DO VOTERS LEARN FROM DYNAMIC TAX COMPETITION?

Ilaria Petrarca[1]
06 May 2011

PRELIMINARY DRAFT
Please do not quote

1. Introduction

The political economics literature stresses the advantages from yardstick competition as a mechanism to constrain the incumbents' rent during the electoral year (Besley and Case, 1995; Bordignon et al., 2003). This paper analyzes strategic interaction among local governments from a different perspective and investigates how persistent is the reduction of selection powers that yardstick competition it is supposed to solve.

When yardstick competition is at work, in fact, voters learn the true type of the incumbent running for re-election only if a separating equilibrium in tax rates is observed. Although yardstick competition reduces the asymmetric information problem of the voters, the empirical literature on yardstick competition suggests a pattern of mimicking both in US (Besley and Case, 1995) and in many European countries (Solè Ollè in Spain, Bordignon et al. in Italy, Revelli in England, and so on). When mimicking occurs a pooling equilibrium is observed and voters receive a signal of high competence regardless the true competence level of the incumbent. This illusion is generated by the strategic behavior of the less competent incumbent who mimics the fiscal decision of the more competent incumbent in the neighborhood when re-election concerns are binding, by setting a tax rate lower than his rent maximizing level. Before elections voters update their electoral preferences with a misleading information, the incumbent's competence is overestimated and his probability of being re-elected is distorted upwards. As a consequence, tax mimicking arising from yardstick competition advantages the bad incumbent to the detriment of voters' selection powers. The re-elected less competent incumbent, moreover, will set higher tax rates than a good incumbent during the next term of office to finance his rent seeking activity and the rent reduction experienced during the electoral year is offset by the decrease of voters' future utility.

Since the existing models of yardstick competition are static, the present work investigates if yardstick competition, once repeated in time, removes the possibility that a pooling equilibrium arises. The key assumption of the model is that the informational capital does not perish every time the game is repeated, but it accumulates up to the point that voters are able to infer if the incumbent is mimicking or not.

The learning process of the voters is generated by the information on the tax rate set in both the domestic and in the neighboring jurisdictions as in the static yardstick competition model, but the longitudinal dimension of the information available in the

---

[1] PhD Candidate at IMT Alti Studi Lucca. E-mail: i.petrarca@imtlucca.it

dynamic setting makes it possible that voters learn the strategy of the past incumbent and infer the probability that the current incumbent is mimicking. The amount of learning is determined by three factors: an exogenous possibility to learn, the probability that voters gather enough information to learn and the weight attached to past experience. The dynamic model proves that when the learning process is maximized voters' beliefs are correctly updated, selection powers are restored and mimicking incumbents would not be re-elected.

The consequences of the dynamic learning on selection powers are thus crucial for the electoral accountability mechanism to work properly. The improvement with respect to the static learning by tax rates is huge, and to make a strong point the graphical solution of the model by Besley and Case (1995) has been replicated in Graph 1. The cost shock level is measured on the horizontal axis while the tax rate level is measured on the vertical axis.

The authors predict that when the cost shock assumes values too low or too high, a separating equilibrium arises because the bad incumbent can either signal good competence while maximizing his ego rent (low cost), or he finds it too costly to seek for votes and he sets the highest tax rate no matter the electoral consequences (high cost). The tax function in this situation is a positive sloping line depending on the cost shock level and the amount of rent diverted R. When the cost shock takes intermediate values, the bad incumbent faces a trade off between vote seeking and rent seeking. The horizontal dotted segment of the tax function represents the mimicking tax level set to signal good competence to voters.

When an incremental learning process occurs, on the contrary, successful mimicking becomes much more difficult to implement because voters learn about the cost function and they infer the incumbents' strategy. The bad incumbent running for re-election would not find it optimal to behave strategically because he would renounce to a share of rent without increasing the probability of re-election. As a consequence a separating equilibrium will be observed also for intermediate values of the cost shock. In the Graph below, this result is represented by the bold continuous segment of the tax function. The same segment indicates the interval of values for which selection powers are enhanced and yardstick competition is effective in improving accountability at the local level.

Graph 1. Dynamic learning and bad incumbent's tax rate decision

2
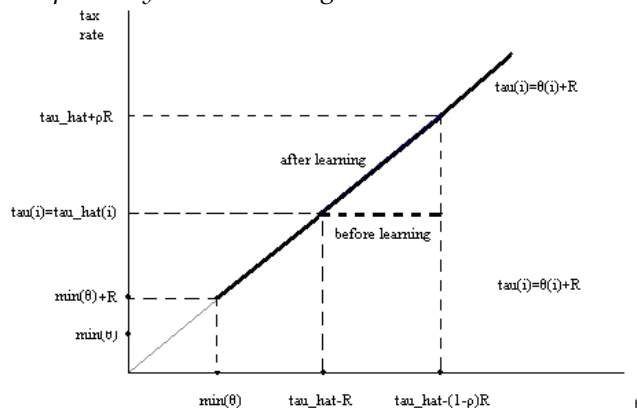
The topic of this work is the main contribution to the literature because the effect of yardstick competition on voter's selection powers is a topic that has not been addressed yet by the scholars. Previous works focused on testing tax mimicking (for a survey see Allers and Elhorts, 2004) or attempted at shaping the incumbents' incentives to mimic (Bordignon et al., 2003; Solè Ollè, 2008; Shaltegger and Kuttel, 2002). Beside this aspect this paper introduces several innovations with respect to the baseline theoretical framework. First, the model of yardstick competition is here developed by the voter's point of view and the utility maximization problem of the citizens is solved with respect to the electoral decision (re-elect the incumbent or not). This choice is motivated by the research question, which focuses on the behavior of the voters. Second, the model is applied to a dynamic setting, expanding the two-period time horizon of the existing models of yardstick competition. The dynamic setting assumes that information does not perish every time period but it incrementally accumulates in time, modeling the evolution of voters' behavior depending on the extent to which the learning process occurs. The existing literature, on the contrary, has so far assumed the mere replication of the game every two periods. Finally, this paper offers an empirical test of the predictions of the dynamic model. Dynamic Bayesian updating of the beliefs and probit estimation have been applied on a dataset of Italian Municipalities in which static yardstick competition has been previously verified.

The rest of the paper is organized as follows. Section Two reviews the contributions in the literature that refer to yardstick competition and learning. Section Three describes the timing, the object and the exogenous conditions for learning to occur. The dynamic model of yardstick competition is presented in Section Four, providing formal results of the effect of the dynamic learning process on selection powers. Section Five describes the dataset on Italian Municipalities and shows the results of the empirical analyses. Finally, Section Six concludes.

2. Related literature

Learning from tax rates has been mainly studied by the local public finance literature. The baseline model of yardstick competition developed by Besley and Case (1995) shares the common view in economics that decentralized jurisdiction are 'local laboratories' in which policies are experimented and the observed outcomes determine the citizens' judgment of the policy makers (Salmon, 1987). Yardstick competition is a mechanism of informational spillover exploited by voters to overcome the agency problem between citizens and politicians regarding the cost of public provision. Since the cost is correlated among neighbors, the relative performance of the incumbent in the region reveals information about the size of his rent seeking activity. The baseline model of yardstick competition predicts a pooling equilibrium in tax rates when a bad incumbent observes lower tax rates in nearby jurisdictions. In such a situation he mimics the neighbors by setting their same tax rate, renouncing to a share of his ego rent to seek for re-election.

The learning by tax game is comparable with the learning by prices models developed by the industrial organization scholars. Benabou and Gertner (1992) studied strategic

competition in a market with two sellers of different types selling a homogeneous good to a customer, assuming the price as a performance indicator revealing the true type of the seller. The authors obtained the same theoretical results as Besley and Case (1993) as the bad seller mimics the good seller by reducing the markup.

The model´s assumption, however, are quite stringent. The prerequisite for static learning from yardstick competition to work is that voters gather and exploit information on the fiscal performance only during the current electoral year. Bordignon et al. (2003) derive the conditions for a successful mimicking of the bad incumbent in the static setting, taking as given that voters gather information about domestic and neighbor´s tax rates. This assumption is not trivial and should not be underestimated since voters' incentives to be informed are small. The change of regime, in fact, is a pure public good and the probability of being pivotal is reasonably close to zero, generating free riding concerns that discourage voters from acquiring information (Schnellenbach, 2005). In the dynamic setting this assumption becomes crucial because if the flow of information is stopped, learning does not occur. Bar-Isaac (2003) investigated the dynamic learning from prices and he proved that when learning occurs in time only the good seller survives in the market. By similarity, in time in the yardstick competition setting only good incumbents find it optimal to choose to run for re-election. The analytical policy literature reached a similar result by studying the diffusion of policy decisions. If several policy makers face a decision and they are exposed to the same stock of information, their beliefs on the performance of the policy converge and they will select the best performing policy among the feasible set of alternatives. The contribution of this strand of literature to this paper, beside the focus on policy decisions that are easily comparable with the voting decision, is the introduction of empirical methodologies to test for the presence of a learning process. The most recent contribution comes from the work of Meseguer (2009), who developed a model in which a government must decide between two alternative policies. In this model the government learns in light of experience and then makes rational choices. Beliefs are updated with the information about own and neighboring past experiences according to the Bayes' rule. Since every agent in the model is exposed to the same information, the performance of each policy decision is common knowledge and the learning process is spurred. The author tests the model to a sample of south-American countries during the 90s, finding that the implementation of institutional and economic reforms has been driven by a learning process consistent with the theory.

The main critique to Bayesian estimation claims that learning is conditional on prior beliefs and new information is taken into account differently from different policy makers, leading to different paths of learning among the agents. As an example, Gilardi (2010) suggest that agents learn from agents sharing their same ideological preferences. Recent papers found support for the political trends hypotheses in yardstick competition, showing that incumbents mimic neighboring jurisdictions ruled by the same political party (Santolini, 2009; Delgado et al., 2011). Although this critique is appealing, it does not apply to voters' behavior. First, at the sub-national level ideology may not play an important role as in the national elections because preferences may be

strictly determined by the local policy platform (e.g. building a bridge, opening new nursery schools, improving the public transportation and so on). Moreover, the smaller the tier of government analyzed the more relevant is the phenomenon of the local lists, which are not formally linked to any political party and are characterized by a platform focusing on territorial issues and not on ideology. Finally, although voters have a preference towards higher public expenditure, if they find it out that the incumbent is mimicking and the share of additional expenditure is wasted in rents, the incumbents' reputation worsens and voters do not trust him anymore. Hence, he would hardly be re-elected.

3. The incremental learning process: a step-by-step description

This Section introduces dynamic learning from tax rates by expanding the model of yardstick competition developed by Bordignon et al. (2003). First, the environment is described then the timing and the object of the learning process is described. Finally, the results are proved to be robust also when the assumptions of a term limited incumbent and constant cost shocks are relaxed. This preliminary argumentation explains what is meant for learning – the object, the timing and the necessary conditions. In particular, the requirements for learning discussed in the final part of this Section will be used in Section 4.2 to define the feasibility of a learning process to occur.

3.1 The environment

Consider a world made of two jurisdictions. Jurisdiction i is assumed to be a neighbor of $-i$ and vice versa. When the number of jurisdictions increases, the set of i's neighbors does not necessarily include all of them. The game lasts for two periods, (t=1, 2). Each period is an electoral term consisting of N years. The last year of each period an election is held between the incumbent and a challenger.

The utility of the voters in each jurisdiction during the period t depends on the consumption of both private (C) and public goods (g):

$$u_{it}^v = C_{it} + g_{it}$$

[1]

where private consumption is the amount of income (y) net of taxes (T):

$$C_{it} = y_{it} - T_{it}$$

[2]

The tax rate proxies the cost of the public provision of goods and services, T:

$$T_{it} = p_t + \theta_{it} - \varepsilon_i$$

[3]

where i refers to the jurisdiction and t refers to time. T_{it} is determined by the observed national price of the public provision (p_t), and by two factors that are observed by the incumbent but not by the voters: a random cost shock (θ_{it}) and the competence level of the incumbent (ε_i). The competence of the incumbent is an individual specific characteristic, constant in time, representing a measure of efficiency in providing public goods. The mayor of each jurisdiction may be competent (good type) or not (bad type) where competence is inversely related with the undertaken rent-seeking activity:

5

$$\varepsilon_i = \begin{cases} \varepsilon_H & if \quad 'good' \\ \varepsilon_L & if \quad 'bad' \end{cases}$$

[4]

such that $\varepsilon_H > \varepsilon_L > 0$ and Prob $(\varepsilon_i = \varepsilon_H) = \varphi$.

The utility level of the voters can be summarized as:

$$u_{it}^V = g_{it} + y_{it} - p_t - \theta_{it} + \varepsilon_i$$

[5]

From equation 5 the link between the electoral decision of the voters and the voters' utility is clear.

Voters are rational agents who choose between re-electing or not the incumbent in period 1 with the purpose to maximize their expected utility (equation 5). The static setting assumes that information is costly and voters gather information about the performance of the incumbent only the year nearest to elections. Furthermore, information fully depreciates every period and voters must begin the process from scratch. The performance indicator considered is the domestic local tax rate, which is benchmarked with the neighbors' tax rate. However, sometimes it is impossible to distinguish a good incumbent from a bad incumbent during the electoral period. To explain why, let us describe the process of tax setting.

When choosing the level of the domestic tax rate, the good incumbent only takes into account the cost shock realization in the first period. When a negative shock occurs ($\theta_{it} > 0$), an additional amount of resources ($\Delta > 0$) is needed to finance the public provision. The good incumbent thus sets $T_{it} = T + \Delta$ when the shock is negative and $T_{it} = T$ otherwise.

The bad incumbent, on the contrary, sets the tax rate to finance both the public provision of goods and services and his private rent seeking activity. As a consequence, he will always – *ceteris paribus* - set a higher tax rate than the good incumbent. Let us define the bad incumbents' tax rate as $T_{it} = T + k\Delta$, where k is the share of additional resources diverted to rents. When the shock is negative $1 \leq k \leq R$, assuming some finite upper bound to the rent extraction, R^2 which is determined by technology constraints or the fact that the size of the rent is so high that the incumbent is unmasked.

The timing of the game is set as follows:

1. At the beginning of period 1 Nature selects a competence level of the incumbent and a cost shock;

2. The incumbent in i observes his competence level, his cost shock realization, then he sets T_{1i};

3. Voters in i observe (T_{1i}) and (T_{1-i}) and update their beliefs on the relative competence level of the incumbent in the neighborhood;

4. At the end of period 1 an election is held between the incumbent and a challenger with a majoritarian electoral rule;

5. At the beginning of period 2 Nature selects a cost shock and the game restarts; if the challenger has been elected his competence level is randomly selected by Nature.

---

[2] Assuming a Laffer curve for the rent extraction of the type $L = k\Delta + k\Delta^2$, the value R that maximizes L is $R = 1/2\Delta$. For k>R as the share of the revenue diverted to rents increases the effective rent received by the incumbent decreases.

Now suppose that in period 1 a bad incumbent runs for re-election and observes a domestic positive cost shock. In the same period in the neighboring jurisdiction a good mayor is experiencing a negative shock. The good incumbent sets $T+\Delta$, while the bad incumbent can choose between setting $T+k\Delta$, his rent maximizing tax rate, or the pooling tax rate $T+\Delta$. In the last case the bad incumbent renounces to the rent, voters observe a signal of high competence, mistake him for a good mayor and he gains re-election. The yardstick competition literature claims that, as long as some conditions are met[3], a pooling equilibrium arises during the electoral year. The tax rates set during the electoral year are the same in the two jurisdictions: $T_{i1} = T_{-i\,1} = T+\Delta$. In such a situation voters observe a signal of good competence, also when the incumbent is in fact bad.

Table 1 summarizes the tax decisions in period 1 conditional on the type of incumbent and the cost shock realization. The cell marked with an asterisk indicates the yardstick competition outcome.

Table 1. The incumbent's tax rate strategies in period 1

	Bad i; Good -i	Bad i; Bad i	Good i; Good -i
Ni; N-i	$T+k\Delta$, $T+\Delta$	$T+k\Delta$, $T+k\Delta$	$T+\Delta$, $T+\Delta$
Ni; P-i	$T+k\Delta$, T	$T+k\Delta$, $T+\Delta$	$T+\Delta$, T
Pi; P-i	$T+\Delta$, T	$T+\Delta$, $T+\Delta$	T, T
Pi; N-i	$T+\Delta$, $T+\Delta$ *	$T+\Delta$, $T+k\Delta$	T, $T+\Delta$

N=negative cost shock, P=positive cost shock; i refers to the domestic jurisdiction, -i to the neighbor(s). The asterisk indicates the yardstick competition outcome.

During the second period of the model no election is held and the assumption of a limit of at most two terms in office removes any reputational concerns of the local incumbent. This is a classical assumption in the literature and it is kept here for simplicity.

In the second period the bad incumbent sets his rent-maximizing tax rate because he cannot run for re-election anymore and he is not incentivized to care about his electoral popularity. Consequently, in period 2 the bad incumbent does not find it optimal to renounce to a share of his ego rent and he would always set $T_{i2} = T+k\Delta$. The tax rate decisions in period 2 are shown in Table 2.

Table 2. The incumbent's tax rate strategies in the period 2

	Bad i; Good j	Bad i; Bad j	Good i; Good j
Ni; Nj	$T+k\Delta$, $T+\Delta$	$T+k\Delta$, $T+k\Delta$	$T+\Delta$, $T+\Delta$
Ni; Pj	$T+k\Delta$, T	$T+k\Delta$, $T+k\Delta$	$T+\Delta$, T
Pi; Pj	$T+k\Delta$, T	$T+k\Delta$, $T+k\Delta$	T, T
Pi; Nj	$T+k\Delta$, $T+\Delta$	$T+k\Delta$, $T+k\Delta$	T, $T+\Delta$

N=negative cost shock, P=positive cost shock; i refers to the domestic jurisdiction, -i to the neighbor(s).

---

[3] See Bordignon et al. (2003) for a detailed description.

3.2 The object and the timing of learning

This Section extends the timing of the existing two-period yardstick competition models, by describing the process through which voters learn the electoral strategy of the incumbent.

Let us consider the third electoral year of the game. A pooling equilibrium arises again, and assuming that the conditions for a successful mimicking hold, the bad incumbent in jurisdiction i sets $T_{i3} = T_{i1} = T + \Delta$. However, the information on the tax rates set in both i and $-i$ during both the first and the second period are now available to the voters. This information triggers the incremental learning process.

As a first step, by comparing the tax rates set in the second period with the tax rate set in the first period voters learn about their incumbent's true type and the neighbor's true type.

Tax rates in the second period are not strategic[4], therefore the bad incumbent will set $T_{i2} = T + k\Delta$ regardless of the cost shock realization while the good incumbent will set $T_{i2} = T + \Delta$ if the shock is negative and $T_{i2} = T$ if the shock is positive.

The additional amount of resources needed to face a negative cost shock is assumed to be the same in all the periods and for both the incumbents: $\Delta_t = \Delta$. This assumption may seem unreasonable because the technological shock may change its magnitude every period and in every jurisdiction. However, even fixing the magnitude of the shock to a constant parameter, the sign of the shock still plays the leading role in determining the strategy of the incumbent. To stress this point at this stage of the discussion $|\Delta_t|$ is assumed to be a constant.

If voters in i observe $T_{i2} > T_{i1}$ during the second period, they know for sure that the incumbent is bad and that he mimicked in the first period. Otherwise, if they observe a tax rate not higher than the tax rate set in the first period ($T_{i2} \leq T_{i1}$) they infer that the incumbent's true type is good.

By observing the tax rates set in the second period voters gather information also about the cost shock realization in the jurisdiction where the incumbent is good. A good incumbent who sets $T_1 = T + \Delta$ in the first period would set the same tax rate $T_1 = T_2 = T + \Delta$ in the second period when facing again a negative cost shock. On the contrary, he would set $T_2 < T_1 = T$ when facing a positive cost shock. To sum up:

$$\theta_2 \mid \varepsilon_H = \begin{cases} N & if \ T_2 = T_1 = T + \Delta \\ P & if \quad T_2 < T_1 = T \end{cases}$$

[6]

The incumbent and his neighbor's true type are valuable information about the first period's tax setting equation, although they are not directly useful in updating beliefs in the third period because the incumbents have changed.

The cost shock is assumed to be spatially correlated among neighbors, reflecting the socio-economic interdependence of the jurisdictions. As an example, the cost of streets maintenance work depends on weather conditions which are similar among neighbors, but they are unknown to laymen because the extent of the damage is difficult to gauge

---

[4] This assumption is relaxed in Section 3.2.1.

without expertise. Moreover, while the local government controls the whole territory of the jurisdiction, voters have reasonably not information on every street condition.

The degree of correlation among jurisdictions is the same in both periods ($\sigma_{i2}=\sigma_{-i2}$) because the technological interdependence between neighboring economies is based on the geographical nearness, common natural resources, possible joint public provision and other factors which are unlikely to unexpectedly change in the short period.

The cost shock is specified as:

$$\theta_{it} = \Delta_{it} + \sigma\Delta_{-it}$$ [7]

where Δ_{it} is the increase in the tax rate due to a cost shock specific to i ($\Delta_{it} = 0$ if a positive shock occurs), $\sigma\Delta_{-it}$ is the portion of cost shock specific to $-i$ which is reflected in i according to the correlation parameter $0<\sigma<1$, and Δ_{it} and Δ_{-it} are assumed to be the same in both the jurisdiction.

Given this setting, during the first period voters ignore both θ_{i1} and ε_i. In the second period the true type of the incumbent ε_i is revealed and the information on θ_{i1} and θ_{-i1} is easily inferred according to:

$$\theta_{i1} = p_1 - \varepsilon_i - T_{i1}$$ [8]

and

$$\theta_{-i1} = p_1 - \varepsilon_{-i} - T_{-i1}$$ [9]

The information on θ_{i1} is not directly useful in updating beliefs in the third period, but it may be used to infer the value of Δ and the correlation coefficient σ.

The correlation of the cost shock is:

$$\sigma = \begin{cases} 1 & if\ \varepsilon_i = \varepsilon_H, \\ 0 & if\ \varepsilon_i = \varepsilon_L, \end{cases}$$ [10]

Equation 10 exploits the fact that during the first period the good incumbent in the neighborhood experiences a negative cost shock ($\Delta_1 = \Delta$) and pooling equilibrium is feasible. If the incumbent in the domestic jurisdiction is the good type, the cost shock increases the unit cost of public provision of goods and services according to reflection of the cost shock from the neighbors ($\sigma\Delta$), while if the incumbent is the bad type and faces a positive cost shock, no additional resources need to be spent.

When the parameter Δ is known, the information on σ is easily inferred and the learning process is complete. The exogenous conditions for the disclosure of the information about Δ are stated in Proposition 1.

Proposition 1. Dynamic learning by tax rates is feasible if the value of the parameter Δ is known or inferred by voters. Voters infer the value of Δ when both the following requirements are met:

- One of the incumbents is good;
- The good incumbent experienced a negative shock in the first period and a positive shock in the second period.

The presence of at least a good incumbent is a condition for having a non strategic benchmark behavior to refer to. The good incumbent's tax rate depends on the cost

shocks realization. Sequential shocks of opposite magnitude in the two periods are a necessary requirement to compute the differential *(T+Δ)-(T) = Δ*. Moreover, since a pooling equilibrium has been observed during the first period, the negative shock must occur before the positive shock.

As shown in Table 3, assuming a pooling equilibrium (implying a negative shock) has been observed in the first period, the required conditions to learn Δ hold only when the incumbent in the second period is good and experiences a positive cost shock. As the Table indicates, Proposition 1 holds in five cases over twelve.

Table 3. The tax rates in period 2 and Proposition 1.

	Bad-Good	*Bad-Bad*	*Good-Good*
NN	Does not hold	Does not hold	Does not hold
NP	**Holds**	Does not hold	**Holds**
PP	**Holds**	Does not hold	**Holds**
PN	Does not hold	Does not hold	**Holds**

N=negative cost shock, P=positive cost shock; the first letter (or word) refers to i, the second to -i

The tax setting equation in the third period is:

$$T_{i3} = p_3 + \Delta_{i3} + \sigma\Delta_{-i3} - \varepsilon_i \qquad [11]$$

When Proposition 1 holds, voters observe the tax rate T_{i3}, the national price of the public provision p_3, the additional resources needed to face a negative shock Δ_3 and the spatial correlation parameter σ.

Voters can now infer the competence level ε_i from Equation 11. In such a situation the bad incumbent would not find it optimal to mimic the good incumbent behavior because the strategic behavior would not increase his probability of being re-elected and a separating equilibrium would arise. Eventually, the bad incumbent would not run for re-election and renounce to the second term rent. On the contrary, if the bad incumbent is not aware of the voters' learning process he would mimic the good neighbor, but this time he will be unmasked and turned down. In both cases, the electoral competition would select only competent politicians and entail an improvement in the quality of the political class.

3.2.1 Bad incumbent with career concerns

This paragraph relaxes the assumption of a non strategic behavior of the incumbent during the electoral period when there is a binding term limit. The motivation for this excurse is the consideration that incentive to mimic the good neighbor may arise even if the incumbent cannot run for re-election. This situation for example occurs whenever the incumbent aspires to continue his public career and be elected or appointed to other offices, or because he is subject to party control and a signal of bad competence would disadvantage the party at the next election.

The findings suggest that conditions stated in Proposition 1 are robust to the introduction of career concerns of the bad incumbent.

The utility function of the incumbent is determined by the exogenous office rent from being in office (m_t), the extra rent diverted from public resources for rent seeking purposes ($0 \leq r_{it} \leq R$), and the discounted value of the future rent (ω_{t+1}):

$$u_{it}^{inc} = m_t + r_{it} + \beta \omega_{it+1} \qquad [12]$$

Career concerns are relevant when the bad incumbent prefers to give up to the rent to increase his possibility to be appointed to another office after the second period. The discounted value of the future rent ($\beta \omega_{it+1}$) is larger than the maximum rent he can expropriate in the second period:

$$(k+1)\Delta < \beta \omega_{it+1} \qquad [13]$$

When a pooling equilibrium is observed again during the third period, under the assumption of a constant cost shock the strategic tax rate $T+\Delta$ equals the tax rate set in period 1. If the pooling equilibrium is driven by mimicking, one of the incumbents is surely bad (the one who mimics) and another is good (the mimicked), and the good incumbent's decision is not affected by the bad incumbents' reputational concerns. Hence, the parameter Δ can be inferred when the tax differential $T+\Delta-T=\Delta$ can be computed, that is when the conditions in Proposition 1 hold. As a conclusion, the results of the previous paragraph are robust to the introduction of career concerns of the bad incumbent.

3.2.2 Non-constant cost shocks

So far Δ has been taken as a time constant parameter. This Section relaxes this assumption allowing Δ to change every period.

Let us consider the two cases in which $\Delta_2 = \Delta'$ where $\Delta' > \Delta$ or $\Delta_2 = \Delta''$ where $\Delta'' < \Delta$.

If $\Delta_2 = \Delta'$, during the second period voters cannot distinguish if the incumbent sets the highest tax rate $T+k\Delta'$ or the pooling tax rate $T+\Delta'$ because both the levels are higher than the pooling tax rate $T+\Delta$ observed in the first period. Voters, in fact, extract a signal revealing a bad incumbent in the jurisdiction where the tax rate is the highest and a good incumbent in the jurisdiction where the tax rate is the lowest. Suppose that the good neighbor experienced a negative cost shock also during the second period. In this situation voters in i observe the domestic tax rate $T+\Delta' > T+\Delta$ and erroneously guess that the incumbent is a bad type. Moreover, voters cannot learn about the public cost function because they infer the cost shock is $\theta_{i1} = \sigma \Delta_1$, they compute $(T+\Delta') - (T+\Delta) = \Delta' - \Delta$ but they cannot learn σ using equation 10.

On the contrary, when the tax rate decreases in the second period ($T_1 = T+\Delta$, $T_2 = T$) voters gather information about Δ, they learn the correlation coefficient σ and they infer the competence level of the past incumbent. However, they don't know the level of Δ' and they cannot learn the true type of the current incumbent. If voters assume $\Delta = \Delta'$, they end up underestimating the competence level of the incumbent.

As shown in Table 4, Proposition 1 holds in the same cases as for a constant Δ, signaled by a an asterisk.

Table 4. Tax rates in period 2 and the respect of Proposition 1 when Δ2= Δ'> Δ.

	Bad-Good	Bad-Bad	Good-Good
NN	T+kΔ', T+Δ'	T+kΔ', T+kΔ'	T+Δ', T+Δ'
NP	**T+kΔ', T ***	T+kΔ', T+kΔ'	**T+Δ', T ***
PP	**T+kΔ', T ***	T+kΔ', T+kΔ'	**T, T ***
PN	T+kΔ', T+Δ'	T+kΔ', T+kΔ'	**T, T+Δ' ***

N=negative cost shock, P=positive cost shock; the first letter (or word) refers to i, the second to -i

Now let us turn to the second case, that is Δ2= Δ'' where *Δ''< Δ*. Since *k>1* two situations may arise:

1. If *kΔ''≤Δ* and the rent maximizing tax rate of the bad incumbent *T+kΔ''≤T+Δ*, voters always observe a lower or the same tax rates as the first period and they cannot learn *σ*;

2. If *kΔ''>Δ* and the rent maximizing tax rate of the bad incumbent *T+kΔ''>T+Δ*, although voters cannot distinguish *T* from *T+Δ''* they understand when the incumbent is bad because he sets the highest tax rate. However, learning *σ* is feasible only when a pooling equilibrium was observed in the first period and the good neighbor incumbent experienced a positive shock in the second period, as in the three cases marked with an asterisk in Table 5.

As in the previous case, however, the learning process is limited to the true type of the past incumbent.

Table 5. Tax rates in period 2 and the respect of Proposition 3 when Δ2= Δ''< Δ.

	Bad-Good	Bad-Bad	Good-Good
NN	T+kΔ'', T+Δ''	T+kΔ'', T+kΔ''	T+Δ'', T+Δ''
NP	**T+kΔ'', T ***	T+kΔ'', T+kΔ''	**T+Δ'', T ***
PP	**T+kΔ'', T ***	T+kΔ'', T+kΔ''	T, T
PN	T+kΔ'', T+Δ''	T+kΔ'', T+kΔ''	T, T+Δ''

N=negative cost shock, P=positive cost shock; the first letter (or word) refers to i, the second to -i

4. Learning from tax rates: the voters' point of view

This Section models the citizens' voting behavior as a rational response to the observed signal of competence. First, the static model is described, then the assumption on the information depreciation is relaxed and the learning process is formalized. Finally, the solution to the utility maximization is derived for both the static and the dynamic games.

4.1 The static model

The voters' problem is:

$$V_{i1}(\varepsilon_i) = \max_j \left(u_{i1}^V(T_{i1}) + \beta(V_{i2}^{VI}); u_{i1}^V(T_{i1}) + \beta(V_{i2}^{VC}) \right) \qquad [14]$$

where *j* is the electoral rule implemented by voters. Since the voter's maximization is the same in each jurisdiction, in what follows the index *i* is suppressed.

The solution of Equation 14 depends on the level of competence of the incumbent. The right hand side of the equation lists the present value of the two-year electoral term

utility when the incumbent is re-elected ($u_1^V(T_1) + \beta(V_2^{VI})$) and when he is defeated at election and a challenger takes his place($u_1^V(T_1) + \beta(V_2^{VC})$). Since the pooling tax rate is observed, the utility during the first period is the same.

On the contrary, the utility level during the second period differs depending on the decision of the voters. Re-electing a good incumbent increases the voters' utility because the tax rate set by a good incumbent is by definition lower than the tax rate set by a bad incumbent ($T_{\varepsilon_H} < T_{\varepsilon_L}$). Utility from defeating the bad incumbent will be higher than utility from re-electing him:

$$V_2^{VI}(T_{\varepsilon_H}) > V_2^{VI}(T_{\varepsilon_L}) \qquad [15]$$

The value of the tax rate set by the good incumbent $T_{1,i} \mid \varepsilon_i = \varepsilon_H$ is approximated by the tax rate set by the neighbors, $T_{1,-i} \sim T_{1,-i} \mid \varepsilon_i = \varepsilon_H$, because it is the assumed benchmark for the voters. The information exploited refers to the tax rates set only in period 1. However, voters' expectations about the expected tax rate set during the second period ($E(T_2)$) are computed as:

$$E(T_2) = \mu_1(T_{\varepsilon_H}) + (1 - \mu_1)(T_{\varepsilon_L}) \qquad [16]$$

where T_{ε_H} and T_{ε_L} are the tax rates conditional on the probability that the incumbent is good or bad and μ_1 is the updated belief on the competence level before voting. Voters' beliefs are updated with the information on the tax rate set in both the domestic jurisdiction i and the neighbors $-i$ according to the Bayes' rule.

The belief μ_1 depends on the relative performance of $T_{1,i}$ with respect to $T_{1,-i}$ and on the ex-ante beliefs that the incumbent is good type φ. The parameter φ behaves as a random variable, $\varphi \sim N(1/2,1)$ and it is bounded between zero and one.

$$\mu_1(\varphi, T_{i1}, T_{-i1}) = \frac{\Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i,-i=H} + \Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i=L,-i=H}}{\Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i,-i=H} + \Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i=L,-i=H} + \Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i,-i=L}} \qquad [17]$$

Equation 17 expresses the conditional probability of a good incumbent in i in terms of the prior probability of a good incumbent in both the jurisdictions ($\Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i,-i=H}$), the prior probability of a bad incumbent in both the jurisdictions ($\Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i,-i=L}$), and the conditional probability of a bad incumbent in i given a good incumbent in $-i$ ($\Pr(T_{i1}, T_{-i1}) \mid \varepsilon_{i=L,-i=H}$).

Voters judge the incumbent as competent if the utility level he provides them is at least as large as the utility provided by the incumbents in the neighborhood, $u_{i,t}^V \geq u_{-i,t}^V$. Simple calculations show that the necessary condition for this inequality to hold is a null tax distance $T_{it} - T_{-it}$:

$$y_{it} - T_{it} + G_{it} \geq y_{-it} - T_{-it} + G_{-it} \qquad [18]$$

Keeping the assumptions y=1 and G homogenous among jurisdictions as in the baseline models:

$$1 - T_{it} \geq 1 - T_{-it} \qquad [19]$$

Rearranging:

$$T_{it} - T_{-it} \leq 0 \qquad [20]$$

When the tax distance is non positive and the pooling equilibrium is observed, the incumbent is considered as a good type and his probability of re-election increases.

In the static setting each electoral year the game repeats itself and the expected tax rate during the following term is obtained through Equation 16. That is, anytime a pooling equilibrium is observed the incumbent is considered as a good type. In such a situation the incentive for a bad incumbent to mimic increases.

The voting rule states that an incumbent is re-elected if his performance is no worse than the performance of a challenger whose ex ante probability of being good is measured by the parameter φ:

$$j_1 = \begin{cases} 0 & if \quad \mu_1(T_{\varepsilon_H}) + (1-\mu_1)(T_{\varepsilon_L}) > \varphi(T_{\varepsilon_H}) + (1-\varphi)(T_{\varepsilon L}) \\ 1 & if \quad \mu_1(T_{\varepsilon_H}) + (1-\mu_1)(T_{\varepsilon_L}) \leq \varphi(T_{\varepsilon_H}) + (1-\varphi)(T_{\varepsilon L}) \end{cases} \qquad [21]$$

which reduces to:

$$j_1 = \begin{cases} 0 & if \quad \mu_1 < \varphi \quad (non\_re-elect) \\ 1 & if \quad \mu_1 \geq \varphi \qquad (re-elect) \end{cases} \qquad [22]$$

Equation 22 claims that voters re-elect the candidate incumbent when the updated belief about his competence is at least as large as the ex-ante probability that the mayor is good type. When the updated belief on the expected tax rate during the second period is higher than the non-updated expected tax rate, the incumbent is considered bad type and he is not re-elected[5].

4.2 The dynamic model

In the dynamics setting, when voters face an incumbent who runs for re-election and sets a pooling tax rate, they update their beliefs considering not only the actual tax rate but also the information on the past performances in both jurisdictions governed by re-elected pooling incumbents and jurisdictions governed by challengers who defeated local incumbents.

If the two-period model is repeated the timing of the baseline model extends to the following stages:

      6. During period 2 the incumbent in *i* is term limited and he sets his rent-maximizing tax rate (T_{i2}); an election is held between two challengers with a majoritarian electoral rule;

      7. At the beginning of period 3 Nature selects a competence level of the new incumbent and a cost shock;

      8. The incumbent in *i* observes his competence level and his cost shock realization and sets a tax rate (T_{i3});

---

[5] The formal conditions for the inequality $\mu_1 \geq \varphi$ to hold are derived by Bordignon et al. (2003). Since the solution of the static model is not the main goal of this work, this Section does not go through the details of the two periods signaling game.

9. Voters in *i* observe the tax rates (T_{i3}) and (T_{-i3}), the realized tax rates (T_{i2}) and (T_{-i2}) conditional on the past electoral decisions, then they update their beliefs on the relative competence level of the incumbent in the neighborhood;

10. At the end of period 3 an election is held between the incumbent and a challenger with a majoritarian electoral rule;

11. At the beginning of period 4 Nature selects a cost shock and the game restarts; if the challenger has been elected his competence level is randomly selected by Nature.

The crucial factor determining the outcome of the election is now the probability that a learning process takes place during the third period. Learning is represented by a function L assumed to be bounded between zero and a maximum value $\overline{L}$, and it depends on three factors: the feasibility of learning (1-q2), the probability of gathering enough information (π) and the weight attached to past experience (ϱ). These three factors are independent from each other, e.g. a variation in the propensity to learn does not affect the realization of the cost shock or the weight attached to new information. Hence, L can be expressed as a product function:

$$L_3 = \pi(\alpha\rho - \beta\rho^2)(1 - q_2)$$

[23]

The feasibility of learning refers to the conditions stated in Proposition 1: if they do not hold, any information is useful in learning the incumbent's strategy[6]. Since the model assumes that a pooling equilibrium is observed during the first period, the respect of Proposition 1 relies on the realization of a positive shock in the jurisdiction governed by the good incumbent during the second period. Defining $0 \le q_2 \le 1$ as the probability of the realization of a negative cost shock during the second period in the jurisdiction governed by the good incumbent, learning is a decreasing function of q_2. The feasibility of learning is a factor exogenous to the model because voters' decisions do not affect it. However, as pointed out, it is a necessary condition for the process to work.

The probability that voters gather enough information to learn π is indeed an endogenous factor shaping L. The probability π depends on the propensity to learn. Learning requires a stock of information P* including the tax rates set in the neighborhood during the second and the third period. Voters are rational agents and they acquire new information when costs are no larger than benefits. The costs of obtaining information are represented by the marginal cost of obtaining both the domestic and the neighbors' tax rate information. The marginal cost of obtaining the information about the domestic tax rate is assumed to be small and constant, since a tax rate is information that the government must periodically release to claim its payment. The marginal cost of obtaining the neighbors' tax rate, on the contrary, is supposed to change depending on the size of the neighborhood. When the number of neighbors is large many pieces of information must be gathered and the total cost of being informed

---

[6] Assuming a constant cost shock learning leads to perfect selection. Relaxing this assumption allows learning the true type of only the past incumbent, which is still a relevant information because it unmasks a possible past mimicking behavior. If the current voting decision depends on past experience, it is a type of learning as well.

about the neighbors' performance increases. However, the information spillover generated by the inter-jurisdictional comparison of citizens may generate social networks giving rise to economies of scale in the diffusion of the information. The marginal cost of the information decreases as the number of neighbors increase. Being the size of the neighborhood constant in time, it affects the slope of the marginal cost function but not its shape. Finally, there is a cost attached to the action of retaining information, implying the effort of storing information in memory and being able to recall it when an election is approaching. The marginal information needs a larger memory capacity, therefore its cost increases with the size of the information stock retained.

The marginal benefit of being informed, on the contrary, is determined by the difference between the realized fiscal performance of the past incumbent during his second period of office, T_2, and the updated belief of the fiscal performance before his re-election, $E(T_2)$. To understand the reason for this specification, assume that the realization of the tax rate set by the past incumbent is higher than its expectation. Voters infer if the incumbent was strategic (bad) during the first period and they attach a larger marginal benefit to new information if compared with a situation in which the incumbent was non strategic (good). In other words, voters find it more convenient to improve their monitoring powers when they realize that their past beliefs have been mistaken and they become more prone to obtaining new information. The slope of the marginal benefit curve is assumed to be negative because voters may come out with a clear idea about the incumbent after having acquired the first pieces of information. In such a situation, their marginal utility from the marginal information decreases.

The following Graph depicts information (quantitatively measured) as a function of the marginal cost and the marginal benefit of gathering information. When the cost is larger than the benefit, voters do not to search for new information. When the benefit is larger than the cost voters find it profitable to gather new information up to the critical level P_1 pinned down by the intersection of the two curves. The quantity P_1 represents the maximum amount of information that voters would gather given the shape of the cost and benefit curves. The probability that voters obtain enough information to learn is the probability that P_1 is at least as large as a critical value P^*, $\pi = \Pr (P_1 \geq P^*)$.

Graph 1. Costs and benefit of gathering information

16
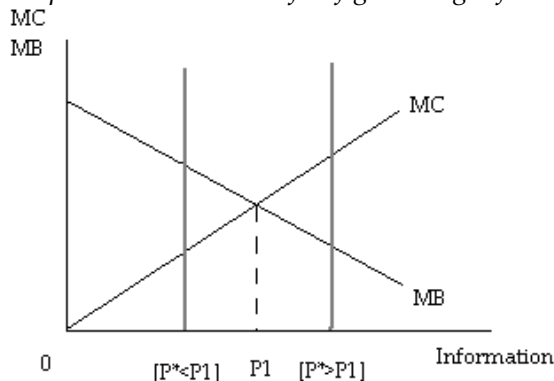
The last factor determining the learning function is the weight attached to the past experience ρ, computed through a Bayesian updating of voters' beliefs taking into account both present and past information. The process to come up with ρ follows Meseguer (2009) and relies on a larger set of assumption than the static updating explained in the previous section.

Assume that the tax distance T to be a random variable normally distributed with an unknown mean M and an unknown variance V. M and V are random variables, and voters learn about them by observing the performance of other incumbents under alternative past voting decisions j. The conditional distribution of the mean is Normal while the conditional distribution of the variance is scaled-Inverse X^2. The decision of these distributions is a classical assumption in Bayesian updating and allows the mean and the variance to be interdependent. Thus,

$$T_j = N(M_j, V_j)$$

$$M_j = N(m_j, \sigma_j^2 / \tau_j)$$

$$V_j = ScaledInv - \chi^2(v_j, \sigma_j^2)$$

[24]

Where m is the location of the mean, σ_j^2 / τ_j is the variation of the mean, v are the degrees of freedom and σ_j^2 is the scale of the variance, τ is the factor that relates the prior variance of the mean to the sampling variance.

During the third period the information available to voters about the performance of the incumbent is the tax distance under alternative voting decisions for all the jurisdiction that re-elected ($j=1$) or did not re-elect ($j=0$) the incumbent during the first period, $T_3^j \mid j$. The information is assumed to be a random variable independent and identically distributed. Hence, the sample mean and the sample sum of squares are sufficient statistics to summarize the information in the sample of countries under each of the alternative voting decisions. When prior beliefs are combined with new information, by applying the Bayes' rule the posterior belief about the mean and the variance of the tax distance is[7]:

$$\omega_3 = \rho\omega_2 + (1-\rho)\bar{x}_3$$

[25]

$$s_3^2 = \frac{S_3}{v_3}$$

[26]

where $0 < \varrho < 1$, ω_2 is the updated belief on the performance of the past incumbent at the end of the second period and $\bar{x}_3$ is the current observed performance of the incumbent; S_3 is the posterior for the sum of squares, v is the posterior for the degrees of freedom.

As Equation 25 shows, although extreme values of ϱ are ruled out, when the parameter is close to zero the past experience has a little influence on the updating process and voters hardly learn the determinants of the public cost function; vice versa, when ϱ tends to one the belief hardly takes into account new information. Learning is here assumed to

---

[7] For a detailed description of how to obtain this result, see Meseguer (2009), Appendix to Chapter 2.

be a quadratic function of ϱ, and its maximum value is associated to a level ϱ^* determined by the shape of the curve[8].

Graph 2. Learning as a function of rho
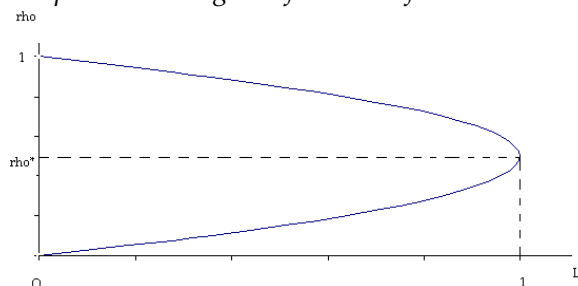
The function L is maximized when the conditions $\pi=1$, $\rho=\rho^*$ and $q_2=0$ jointly hold. On the contrary, no learning occurs if $\pi=0$ or $q_2=1$, that is if voters do not want or they cannot learn.

It is noteworthy to stress that the location of ρ^* indicates the optimal weight to attach to past experience. This parameter is fundamental when the electoral rule prescribes a term limitation, because the past incumbent is a different person than the current incumbent and the information gathering process may be stopped. Competence is in fact an individual specific characteristic, and if voters believe that the electoral strategy of the past incumbent does not affect the electoral strategy of the current incumbent in any possible way, ρ is close to one. The probability that the current incumbent is strategic, however, is not independent from the probability that past incumbents have been strategic. If a bad incumbent knows that his predecessor mimicked and he was re-elected (incumbents know the performance of the past incumbents), it is likely that he would play the same strategy. As a consequence, voters always gain positive utility from the marginal information since that is the only way to come out with a distribution of the type of the past pooling incumbents. The maximum value of the learning function is such that $1-\rho^*$ equals the probability that a bad incumbent in period 3 behaves as a bad incumbent in period 1 (equals to one) times the distribution of the type of the incumbents[9]. This condition is taken as an assumption here.

During the third period the problem of the voters is:

$$V_3(\varepsilon_i) = \max_j \left(u_3^V(T_3) + \beta \left(I_L * V_4^{VI}(\omega_3) + (1 - I_L) * V_4^{VI}(\mu_3) \right); u_3^V(T_3) + \beta(V_4^{VC}) \right) \qquad [27]$$

This maximization is similar to the maximization during the first period, but the function now depends on the belief updated using Equation 25 (instead of using Equation 16). Specifically, voters' expectations about the tax distance set during the fourth period $E(T_4)$ are:

$$E(T_4) = \rho T_3 + (1 - \rho)T_2 \qquad [28]$$

---

[8] Formally, $L = \alpha\rho - \beta\rho^2$. The function is maximixed when $\alpha/2\beta = \rho$.
[9] This condition suggests an alternative interpretation of ρ as the share of good incumbents in the population of incumbents during period 3. Of course voters lack this information.

Another innovation introduced in Equation 27 with respect to its counterpart in the static model is the presence of the probability function L in the second term of the inter-temporal utility function. The indicator function I_L indicates if the learning function has been maximizes. When voters optimally learn about the incumbent's true type in the dynamic setting, in fact, the updated beliefs on competence do not depend exclusively on the tax distance observed during the current electoral year (T_3). For $I_L = 0$, learning does nor occur, the static setting is restored and given a pooling equilibrium during period 3, he will be re-elected as long as the condition for successful pooling hold.

If learning occurs $I_L = 1$, two situations may arise. In the first one the past incumbent played a pooling equilibrium during the second period, $E(T_4) = 0$ and the current pooling incumbent will be re-elected in period 3. In other words, perpetual pooling does not provide voters with any useful new information. Of course, this situation is driven by a factor $q_2 = 1$, indicating that a positive shock to the good incumbent occurred during the second period. Optimal learning $I_L = 1$ implies $q_2 > 0$, and $E(T_4) = 0$ is ruled out by definition.

In the second situation the past incumbent did not play a pooling strategy during period 2 and T_2 conveys with useful information. Specifically, the past incumbent is recognized as bad type and his behavior has consequences on his successor. The updated tax distance $E(T_4)$ becomes positive and voters do re-elect the incumbent in period 3 only if $I_L * V_4^{VI} \geq V_4^{VC}$. The equality is included in the condition because we assume that voters have a strong preference for dismissing an incumbent who cheated them[10].

Given that the updating process provided voters with posterior beliefs on both the mean and the variance of the tax distance set by the incumbent during the fourth period, the expected utility V_4^V reasonably takes into account both these information.

$$V_4^j = EU(\omega_3, s_3) \mid j = \beta_1 \omega_3^j + \beta_2 s_3^j + \varepsilon_3^j \tag{29}$$

Where ε is an error term included to account for the inaccuracy of the voters in updating their beliefs and j is the re-election probability. Voters re-elect the incumbent if:

$$EU(\omega_3, s_3) \mid (j=1) \geq EU(\omega_3, s_3) \mid (j=0) \tag{30}$$

In case of a defeat of the incumbent, a challenger is elected. The right hand side of the equation can thus be interpreted as the expected utility from electing a challenger.

Substituting Equation 29 in Equation 30 and rearranging terms we obtain:

$$\beta_1 \left(\omega_3^I - \omega_3^C \right) + \beta_2 \left(s_3^I - s_3^C \right) \geq \varepsilon_3^C - \varepsilon_3^I \tag{31}$$

where I stands for the incumbent and C for the challenger. Denoting the differences with a d, $\mu_3^d = \mu_3^{j=1} - \mu_3^{j=0}$, $s_3^d = s_3^{j=1} - s_3^{j=0}$ and $\varepsilon_3^d = \varepsilon_3^{j=1} - \varepsilon_3^{j=0}$, we obtain:

$$\beta_1(\mu_3^d) + \beta_2 s_3^d \geq \varepsilon_3^d \tag{32}$$

The probability that a voter re-elects the pooling incumbent during the third period is:

---

[10] The incumbent in this case has a probability of being good equal to zero, while the probability that a challenger is good is above zero.

$$P(j=1) = P\left(EU_3^I \geq EU_3^C\right) = P\left(\varepsilon_3^d \geq \beta_1(\mu_3^d) + \beta_2 s_3^d\right) =$$
$$= 1 - F\left[-\left(\beta_1\mu_3^d + \beta_2 s_3^d\right)\right] = F\left(\beta_1\mu_3^d + \beta_2 s_3^d\right)$$

[33]

Equation 33 allows us to summarize the theoretical predictions of the dynamic model.

If the incremental learning by tax rates presented in this paper occurs, β_1 and β_2 are significant and negative coefficients. A large mean of the tax distance is associated with an incumbent extracting rent, while a large volatility of the tax distance is associated with an ambiguous outcome. Assuming that voters are risk averse and prefer certainty of policy outcomes rather than uncertainty, also β_2 is expected to be negative.

If the learning process does not take place, due to the failure of Proposition1, a stock of information insufficient for learning, or almost extreme weight attached to past experience, updated beliefs on the tax distance do not have any influence on the decision to re-elect the incumbent. As a consequence, the coefficients in Equation 33 will be not statistically significant.

5. An empirical test of the dynamic learning from tax rates

5.1 Italian Municipalities: institutional setting, accountability system and yardstick competition

Municipalities are the lowest tier of government in Italy, and they are a suitable framework for this paper's empirical test. In the early 1990s, in fact, an institutional reform aimed at strengthening local accountability introducing tax decentralization and reforming the electoral rule, setting the framework for yardstick competition to arise.

The local property tax rate (ICI, Imposta comunale sugli Immobili), introduced in 1993, increased the tax autonomy of local governments and in the period 1993-2007 accounted for more than 55% of total Municipality revenue and more than 25% of local expenditure11. ICI is a highly autonomous tax rate, specifically a level b in the OECD tax autonomy scale ranging from ´a´ to ´e´ (OECD, 1999). The previous setting was characterized by the lowest degree of tax autonomy, level e, being the tax rate and the tax base both set by the central government. In 1995 the tax rate has been differentiated between the house tax rate applied to the main living property and the business tax rate applied to holiday houses, offices, shops, and so on. Local house property taxation accounts only for 6% of local tax revenues, but it is a cost that voters directly link to the house and makes it clear to the citizens the relationship between the costs and the benefits of local public services in a certain jurisdiction. In addition to this, more than 80% of the residents in Italy are home-owner12, making the local house tax rate the main indicator of jurisdictional performance. Since the tax base is fixed and property value reassessments are nationally implemented, local autonomy is restricted to only one dimension, the tax rate level. The tax rate can be set in a range between 4‰ and 7‰. Although the tax interval is small, a marginal variation of the tax rate determines a consistent variation in the per capita tax paid by the citizen and in the overall tax

---

[11] Source: *ANCI*, National Association of Italian Municipalities.

[12] Source: ISTAT, *L'abitazione delle famiglie residenti in Italia* ‑ *Anno 2008*, published in Spring 2010.

revenue[13]. Moreover, the single dimension of the decision makes it easier for the voters to exploit this information when forming their voting preferences.

Regarding election, the Italian local electoral rule has been reformed in 1993 from proportional to majoritarian, introducing the direct election of the mayor according to the plurality rule in Municipalities with less than 15000 inhabitants (9% of the total number of Municipalities) and according to the majority rule with runoff elections in the others. The local legislature lasted four years before 1999 and it has now been extended to five years, and a term limit is fixed to two terms. In case of motion of no confidence both the mayor and the council must resign and new elections are held. Because of the early fall of many executives in the past Italian Municipalities hold elections in different years. There is, however, a concentration of local elections in 1995, 1999 and 2004, when more than 60% of the jurisdictions are called to the ballot. These three years are considered as 'first order electoral years'.

There is evidence of strategic tax setting among Italian Municipalities, as studied by Bordignon et al. (2003), Santolini (2007), Bartolini and Santolini (2009)[14]. Given these facts it is natural to guess that selection powers have been constrained and a certain number of mimicking incumbents have been re-elected to the detriment of their citizens. However, the results of the model in Section 4 predict that election after election, voters may learn the incumbents' strategy and since the third electoral period they can correctly judge the fiscal performance. The next paragraph tests this hypothesis.

5.2 Methodology and data

The methodology applied in this analysis stems from the model of learning from economic policies by Meseguer (2009). This section adapts the original cross-countries economic policy decision setting to the sub-national electoral learning environment.

During the third electoral year of the model of dynamic learning from tax rates voters choose among two different electoral policies, which are not re-elect the incumbent who pools (j_3=0) or re-elect the incumbent who pools (j_3=1). Pooling is defined as setting a tax rate which is no higher than the tax rate in the neighborhood. Equation 33 makes the empirical link between re-election and the updated beliefs.

The method suggested by Meseguer (2009) includes three-steps:

---

[13] The average value of the house properties in Italy was 182000 euro in 2008 (source: Dipartimento delle Finanze and Agenzia del Territorio, *Gli Immobili in Italia,* published in 2010). Using this value as a proxy for the tax base of ICI, a marginal variation in the tax rate leads to a variation of 182 euro of the individual tax burden. In turn, this amount accounts for a 7‰ of the he average yearly income of an employee in 2009 (ISTAT).

[14] The findings of Chapter 1 of the author's Ph.D. Dissertation confirm the previous results in the literature, since neighboring jurisdictions set similar tax rates during the electoral year. The dataset implemented, however, is significantly wider and longer, including all the Municipalities belonging to the 15 Italian Ordinary Regions from 1995 to 2004. The same work estimated both a local vote popularity equation and a local tax distance equation – tax distance is computed as the difference between the domestic tax rate and the neighbors' tax rate - accounting for the simultaneity between the local tax distance and the local win margin. The results are consistent with the static model of yardstick competition, since a decrease in the electoral tax distance is associated with an increase in the electoral popularity of the incumbent, and an increase in the incumbents' expectation of popularity is associated with a decrease of the tax distance during the electoral year.

1. the posterior beliefs μ_3^d *and* s_3^d are calculated using Bayesian updating;
2. posterior beliefs conditional on the voting decision are compared. The prediction is a worse performance (higher expected average and a higher expected volatility) associated to the non re-elected incumbents;
3. a regression is estimated, using as dependent variable the voting decision j.

The function estimated in the original model is:

$$j_3 = \beta_1\mu_3^d + \beta_2 s_3^d + \beta_3 X_3 + \xi_3^d \qquad\qquad [34]$$

The main independent variables are the differences in posterior beliefs about average results and variability of results under each status, $\mu_3^d = \mu_3^{j=1} - \mu_3^{j=0}$ and $s_3^d = s_3^{j=1} - s_3^{j=0}$, and X is a vector of explanatory variables.

Meseguer updates the agent's beliefs conditional on policy decision using data on GDP growth, a variable which is yearly observable.

The electoral scenario of Italian Municipalities, however, is much more complicated. The dataset is a cohort of 228 obbservations. The mean and the variation of the tax rate is observed during every electoral period, but the neighbors do not hold necessarily belong to the cohort, therefore they do not hold elections in the same years. As a consequence, the descriptive statistics for the realization of the tax rate conditional on the past voting decision is not feasible if not picking different electoral years for the neighbors. This option is not followed at this stage of the analyses for two main reasons. The first one is theoretical, because this approach would catch different years of the political electoral cycle, making the values not comparable. The second one is technical, because such a computation complicates the calculus of the beliefs and would not be possible for data before 1995.

Hence, the empirical test of this paper estimates Equation 29 separately for each policy decision, focusing on the domestic past experience. Since the dependent variable takes values zero or one, and since the available data refer to only one year, the chosen model is a static probit. The null hypothesis states that voters did not switch from re-electing the pooling incumbent to not re-electing him as a result of dynamic learning from tax rates. The empirical predictions are that β_1 and β_2 should be significantly negative because both a high average and a high volatility of the tax rate reduce the voters' utility.
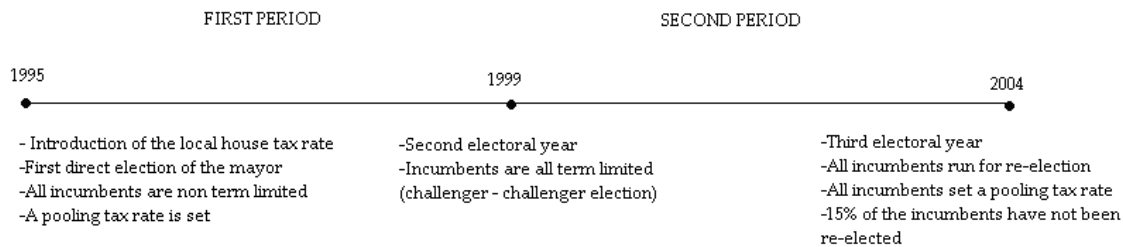
The data used for the estimation come from a comprehensive dataset of Italian Municipalities (Padovano, 2007). The considered observations are 227 Municipalities meeting the following requirements:

- are members of the cohort of Municipalities that held local elections in 1995, 1999 and 2004;
- a local house tax rate set in 1995 at most equal than the average tax rate set by its neighbors (defined as a 'pooling' tax rate);
- a local house tax rate set in 1999 higher than the average tax rate set by its neighbors (defined as 'non pooling' tax rate);
- an incumbent running for re-election in 2004;

- a local house tax rate set in 2004 at most equal than the average tax rate set by its neighbors ('pooling' tax rate).

As the following graph shows, the selected observations are in their third electoral year since the local fiscal and electoral system has been reformed, and they belong to a cohort of jurisdictions experiencing two full local legislatures (1995-1999, 1999-2004). Among them, in 2004 the incumbent was defeated in 33 Municipalities (about the 15% of the sub-sample) while in the remaining 194 Municipalities he was re-elected.

Graph 4. Electoral dynamics of the 227 Municipalities in the dataset

FIRST PERIOD   SECOND PERIOD

1995   1999   2004

- Introduction of the local house tax rate   -Second electoral year   -Third electoral year
-First direct election of the mayor   -Incumbents are all term limited   -All incumbents run for re-election
-All incumbents are non term limited   (challenger - challenger election)   -All incumbents set a pooling tax rate
-A pooling tax rate is set   -15% of the incumbents have not been
                                                                                              re-elected

5.3 Estimation and results

The first empirical step implies computing the posterior beliefs in 2004 using Bayesian updating.

Several set of priors have been used to calculate different updated beliefs.

The first update considers as priors the average and the variability of the tax distance in the dataset (measured as the standard deviation from the possible interval of values of the tax distance). The tax distance is measured as the difference between the domestic tax rate and the average tax rate in the neighborhood. This set of priors (*uptd*) is closer to the specification of the model presented in this paper, but since the existing literature on yardstick competition focused separately on the domestic and the neighbors' tax rate, alternative sets of priors have been investigated.

The alternative sets of priors calculate updated beliefs with respect to the average and the variation of the domestic tax rate, taking as priors the average and the variation from the possible interval (*up1*) or from the average and the variation from the observed values in the neighborhood (*up2*).

Summary statistics for the posterior point estimates for the location and the scale are reported in Table 7.

Table 7. Posterior beliefs using different sets of priors

own xp	priors	Obs	Mean	Std. Dev.	Min	Max
Updated average, μ3		227	4.776	0.477	4.000	5.880
Updated variance, s3	UP1	227	0.894	0.953	0.000	2.638
P		227	0.364	0.026	0.333	0.4
1-ρ		227	0.636	0.026	0.595	0.67
Updated average, μ3		222	4.769	0.485	4.000	5.878
Updated variance, s3	UP2	222	0.331	0.273	0.020	2.024
P		223	0.351	0.012	0.334	0.41
1-ρ		223	0.649	0.012	0.594	0.67
Updated average, μ3		227	-0.504	0.372	-1.696	0.159
Updated variance, s3	UPTD	227	0.497	0.741	0.000	5.768
P		227	0.353	0.022	0.333	0.43
1-ρ		227	0.647	0.022	0.573	0.67

The mean updated domestic tax rate using both the sets of priors is about 4.77 (note that the tax rates are scaled between 4 and 7), but the variation is smaller when using the set of priors exploiting the neighbors' information. These figures suggest that benchmarking the domestic performance with the neighboring performance allows voters to have a more precise expectation of the future performance.

When a direct indicator of performance comparison is used and voters' belief are updated with the priors on the tax distance, $\mu3$ ranges from -1.696 to 0.159, with a negative tax distance mean of -0.504. In other words, in some Municipalities voters expect a bad performance (positive tax distance) and in other Municipalities they expect a good performance (non positive tax distance)

From these results we can also see that the contribution of past information to the updating process is stable at about 35% regardless the specification of the priors[15]. These results for ϱ suggest that voters form their electoral beliefs taking into account both the current incumbents' performance and the past performance. Meseguer (2009) argues that a low value of ρ indicates that the learning process has already occurred, while a high value tells that new information is still relevant for voters and in time they will complete the learning. We can comment that a learning process had started in the analyzed sample, but we cannot say if this is the level of ρ that maximizes the learning function. Consequently, this information does not answer the question if a learning process took place or not.

A closer look at the posterior beliefs is helpful in asking this question. Table 8 reports the posterior beliefs conditional on the voting decision.

---

[15] In the updating with the neighbors priors the number of observations has been set to one and does not equal the number of neighbors. The reason for this choice is that the information included in the formula is the average tax rate in the neighborhood and not the single tax rates in the contiguous jurisdictions. Specifying the number of neighbors leads to an underestimation of rho and to updated average neighbors tax rates falling outside the feasible set.

Table 8. Posterior beliefs conditional on voting decision

re-elected incumbent				not re-elected incumbent			
Variable	*priors*	*Obs*	*Mean*	*Variable*	*priors*	*Obs*	*Mean*
Updated average, μ3	UP1	194	4.79	Updated average, μ3	UP1	33	4.71
Updated variance, s3		194	0.89	Updated variance, s3		33	0.89
Updated average, μ3	UP2	189	4.78	Updated average, μ3	UP2	33	4.71
Updated variance, s3		189	0.33	Updated variance, s3		33	0.35
Updated average, μ3	UPTD	194	-0.50	Updated average, μ3	UPTD	33	-0.53
Updated variance, s3		194	0.49	Updated variance, s3		33	0.55

This comparison of the updated beliefs on the tax levels does not support the learning hypotheses since the level of the posterior belief about the average tax rate for the incumbent re-elected in 2004 is always higher than the level of the posterior for the incumbent non re-elected in 2004 (4.79>4.71 using UP1, and 4.78>4.71 using UP2). The tax distance, however, is expected to be larger for non re-elected incumbents. An explanation for this results is that voters consider the fiscal comparison and not the domestic tax level as a performance indicator.

The results regarding the variation of the updated beliefs disaggregated by the incumbent status indicate as expected that the re-elected incumbent is always associated with a smaller or equal variation than the non re-elected incumbent. Voters, thus, should behave as risky adverse agents.

At this stage of the analyses it is interesting to perform a comparison based on the history of voting decision. If a learning process occurred we expect that the average updated beliefs in the jurisdictions switching from re-election in 1995 to not re-election in 2004 (coded as ´RNR´) should be higher than the updated beliefs in the jurisdictions that re-elected the incumbent in 2004 (coded as ´NRR´). The summary statistics in Table 9 support this hypothesis only when the updating process exploits the set of priors UP2, (column 8). This figure suggests that a learning process may have occurred if voters updated their beliefs based on the performance in the neighborhood.

Table 9. Comparison of posterior beliefs with respect to the history of voting decisions

						RNR>NRR
Variable	*priors*	*RR*	*NRR*	*NRNR*	*RNR*	*Column 8*
Updated average, μ3	UP1	4.792	4.775	4.507	4.769	FALSE
Updated variance, s3		0.895	0.893	1.324	0.748	FALSE
Updated average, μ3	UP2	4.785	4.763	4.506	4.769	**TRUE**
Updated variance, s3		0.326	0.337	0.634	0.253	FALSE
Updated average, μ3	UPTD	-0.493	-0.521	-0.560	-0.525	FALSE
Updated variance, s3		0.467	0.559	0.796	0.473	FALSE
Observations		*150*	*44*	*8*	*25*	

Notes: Rr=re-elected in both 1995 and in 2004; Nrr=not re-elected in 1995 and re-elected in 2004; Rnr=re-elected in 1995 and not re-elected in 2004; Nrnr = not re-elected in both 1995 and in 2004. 227 total observations.

These results still give ambiguous indication about the presence of a learning process in the sample, therefore equation 34 must be estimated to measure the extent to which the

updated beliefs determine the voting decision of the citizens. Specifically, the vote decision is regressed upon the posteriors for the tax distance and both own and regional experience. The time period available, however, poses some problems in the specification. A time period of three subsequent elections tightly fits the model but the experience observed inside each Municipality is either re-election in 1995 or not re-election in 1995. This characteristic of the data does not allow us to compute the differential between the updated beliefs conditional on the policy decision, but we still can analyze it separately.

Table 10 presents the results from probit estimations obtained without covariates (Model 1-4) and with covariates[16] (Model 5-12).

---

[16] The covariates included are considered relevant determinants of the voting decision in the literature (see Petterson-Libdom, 2006).

Table 10. Dynamic learning from tax rates, probit regression, marginal effects

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx	dy/dx
tax(i, 04)	0.006					-0.002							
tax(-i, 04)	0.040					0.029							
tax distance(i,04)		-0.001					-0.006						
μ3UP1*reel95			0.001						-0.003				
s3UP1*reel95			0.020						0.027				
μ3UP2*reel95				-0.009					-0.014		-0.019	-0.013	
s3UP2*reel95				0.211					**0.229 .**		**0.240***	0.206	
μ3UPTD*reel95					0.073					0.097			0.110
s3UPTD*reel95					0.50					0.057			0.070
right party						-0.957	-0.062	-0.064	-0.061	-0.063	-0.064	-0.060	-0.063
unemployment rate 03						-0.195	-0.191	-0.246	-0.323	-0.204	-0.248	-0.267	0.419
popularity 99						0.073	0.071	0.073	0.100	0.083	0.067	0.074	0.065
past performance tax level											**0.097***		
past performance tax distance												0.003	0.010
performance tax distance												-0.089	-0.091
Indep. Vars.	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Updating	Static	static	dynamic	dynamic	Dynamic	static	Static	dynamic	dynamic	dynamic	dynamic	dynamic	Dynamic
level or distance	Level	distance	level	level	distance	level	Distance	level	level	distance	level	level	Distance

Note: dependent variable: binary variable equal to one if incumbent re-elected in 2004 and zero otherwise.

The fit of the model is very limited, and the coefficients do not show a high degree of significance. However, there are some expected results.

Model 1-2 and Model 6-7 test the static yardstick competition model. The domestic tax rate shows the expected negative sign only when the covariates are include (Model 6). The neighbors' tax rate shows an unexpected positive sign, while the coefficient for the tax distance is negative as the theory predicts.

Models 3-4 and Models 8-9 tests the dynamic learning from tax rates including the updated beliefs UP1 and UP2 interacted with the dummy indicating re-election in 1995. the interaction term has been chosen to control the effect of the updated belief conditional on past re-election or not. The results using UP2, which seems to use specification of the priors more appropriate in this frameowrk, confirm that an increase in the expected average tax rate in the Municipalities re-electing a pooling incumbent in 1995 reduces the probability of being re-elected in 2004 with respect to the Municipalities not experiencing a past pooling incumbent re-elected, while an increase in the variation of the tax rate in the same Municipalities is associated with an increase in the probability of being re-elected with respect to the control group. In Model 9, moreover, the updated belief on the variation is statistically significant at the 10% level.

Model 5 and Model 10 control for the updated beliefs on the tax distance, but the results are both not significant and unexpectedly positive.

Model 11 introduces as covariate the past performance (tax rate 1999 – tax rate 1995). The results are significant but at odds with the theory because voters do not seem to punish a term limited incumbent performing badly. The last two Models, 12 and 13, replicate the results of Model 9 and 10, but including the past performance on the tax distance (td99-td95) and the current performance on the tax distance (td04-td99). The results indicate that the probability of being re-elected increases with the increase of the past performance and decrease with an increase of the current performance. In other words, voters seem to punish the first term incumbent performing worst than the past term limited incumbent, but not the present pooling incumbent if his predecessor was unmasked as a mimicking bad incumbent.

To sum up, the coefficients of the regressions are rarely significant and they show a pattern of learning when using updated beliefs on the average tax rates based on the priors UP2 (Model 4, 9, 11, 12), but this result is not backed by a negative coefficient of the updated belief on the variation using the priors UP2.

6. Concluding remarks

The political economics literature recognized that re-election mechanism is an imperfect device to select good politicians when the candidate incumbent behaves strategically. This paper analyzes how persistent is the reduction of selection powers caused by yardstick competition. The theoretical results of the model show that a learning process restores selection powers by granting the correct update of voters' beliefs also when a pooling equilibrium in tax rates arises during the electoral year. When learning occurs voters overcome the asymmetric information on the cost of the public production of goods and services and they infer the competence level of the current incumbent. The

learning process, however, relies on a set of exogenous and endogenous factors that hardly take place together, as the feasibility of learning, the willingness to learn, the weight attached to past experience. Furthermore, the assumption of a Δ constant in time is required to learn the true type of the current incumbent. If this assumption does not hold, voters realize the extent to which mimicking occurred in the past and if they have been deceived by the past incumbent, but the way in which this consciousness affects their voting behavior cannot be univocally derived.

In order to test the model of dynamic learning from tax rates an empirical test has been conducted on a sample of Italian Municipalities characterized by a pattern of yardstick competition. The results reject the presence of a voters' learning process in the data because when voters face a pooling equilibrium the re-election of the incumbent is not influenced by the updated beliefs on his fiscal performance, as suggested by the non significant marginal effects. The results confirm that the advantages coming from yardstick competition, that is the reduction of the incumbent's rent during the electoral year, has to be weighed against a reduction of voters' selection powers that the informational spillover is not able to remove, even in the long run.

This paper is the first attempt at analyzing the consequence of yardstick competition on selection powers, therefore it calls for future research. A natural extension of the dynamic model is investigating voters' learning from different stock of information than the ones considered so far. The relevant information, as an example, could be the whole term performance of the candidate incumbent measured as the intra-term mean and variation of the tax rate. The literature on strategic interaction in expenditure levels should also suggest to exploit expenditure measures as performance indicators. These perspectives could shed light also on the interaction between the political business cycle in taxation with the cycle in expenditure.

REFERENCES

Bar-Isaac (2003), 'Reputation and Survival: learning in a dynamic signaling model', *Review of Economic Studies, Blackwell Publishing, vol. 70(2), pages 231-251, 04.*

Benabou, R., Gertner, R., (1993) 'Search with learning from prices: does increased inflationary uncertainty lead to higher markups', *The Review of Economic Studies,* Vol.60, N.1, 69-93.

Besley, T., Case, A. (1995a), 'Incumbent behaviour: vote seeking, tax setting, yardstick competition', *The American Economic Review, vol. 85, n. 1.*

Bordignon M., Cerniglia F. and Revelli F. (2001), 'In search for yardstick competition: property taxes, tax rates and electoral behaviour in Italian cities', *Workshop on: strategic interaction among local governments: empirical evidence and theoretical insights,* Milan, Università Cattolica del Sacro Cuore, 11/12 May.

Bose S., Orosel G., Ottaviani M., Vesterlund L. (2008), 'Monopoly pricing in the binary herding model', *CEPR Discussion Papers 5003.*

Delgado, F. J., Lago-Peñas S., Mayor M., (2011), 'On the determinants of local tax rate: new evidence from Spain', *IEB Working Paper*, 2011/4.

Gilardi, F. (2010), 'Who Learns from What in Policy Diffusion Processes?' *American Journal of Political Science, 54(3): 650-666.*

Meseguer C. (2006), Learning and Economic Policy Choices, *European Journal of Political Economy Vol. 22 (2006) 156– 178.*

Meseguer C. (2009), Learning, Policy Making, and Market Reforms, *Cambridge University Press.*

Ottaviani M. (1999), 'Monopoly Pricing with Social Learning', *Doctoral Dissertation, MIT.*

Padovano, F. DATASET

Santolini, R. (2009), `The political trend in local government tax setting´, *Public Choice*, 139, 125-134.

Schaltegger C. and Kuttel D. (2002), 'Exit, Voice, and Mimicking Behavior: Evidence from Swiss Cantons', *Public Choice, Springer, vol. 113(1-2), pages 1-23.*

Solè Ollè A. (2003), 'Electoral accountability and tax mimicking: the effects of electoral margins, coalition government, and ideology', *European Journal of Political Economy Vol. 19 (2003) 685–713.*