending on the size of the tax rate, 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.

Notice that in the above setting neither a (the probability of an audit) nor θ_t (the probability that non compliance is fined in a fiscal trial) depend on the individual history of compliance.[7]

Hence taxpayer i in district d as of time t (the latter two subscripts dropped), perceiving a probability $\widehat{p}_i$ of apprehension, will solve the following problem:

$$\underset{\delta_i \in [0,1]}{Max} \quad U(\delta_i) \equiv (1-\widehat{p}_i)Y_i(1-\tau\delta_i) + \widehat{p}_i(Y_i - \tau\delta_i Y_i - \phi\tau Y_i(1-\delta_i)) + \varepsilon_i(\delta_i) \quad (6)$$

Where $\delta_i \in [0,1]$ denotes the percentage of income reported by taxpayer i, Y_i denotes her income, $\widehat{p}_i$ denotes her perceived probability of apprehension in case of evasion, τ denotes the tax rate, ϕ denotes the fine proportional to evasion in case tax evasion is discovered by the audit and successfully finalized by the fiscal court.

The parameter ε_i measures the additional utility from compliance measuring the individual tax morale. We assume $\varepsilon_i(\delta_i) = \varepsilon_i\delta_i Y_i$ so that the additional utility associated to tax morale is proportional to the tax bill. Notice that that $U(\delta_i)$ is linear as a consequence. Moreover ε_i is taken uniformly distributed on $[0, \bar{\varepsilon}]$. Finally, to simplify the analysis, we assume $\phi\tau = 1$ (maximal fine).

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

Proof. Linearity of eq. (6) and $\delta \in [0,1]$ imply that the solution exists, it is unique and it can only be at the boundaries, i.e. $\delta_i \in \{0,1\}$. Specifically, 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 $U(\delta_i = 1) = (1-\tau)Y_i + \varepsilon_i Y_i$, it is immediate to see that $\delta_i = 0$ obtains if $\varepsilon_i \leq (\tau - \widehat{p}_i)$ and $\delta_i = 1$ otherwise.[8] ■

[7] This is equivalent to assume that the cost of current evasion being discovered has no future consequence neither in terms of liabilities on future streams of income, nor for the probability of future audits or convictions. In a more sophisticated model of tax evasion in which these feature had been modeled, the source of persistence due to learning and the consequences of the externality for convergence would not be eliminated. Clearly the dynamics of the equilibrium distribution of perceptions and enforcement would be more complex. No additional insights on the issue under investigation would emerge if a more effective auditing strategy was implemented at the nation-wide level.

[8] Notice that for $\widehat{p}_{i,t} > \tau$ the incentive effect of expected punishment is large enough to guarantee full compliance by those individuals with such a perception. Although no particular problem would arise if this possibility is allowed (just take $\varepsilon_i \leq min\{0, (\tau - \widehat{p}_i)\}$. No problems

Lemma 1 above characterizes individual compliance for a given individual perception $\widehat{p}_i$. Intuitively, given a perception $\widehat{p}_i$, an individual with low enough (moral) benefit from compliance will evade the whole income. For given fiscal morality the incentive to evade are increasing in the tax rate and decreasing in the perceived probability of apprehension.

Before turning to the dynamics of perception, for later use, we compute here the aggregate compliance rate for a given $\mathbb{P}_t(\widehat{p}_i)$, with an average being denoted by $\widehat{p}_{m,t}$. Then the average non compliance rate can easily be obtained as follows:

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

$$e_t = \frac{(\tau - \widehat{p}_{m,t})}{\bar{\varepsilon}}. \tag{7}$$

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

Proof. From Lemma 1 we know that the tax evasion decision is $\delta_i = 0$ if and only if $\varepsilon_i \leq (\tau - \widehat{p}_{i,t})$. So the measure of non compliant taxpayers in a district will be given by $e_t = \Pr[\varepsilon_i \leq (\tau - \widehat{p}_i)]$, where ε_i is uniformly distributed on $[0, \bar{\varepsilon}]$ by assumption. Therefore $e_t = \int_{\widehat{p}_i=0}^{1} \frac{(\tau - \widehat{p}_i)}{\bar{\varepsilon}} d\mathbb{P}_t(\widehat{p}_i)$ and eq. (7) follows ■

Eq. (7) describes the equilibrium aggregate evasion rate corresponding to an average perception computed for a given $\mathbb{P}_t$. Notice that for $\widehat{p}_{m,t} \to 0$, $e_t \to \frac{\tau}{\bar{\varepsilon}}$, so that average compliance does not disappear even in the presence of no enforcement due to the presence of moral benefits from compliance.

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

Summarizing, Lemma 1 and Lemma 2 describe individual evasion choices as a function of individual perceptions $\widehat{p}_i$ in each district and the average evasion rate as a function of the average perception for any given probability distribution $\mathbb{P}_t$. This latter is to be determined in the equilibrium dynamics to be studied in the following.

2.2 The dynamics of the distributions of perceptions

We move now to the analysis of 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 and in particular by its average $\widehat{p}_{m,t-1}$ with the associated tax evasion decision by each taxpayer in the population and the ensuing aggregate evasion rate e_{t-1}, for the exogenously given learning rule and the signal, $s_{i,t}$, obtained by each agent i as of time t.

In the following we start with the formal description of the stochastic process followed by the taxpayers' perceptions about the probability of apprehension,

with atoms in the equilibrium distribution of perceptions should arise since the signal has a diffuse support. In any case conditions can be provided to avoid this case so that the average perception is strictly interior,i.e $\widehat{p}_{m,t} \in (0, \tau)$ and hence $e_t \in (0, 1)$.

then we prove existence and convergence of the sequence of $\mathbb{P}_t(\widehat{p}_i)$ to $\mathbb{P}^*(\widehat{p}_i)$, starting from an arbitrary initial distribution $\mathbb{P}_0(\widehat{p}_i)$. The analysis proceeds by adapting standard results in the theory of Markov Process as in Stockey, Lucas and Prescott (1989), SLP (1989) henceforth.

Let us start by noticing that the transitions of individual perceptions are driven by the lagged aggregate level, r_{t-1}. By using equation (2) lagged one period, $\widehat{p}_{i,t}$ can be written as:

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

By recovering $\widehat{p}_{m,t-1}$ from (7) and by replacing its value into the equation above the stochastic difference equation can be written as follows:

$$\widehat{p}_{i,t} = \Phi(\cdot) = \begin{cases} \Phi_h = \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\overline{\theta} + \eta_{i,t}) & \text{for } \widehat{p}_{m,t-1} \geq \widetilde{p} \\ \Phi_l = \alpha\widehat{p}_{i,t-1} + (1-\alpha)(a\underline{\theta} + \eta_{i,t}) & \text{for } \widehat{p}_{m,t-1} < \widetilde{p}, \end{cases} \tag{8}$$

where $\widetilde{p} = \tau - \overline{\varepsilon}\widetilde{e}$.

Hence the dynamics of individual perceptions are described by the stochastic difference equation in (8) as a function of their past value $\widehat{p}_{i,t-1}$ and the signal given by (4) according to the past average perception $\widehat{p}_{m,t-1}$ and the idiosyncratic component $\eta_{i,t}$.

To avoid uninteresting cases we will assume that $\overline{\varepsilon}$ and $\widetilde{e}$ are such that $\widetilde{p} \in (0,1)$. The notation emphasizes that Φ is a stepwise linear correspondence $\widehat{p}_{i,t-1} \to \Phi_i(\widehat{p}_{i,t-1}, .)$ mapping the set of equilibrium perceptions achievable by agent i at time t starting from 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 in the previous period. For future reference we denote $\overline{\Phi}_h \equiv \max_{\eta} \Phi_h(.)$, $\underline{\Phi}_h \equiv \min_{\eta} \Phi_h(.)$, $\overline{\Phi}_l \equiv \max_{\eta} \Phi_l(.)$ and $\underline{\Phi}_h \equiv \min_{\eta} \Phi_h(.)$. A graphical representation of this difference equation is provided at Figure 1.

Figure 1 here

Suppose, for the moment, that $\overline{\eta}$ is small and it satisfies the following condition $\overline{\Phi}_l(\widetilde{p}) \leq \underline{\Phi}_h(\widetilde{p})$ or $\alpha(\overline{\theta} - \underline{\theta}) \geq \overline{\eta} - \underline{\eta} = 2\overline{\eta}$.

Clearly, the stochastic difference equation defined by (8) satisfies the conditions of theorem 8.9 in SLP (1989) and hence it defines a transition function for the Markov process denoted by $Q(p, S)$[9]. The transition function defines the probability that individual taxpayers end up having perception $x' \in S$ starting from perception $\widehat{p}_{i,t-1} = x$, with Q satisfying the usual definition. [10]

[9] A transition probability on the state space $\widehat{p}_i \in Z$ is a function $Q : Z \times \mathcal{Z} \to [0,1]$ 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.

[10] See Theorem 8.9 SLP (1989). The only non negligible difference here is that Q at each period depends on r_{t-1} in general. However, the structure of the enforcement externality as defined at eq. (2) clearly implies that for any given $\mathbb{P}_0$ the standard definition applies. Indeed, if $\mathbb{P}_0$ is such that $p_{m,0} \leq \tau - \widetilde{e}$ then $\widehat{p}_{i,t} = \Phi_h(\widehat{p}_{i,t-1}, \eta_{i,t}; r_{t-1} = \underline{\theta})$ for all $t = 0, 1, 2, ...$; on the other hand if $\mathbb{P}_0$ is such that $p_{m,0} > \tau - \widetilde{e}$ then $\widehat{p}_{i,t} = \Phi_l(\widehat{p}_{i,t-1}, \eta_{i,t}; r_{t-1} = \overline{\theta})$ for all $t = 0, 1, 2, ...$

It follows that the sequence of distribution functions $\mathbb{P}_t$ satisfies the following recursive relation

$$\mathbb{P}_t(x') = \int Q(x', x; r_{t-1}) d\mathbb{P}_{t-1}(x) \tag{9}$$

starting from any initial $\mathbb{P}_0$.

Hence, an equilibrium distribution of perceptions as of time $t+1$ is a probability measure satisfying $\mathbb{P}_{t+1} = T^*\mathbb{P}_t$ where T^* is the self adjoint Markov operator associated to the transition $Q(p, S)$. $\mathbb{P}_t$ may converge to an invariant distribution, possibly depending on $\mathbb{P}_0$. A steady state of the fiscal system (invariant distribution of perceptions and induced evasion behavior) is a measure $\mathbb{P}$ on $[0, 1]$ satisfying $\mathbb{P}^* = T^*\mathbb{P}^*$. Endowed with this definition we will now provide some results about the existence of $\mathbb{P}^*$, its convergence and the dependence of the dynamics on $\mathbb{P}^0$.

In the next subsection we first discuss the steady state assuming that $\mathbb{P}^*$ exists, then we will prove and characterize convergence of the distribution of perceptions $\mathbb{P}_t$ to $\mathbb{P}^*$.

2.3 Steady States and convergence of perceptions. The role of local history

Suppose, for the moment, that the fiscal system converges to a distribution of perceptions (not necessarily unique, for any initial distribution $\mathbb{P}_0$) 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 at steady state. The steady state level of the average perceptions must satisfy $\widehat{p}^*_m = \widehat{p}_{m,t} = \widehat{p}_{m,t-1}$ and from equation (8) it necessarily holds:

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

with $\tilde{p} = \tau - \bar{\varepsilon}\tilde{e}$. Then it follows

$$\widehat{p}^*_m = \begin{cases} a\overline{\theta} & \text{for } \widehat{p}^*_m > \widetilde{p} \\ a\underline{\theta} & \text{for } \widehat{p}^*_m \leq \widetilde{p}. \end{cases} \tag{10}$$

From the agents' optimal evasion choice at equation (7) the steady state aggregate evasion rate is given by $e^* = \frac{(\tau - \widehat{p}^*_m)}{\bar{\varepsilon}}$ and hence, by definition of $\widehat{p}^*_m$, it necessarily holds

$$e^* = \begin{cases} e^*_l = \frac{(\tau - a\overline{\theta})}{\bar{\varepsilon}} & \text{for } \widehat{p}^*_m > \widetilde{p} \\ e^*_h = \frac{(\tau - a\underline{\theta})}{\bar{\varepsilon}} & \text{for } \widehat{p}^*_m \leq \widetilde{p} \end{cases} \tag{11}$$

Moreover, the steady state probability of apprehension is obtained by setting $r_t = r_{t-1}$ at equation (2)

$$r^* = \begin{cases} a\overline{\theta} & \text{for } \widehat{p}^*_m > \widetilde{p} \\ a\underline{\theta} & \text{for } \widehat{p}^*_m \leq \widetilde{p} \end{cases} \tag{12}$$

From equations (10), (11) and (12) it follows that there are at most two steady states. The result 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* $\overline{p}^*_m = r^* = a\overline{\theta}$ *and* $\underline{p}^*_m = r^* = a\underline{\theta}$ *and at most two associated aggregate evasion rates* e^* *given by eq. (11). Depending on the tax rate* (τ) *and the enforcement system* $(\widetilde{e}, a, \theta\ )$ *there are three possible regimes:*

i) low tax level: if $\tau \in [0, \bar{\varepsilon}\widetilde{e} + a\underline{\theta}]$ *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 $\tau \in [\bar{\varepsilon}\widetilde{e} + a\underline{\theta}, \bar{\varepsilon}\widetilde{e} + a\overline{\theta}]$ *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: $\tau \in [\bar{\varepsilon}\widetilde{e} + a\overline{\theta}, 1]$ *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 state aggregate evasion rates exist clearly follows from (11). 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 fixed points of the function

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

Where it is easy to see that for $\tau \in [0, \bar{\varepsilon}\widetilde{e} + a\underline{\theta}]$ only $\overline{p}^*_m = a\overline{\theta}$ satisfies $f(a\overline{\theta})$ and hence $e = e^*_l = (\tau - a\overline{\theta})\bar{\varepsilon}\widetilde{e}$. For $\tau \in [\bar{\varepsilon}\widetilde{e} + a\overline{\theta}, 1]$ only $\underline{p}^*_m = a\underline{\theta}$ satisfies $f(a\underline{\theta})$ and $e = e^*_h = (\tau - a\underline{\theta})\widetilde{e}$. Finally, for $\tau \in [\bar{\varepsilon}\widetilde{e} + a\underline{\theta}, \bar{\varepsilon}\widetilde{e} + a\overline{\theta}]$ 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 (11) 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 distributions $\mathbb{P}^*$. ■

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 can be consistent with steady state; which one of the two will privail as a selected equilibrium will depend on historical initial conditions described by $\mathbb{P}_0$ or by the permanent effects of transitory shocks to local parameters measuring tax morale ($\bar{\varepsilon}$), and the measure of local efficiency of the enforcement system ($\tilde{e}$).

These results capture the intuition associated to the enforcement externality as in the quote in the introduction, 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 tax courts 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 and the role of history becomes crucial of the selection of the relevant equilibrium.

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 a rational expectation equilibrium, 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.

Our next task is to prove that, indeed, the learning model with the enforcement externality converges to the rational expectations equilibrium over time, as characterized in Proposition (1). In the following we prove that the Markov process defined by the stochastic differential equation at equation (8) converges to $\mathbb{P}^*$ starting from any $\mathbb{P}^0$, in all of the three regimes in proposition (1) as defined by the values of the parameters of the fiscal system. 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 are induced by the fiscal enforcement system will depend on the initial distribution of perceptions $\mathbb{P}_0$.

Proposition 2 *Starting from an arbitrary distribution of perceptions $\mathbb{P}_0(\widehat{p}_i)$, the Markov process defined by Φ at equation (8) will converge to a unique $\mathbb{P}^*$, possibly depending on $\mathbb{P}_0$.*

Proof. We proceed by proving the result in a few steps. That the stochastic differential equation defined by Φ defines a Markov process was already stated in the text as an immediate consequence of theorem 8.9 in SLP (1989). Although individual transitions depend on the past distribution the key observation is that, depending on $\mathbb{P}_0$ and hence on initial average $\widehat{p}_{m,0}$, dynamic transitions for individual perceptions will be defined either by Φ_h or by Φ_l. In particular if $\mathbb{P}_0$ is such that $\widehat{p}_{m,0} \leq \widetilde{p}$ then individual transitions at eq. (8) will be governed by Φ_l for all $t > 0$; on the other hand if $\widehat{p}_{m,0} > \widetilde{p}$ then individual transitions at eq. (8) will be governed by Φ_h for all $t > 0$. Clearly (see fig.1) both Φ_l and Φ_h satisfy conditions for contraction and condition M in SLP p.348 for convergence, hold in both cases. Whether ergodicity or history dependency obtains depending on the parameters of the system depending on the value $\tilde{p}$ as in the three cases in Proposition 1. The details of the proof for each of the three cases is immediate and it is not reported for brevity. ■

The case where equilibrium multiplicity can arise is particularly relevant for our investigation. The model shows that an enforcement externality in fiscal jurisdiction arising at the local level can support long-run persistence of geographic dispersion of evasion rates that is likely to emerge if the tax rate is

large enough (not so large to disrupt the fiscal state all together) compared to a measure of administrative efficiency.

2.4 Convergence when the support of η_i is large: the role of experts.

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 $\overline{\Phi}_l(\widetilde{p}) \leq \underline{\Phi}_h(\widetilde{p})$ 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 around its average is not enough precise.

This situation can occur either when the information technology of the taxpayers features lower precision or when there is an intrinsically large noise component in the announcement that the fiscal authority makes about the enforcement system, so that taxpayers strive to learn about the probability of punishment. This is interesting to emphasize since it highlights the role of experts in the reinforcement of the fiscal externality. More precisely experts provide information to taxpayers, more likely is the possibility that the enforcement externality will make history dependent on initial conditions.

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 $\overline{\Phi}_l(\widetilde{p}) > \underline{\Phi}_h(\widetilde{p})$. 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 (9). Consider then case II in Proposition (1) then for $\overline{\eta} \geq \frac{\alpha(\bar{\theta}-\underline{\theta})}{2}$ 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 the above result easily follows from general principles in Markov processes they highlight an interesting role of experts, or more generally of the factors that influence the precision of the signal in the learning process. 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 A historical riconstruction of the Italian fiscal administration: a primer on taxes in the newborn kingdom of Italy (to be completed)

In this section we provide a reconstruction of the institutional traits of the fiscal system created in Italy after unification. It will be documented the high level and the geographical dispersion of tax evasion since unification. As we shall consider, the experts of the time point to the high level of taxes and to the local inefficient organization of enforcement as the main explanations for the high levels of non compliance. This historical evidence lends support to the main insights of our theoretical model. Namely the prediction that history matters for the equilibrium outcomes if the tax rate is large enough compared to a measure of the effectiveness of audits and the prediction that if the initial dispersion in tax evasion is large enough, such a geograpical dispersion will be persistent. We shall further investigate the relationship between tax evasion in 1870 and tax evasion in 2006 in the following section.

Italy was unified in 1861. This was the final outcome of a social, political and military campaign deeply rooted in the geopolitics of Europe in the XIXth century. The process was led by the Savoy monarchy based in the North West (Piemonte) and subsequently affected regions in the Centre and in the South (Kingdom of Naples and the Papal State). The socioeconomic 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.

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. 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 the fiscal regime of Piemonte (part of the Sardinia Kingdom) to the newly born country was seen as a necessary step toward economic integration for a number of reasons: it was felt that a modern state had to rely on a sufficiently larger amount of resources in order to be able to finance external debt (this was the case of Piemonte but not, however, of the Two Sicilies Kingdom - the south of Italy). It was also thought that, even if the regional differences were clearly recognizable, the homogeneity of fiscal duties would have reinforced the faith in the newly born State. 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 decentralized 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.

One important characteristic of the fiscal regime which was introduced in 1864, based on a tax on "Ricchezza Mobile", the first form of income tax, 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. The tax was an apportioned one, in order to secure a definite revenue. The system was such that, given the total amount of fiscal revenue the government aimed to raise, then the amount was apportioned among different provinces. Each provincial quota was apportoned in the same way among the municipalities, and the amounts were then levied upon individuals according to their income. In 1866 the tax was changed to a percentage tax, the rate being made 8.8 percent. In 1869 the tax rate was 13.2%, compared to a rate of 2.46% in the UK in the same year, and at the beginning of the XX century it reached 20 percent.In his analysis of the Italian Fiscal system Giulio Alessio (1883) has an interesting discussion of tax evasion. Figures reported at 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 452 millions Liras of which 266 are wages to public employees. By adding wages to public employees in local administrations, the total 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 reveals high level of tax evasion. According to the author, the causes of tax evasion are to be identified in the large tax rate (and very low exemption level of income) and in the local organization of enforcement. For the enforcement, the collection of disputed unpaid taxes was delegated to "Commissioni di Accertamento" (the original form of a modern tax court), which originally had both initiative in the estimation of individual income and a role of inspection. Despite the original aim to design a decentralized system to create interactions between local authorities and taxpayers, these local commissions made the fight against tax evasion more difficult. 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) and the others were elected by the Town Councils, where rich land owners and tradesmen predominated among those who had the right to vote. Hence there were important conflicts of interest, as the members of these commissions represented the interests of rich land owners and tradesmen, the

same categories of taxpayers engaged with widespread tax evasion. Lobbying to affect enforcement was also made possible by the fact that the members of the fiscal commissions were in charge for several mandates, with no bans on re-election. Also, there were very limited means to verify declarations, above all for individual businesses and professionals, who were not required to provide a balance-sheet. 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. Hence this evidence suggests that the culprit of high tax evasion observed since unification is rooted in high tax rates and in flaws in the administration of the enforcement process.

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). 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 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 having 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 Trade Unions and by the Provincial Council presided by the prefect (Lignani 2005, p.61).

In the Republican era the rules were formally changed again: mayors 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 first important reform in the Republican era was that of 1972 where the system of appeals (ordinary law courts and fiscal law courts) was disciplined. No major changes occurred in the composition and in the appointment of the judges in the Commussione Tibutaria and in the Commissione Povinciale (see Trotta, p.37).

The last reform, still in force, was made in 1992 and the Commissioni Tributarie were replace by the Commissioni Provinciali, who judge tax disputes at the first instance. There are two degrees of appeal: decisions of a Commissione Provinciale can be appealed at the Commissione Regionale, who judge finally on appeal. Appeals to the Supreme Court (Corte di Cassazione) are only for

legitimacy issues. The reform also prescribed that the components of the tax commissions ceased from office upon the completion of the 75th year of age, and could not be assigned to the same section for more than 5 consecutive years, although this rule does not seem to be applied, as it appears in the lists of the members of these commissions publicly available on line.

4 Empirical analysis on the persistence and geographic dispersion of tax compliance in Italy since unification

4.1 Data

As it is well known, measuring tax compliance is not an easy task; this is mainly true when a historical perspective is pursued. We shall consider a measure of unpaid taxes for all Italian provinces over the period 1868-1870, after Italy unification, and an equivalent measure of unpaid taxes at provincial level in 2006.[11] The province and region levels turns out to be the relevant ones in case of dispute between the taxpayer and the tax administration and, in turn, to verify the implications of the model.

The historical measure of tax evasion we use is the official difference—as reported by Ministry of Finance of the Kingdom of Italy—between the total tax bill notified, during 1868-70, by the various municipality majors to all taxpayers in a province and the total tax revenue, that is the amount actually paid as taxes in the same period. In particular, we use data from the *Annuario Statistico delle Province Italiane* where information is reported about the amount of unpaid taxes per one hundred Italian Lira of assessed direct taxes in the period 1868-70. Assessed taxes include all the direct taxes as determined by the relevant public official and the Province Tax Commission.[12] They represent the official aggregate measure of the amount of direct tax bill due by all taxpayers in any province. The unpaid taxes for the years 1868-1870 represent the measure of all the assessed taxes that had not been collected.

Given the different tax assessment system of today in comparison with the period 1868-1870, it is quite difficult to identify an equivalent measure of unpaid taxes. Only a subset of individual declarations are assessed, and so it is not possible to get an aggregate measure of the amount of the tax bill due by all liable taxpayers, as we have for the unification period. We use data provided by the Italian Tax Revenue Agency on total tax gap and spontaneously paid taxes, that is the taxes that are actually paid. The sum of the two figures is the potential tax revenue amount. The tax gap is calculated using the top-down approach, based on the discrepancy between national accounts data of the Italian National Institute of Statistics (ISTAT) and fiscal data. ISTAT provides an estimate of the unobserved economy (mainly composed of underground economy - understating of income/profits or overstating of costs and

[11] Note that no data are available at the municipality level.

[12] Including Land Taxes, Property taxes and the form of labor and capital income tax used at that time (imposta di ricchezza mobile)

irregular work - and illegal economy – drug production and trafficking, prostitution and tobacco smuggling) and on this basis it calculates the theoretical tax base for different taxes (IVA, IRAP, IRES, IRPEF). By applying an estimation of what should be the appropriate average tax rate, a measure of the theoretical tax liability for each category of taxes is derived. The tax gap is defined as the difference between the theoretical tax liability and taxes paid as resulting from fiscal data. This measure of the tax gap is a crude estimate of intentional tax evasion, as for example it includes non-intentional mistakes by taxpayers (non evasion) and it excludes overstating of deductible expenses (e.g. charitable contributions) and hidden income in tax heavens or in foreign banks. In our analysis we use the tax gap for all taxes administered by the Italian Revenue Agency, as a proxy for the total missing tax revenues. The ratio between the tax gap and potential tax revenue amount in 2006 provides the measure of non compliance which is comparable to the measure of unpaid taxes in the period 1868-1870. We also use data on the efficiency of audits in 2006, as collected from the Italian Revenue Agency. In particular we consider data on the *weighted worked hours*[13] on audits and on the *equivalent output*[14] of audits. We take the ratio of equivalent output to weighteed worked hours to get a measure of audits efficiency.

At the beginning of the current century Italy was divided into one hundred and six provinces while in the aftermath of the unification the corresponding number was sixty-eight, as thirty-eight new provinces had been created in the meantime. With the aim to compare tax evasion in the two periods, we have defined our unit of observation on the basis of the old provinces. In particular, for twenty-eight cases the new province consists of land area that was entirely part of a single province at the time of the unification. Thus, in these cases we have simply aggregated the current provinces so as to restore the historical territories. For eight current provinces—Trento, Bolzano, Trieste, Roma, Viterbo, Latina, Frosinone, Rieti—we do not have historical data since these provinces were not part of the new-born Kingdom of Italy; hence we have dropped them. Finally, there are five cases such that the land area of the current province was part of two different historical provinces. This is so for Barletta-Andria-Trani, Varese, Pescara, Nuoro, and Enna. We have solved the indeterminacy either by aggregating the couple of historical provinces or by assigning the new province to one of the two historical ones according to the size of the land area parcelled out. In the former case we end up with sixty-three observations, while in the latter case with sixty-eight. We anticipate, however, that qualitative results are not at all sensitive to the chosen strategy.

[13]These are the number of hours worked on 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.

[14]Equivalent output is the actual product of a given process, in this case of audits, multiplied by the average time requested to produce one unit of that item. See Alborino et al. (2008) for further details.

4.2 Results

4.2.1 The persistence of tax evasion in Italy since unification

At the end of the XIX century the whole tax system was honeycombed with frauds, and abuses of various kinds set in (Selingam 1904). Alessio (1883) also reported high amount of tax evasion in 1877, as we considered in the previous section. Few years after the formation of the Kingdom of Italy the average rate of evasion across Italian provinces was about 25%. *(citation?)* Around this average value the tax gap was characterized by huge variations, ranging from 1% to 79% (see Table 1).

Table 1 here

Provinces that had been dukedoms in the past were characterized by the highest rate of evasion, higher than 40% on average. The two former big kingdoms—that is, the Kingodm of Sardinia (mainly consisting of the current regions of Piedmont, Liguria, and Sardinia) and the Kingdom of Two Sicilies (roughly corresponding to the south of Italy and the Sicily region)—both shared values of the size of evasion higher than the country average that is 34% and 30%, respectively. Provinces with with at least one large urban centre, that is with population higher than 100,000 inhabitants in 1870, had much higher rates of evasion than the rest of the country—37% instead of 23%. Among them, Venezia had the lowest measure of tax gap; Torino, Genova, and Naples all shared values of the tax gap much higher than the country average; the highest value was attached to Palermo.

Figure 2 here

Strikingly enough, the measure of the tax gap provided by the Italian Revenue Agency after one century and half from the unification of Italy is virtually the same as the historical one: about one euro out of four euros of taxes is still unpaid on average (see second row of Table 1). The main difference between the two distributions of tax gap attains at their dispersion as measured by the standard deviation, which halves from 0.18 to 0.09. This drop is perhaps much more evident by looking at Figure 2 where we report the change in the tax gap against the historical level. Clearly, provinces characterized by relative high (low) rates of evasion in 1870 have recorded increments in the rate of evasion lower (higher) than the country average. Moreover, it is also evident from the figure that points of the scatter plot align very well along the regression line, suggesting that there is no specific group of provinces which drives the overall result. Yet, the current value of the standard deviation still implies substantial spatial differences at province level in the size of unpaid taxes—actually, from 12% to 46%; this is consistent with the slope of the regression line which is higher than -1. In summary, by comparing the historical and actual distributions of tax gap over one hundred and fifty years it follows quite stable average values and rankings across provinces.

Table 2 reports more sistematic evidence on persistence in the ranking of the tax gap rate. The dependent variable of the regression is the logarithm

of the ratio ratio between the tax gap and potential tax revenue amount in 2006. For the explanatory variables, the Tax Gap in 1870 is the logarithm of the ratio between the amount of unpaid taxes and the overall amount of taxes in the period 1868-1870. High Population 1870 is a dummy variable taking the value of 1 for provinces characterized by at least one large urban centre (with population greater than 100,000 inhabitants). Sardinia Kingdom is a dummy variable for provinces that were part of the kingdom of Sardinia before the creation of the kingdom of Italy. A similar definition applies for the dummy variables Dukedom (dukedoms of Toscana, Parma and Modena) and Two Sicilies Kingdom (south of Italy and Sicily). The reference category is the Kingdom of Lombardo-Veneto. IRS efficiency is a measure of the efficiency of audits, as described above, for the year 2006. *(put a map of the dingdoms in the appendix?)*

In the first column of table 2 we show the point estimate of the relatioship between past and current evasion: as expected it is positive, less than 1, and statistically different from zero. Results in the second column confirm that persistence applies at province level, that is it is not driven by persistence in the average evasion of groups of provinces whose characteristics tend to persist over time, such as provinces with large cities, or provinces that were part of different states before the creation of the kingdom of Italy. The point estimate of the coefficient attached to the historical measure of tax evasion increases a bit and remains statistically significant even when we allow for a dummy variable identifying provinces with large urban centres and three dummy variables picking the former Kingdom of Sardinia and the Kingdom of Two Sicilies as well as the dukedoms located in the centre of Italy. Results in column 3 also take into account differences in current efficiency of tax assessment by the Internal Revenue Service. The coefficient of IRS Efficiency is negative, suggesting that the productivity of monitoring depresses the incentive to evade; however, it is not responsible for the persistence of tax evasion. Finally, we note that provinces in our data set have different sizes. To account for this heterogeneity, regression reported in the last column is weighted by land area of the province. Results suggest that on average higher value of the tax gap rate in 1870 by 10 percent matches with higher tax gap after one century and half by about 2 percent.

Table 2 here

4.2.2 Using the shock in the tax burden at unification to instrument tax evasion in 1870.

Previous evidence suggests persistence of tax evasion at local level even when we control for north-south or small-large cities differences. In his treatment of the Italian fiscal system Alessio (1883) suggests that the main cause for the differential diffusion of tax evasion in Italy was to be identified in the wide variability of the tax increments, around a strong average increment of the tax burden, that characterized Italy after the unification.[15] This view was shared

[15] As said before in 1869 the tax rate was 13.2%, compared to a rate of merely 2.46% in the UK in the same year.

by Selingam who noticed that "The tax rates are so enormously high that evasion and fraud are almost universal[...] with a tax rate four to five times as high as in England or Germany, the total yield is less than half of what it is in Germany and less than a third of what it is in England."We now exploit this conjecture in a formal way by constructing an index of the shock in the tax burden due to the Italian unification. In particular, we use information relative to the land and capital-labor taxes to derive a proxy for the differential increments in the tax burden across areas of Italy after the unification.

As reported before, when the new income tax was introduced by the Kingdom of Italy in 1866, an equivalent tax only preexisted in the Sardinia Kingdom, the Lombrado-Veneto, the dukedoms of Parma, of Modena and of Tuscany. Thus, for these areas as a proxy for the increment in the tax burden due to the new tax, we have calculated the ratio between the per capita tax revenue in 1872 and the corresponding measure in 1861, that is before the introduction of the new income tax—data from Zamagni (2001). The ratio has then been rescaled so that to have an index ranging from 1 to 5. A value of 5 is also assigned to the former Papal State and the Two Sicilies Kindom where an equivalent income tax did not exist. The same procedure has been used to have an index on the increments in the tax burden due to the land income tax after the reform in 1867. In particular, in this case the before-after ratio in the tax revenue is relative to the per capita revenue forecasted by the government for the 1867 and the corresponding actual value in 1860—data from Zamagni (2001). Since this tax already exhisted in all areas of Italy before the unification, variability of the index is more smooth than in the previous case. Finally, to have an overall index of increments in the tax burden just after the unification of Italy we have simply averaged the two elementary indices (see Table 3).

Table 3 here

Table 4 reports evidence when the index of tax rate variation is used as an instrument for tax evasion in 1870. The first column shows results of the first stage: there is a strong positive relationship between the tax gap and our index of change in the tax burden. In particular, the t-ratio, whose value is around 5, is well above the lower bound for not incurring in the weak-instrument concern. The last two columns—where we report the 2SLS results—confirm previous evidence relative to the persistence of tax evasion over one century and half. In fact, the 2SLS estimate is higher than the OLS one, more so when we weigh observations by using the land area of the provinces.

Table 4 here

5 Conclusion (to be completed, in particular extend the policy implications on political integration processes)

We provide a model for explaining the persistence and geographical dispersion of tax compliance. In the presence of high tax rates and of a congestion externality on the enforcement process, the model predicts multiple equilibria and history dependence. Historical evidence for the case of Italy supports the insights of the model and empirical analysis suggests that the role of the local enforcement externality has been important for the emergence of history dependence. Results suggest that persistent heterogeneity of tax compliance may have to do more with apparently small details of the design of tax collection and jurisdiction than with culture, habits, social norms and social capital. There is support for the idea that the convergence process relies on the importance of formal sanctions (Posner, 1997) for the establishment of a good social norm. Some policy implications on political integration processes can be drawn, in that political integration should take into consideration historical conditions and the right degree of decentralization of the process of law enforcement.

6 References

Acemoglu, D. and Jackson, M., (2017), Social norms and the enforcement of lows, *Journal of the European Economic Association*, Vol.15 (2), pp. 245–295.

Acemoglu, D., Johnson, S. and Robinson, J. (2001), The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, Vol. 91 (5), pp. 1369-1401.

Acemoglu, D., Johnson, S. and Robinson, J. (2005), Institutions as a fundamental cause of long-run growth, Chapter 6 in *Handbook of Economic Growth*, vol. 1, Part A, pp 385-472, Elsevier.

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.

Antonielli, E. (1872) Annuario Statistico delle Provicie Italiane per l'anno 1872, Tofani, Firenze.

Becker, G.S., (1968), Crime and Punishment: An Economic Approach, *Journal of Political Economy,* Vol. 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.

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

Dominici, R. and G. Marongiu, (2005), "La giustizia tributaria dalla fondazione dello Stato unitario alla Seconda guerra mondiale" in "La Giustizia Tributaria Italiana e la sua Commissione Centrale", Paleologo, G., Giuffrè editore, Milano.

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

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

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

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

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

Guiso L., Sapienza P., Zingales L., (2016), Long-term Persistence, *Journal of the European Economic Association*, Vol. 14(6), pp. 1401-1436.

Lochner, Lance (2007) Individual Perceptions of the Criminal Justice System, *American Economic Review*, Vol. 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.

Nicola Francesco, (1897), Il Sistema del Contingente nella Imposta di Ricchezza Mobile, *il Giornale degli Economisti*, Vol. 15 (8), pp. 531-551.

Nuzzo Margherita (2011), Il sistema di riscossione coattiva delle imposte dirette tra funzione pubblica perseguita e tutela del contribuente. Unpublished manuscript. Universita' degli Studi di Milano-Bicocca.

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*.

Posner, Richard (1997), Social Norms and the Law: An Economic Approach. *The American Economic Review*, Vol. 87 (2), pp. 365-369.

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

Simone, Antonio (2014), Il giudice tributario: passato, presente e futuro. Unpublished manuscript, Ordine dei Dottori Commercialisti e degli Esperti Contabili, Milano. Scuola di Alta Formazione Luigi Martino.

Traxler, Christian (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*, Vol. 1, pp. 423-436