CENTRAL AND LOCAL TAX AUDITING

DAVID BARTOLINI AND FABIO FIORILLO
CENTRAL AND LOCAL TAX AUDITING

DAVID BARTOLINI AND FABIO FIORILLO

Università Politecnica della Marche and OPERA, Ancona

Abstract. In a federal state the process of fiscal decentralisation assigns to regions the provision of local public goods/services and the power to impose taxes. Little is said, however, about the enforcement of such local taxes. In particular, we focus on the following question: should the enforcement of a local tax be assigned at the local or central level? We propose a model where local and central tax authorities need to enforce tax compliance through a costly monitoring and auditing policy. We show that the assignment of tax enforcement may not follow the same criteria as tax assignment: it might be optimal to assign the enforcement and the tax to two different levels of government.

JEL Classification:
Key words: fiscal federalism, tax auditing

1. INTRODUCTION

The process of fiscal decentralisation consists in assigning government functions at lower levels of government, with the explicit goal of increasing social welfare. This topic has been widely investigated by the economic literature with several contributions that shed light on the benefits and costs of assigning taxes and the provision of public goods at the local level (Oates, 1999). As regards optimal taxation a key factor is the mobility of agents across jurisdictions, which, on one hand, can help matching public goods to preferences (Tiebout), but on the other hand makes the application of local taxes more complex (fiscal externality, tax competition, etc.). According to Wellisch (2000) a complete (optimal) tax system finances the efficient level of public goods without distorting locational patterns.

This approach does not take into account the possibility that people evade taxes. Most of the works implicitly assume that the amount of taxes imposed is costlessly raised. Obviously this is not the case, in reality tax authorities engage in expensive auditing activities. Since the objective of the optimal tax system is to raise revenues minimising distortions, the possibility to evade would undermine the whole tax structure because citizens would face a different marginal tax rate. The tax authority can actually raise taxes either by increasing the legal tax rate or tightening enforcement of the existing tax system (Kaplow,
1989). However, even the literature that considered the effect of evasion on the optimal tax system has not considered the problem of tax enforcement in a decentralised setting. When we consider a federal state the question of who should enforce taxes, naturally arise. In a scenario in which taxes are assigned to the central and local level according to the standard principles of fiscal federalism, we ask what rules should be followed regarding tax enforcement. In particular, should the central government enforce all taxes? Should a local tax be enforced at the local level? In the latter case the local government would be in charge for the enforcement of the local tax. In other words, our objective is to investigate the best assignment of auditing functions – centralised, decentralised or separate – to raise tax revenue at the lowest cost.

We consider a federal state organised in a central government and two local governments. The tax system consists in a central tax and two local taxes – one for each local jurisdiction. The amount of tax base reported by citizens depends on the probability of being audited. In order to enforce taxation the authority in charge (it would depend on the scenario we consider), uses some resources for monitoring and auditing activity. Therefore the amount of taxes collected is gross of the cost of enforcement.

In this setting information plays a crucial role. The “productivity” of the resources devoted to the enforcement activity depends on the availability of information. We consider two types of information: direct information and information spillovers. The former stems from the possibility to directly observe citizens’ characteristics (e.g. standard of living, number of children, housing etc.). This information would impact on the enforcement cost of the central and local level according to the degree of mobility of the tax base: a mobile tax base is better monitored by the central level, while a less mobile tax base is better audited at the local level.

Information spillovers arise across tax bases when they are correlated, so that the auditing activity on the local tax can bear useful information about the central tax base, or viceversa. The magnitude of this externality depends on the connection between tax bases, for instance the base of an income and a wealth tax are strictly linked. A necessary condition to exploit this spillover is either full cooperation among tax authorities or the presence of a single tax authority (Esteller-Moré, 2004).

The optimal tax assignment relies on the degree of mobility in order to assign taxes at the different levels of government. It seems reasonable to follow the same principles also

---

1As a reference for the study of optimal taxation with tax evasion, see Sandmo (1981) for income taxation, and Cremer and Gahvari (1993) for commodity taxation.
for the assignment of the enforcement power, however, the presence of spillovers among tax bases may lead to a different assignment of auditing responsibilities. For instance, it might be that an immobile tax base is strongly linked to a more mobile tax base, so that, even though the optimal allocation of taxes is achieved by decentralisation, it might be optimal to have a centralised tax enforcement.

Our analysis does not exhaust all the issues related with tax enforcement. For instance, we do not consider agency problems that may arise when functions are delegated and different agents have different objective. Furthermore, we do not touch the problem of fiscal competition\(^2\).

1.1. Related literature. As mentioned in the previous section the economic literature has focused on fiscal decentralisation assuming away any issue related with tax enforcement. A partial exception is Cremer and Gahvari (2000) which focus on a two country model where the enforcement activity is conducted independently by the two countries. They consider fiscal integration taking into account that countries have an alternative instrument to conduct fiscal competition: tax enforcement. Our model is substantially different since we look at a federal system where citizens face both a central and a local tax, and where there is less scope for tax competition.

A work closer to our paper is Esteller-Moré (2004) which consider a model with two interrelated taxes. The idea is that auditing one tax base provides information on the other tax base. He assumes that the possibility to check congruence in tax reports is limited by imperfect collaboration between tax administrations. Indeed, if collaboration was perfect congruity would be the optimal tax payer’s strategy. This model differs from ours in two main aspects. Firstly we consider a degree of externality between the two tax bases that depends on their similarity (irrespective of collaboration), and, secondly, we consider a costly auditing activity.

2. Structure of the model

We consider a federation of two regions \((i = 1, 2)\), with a central and a local tax (in each region). The two regions are identical in terms of population and income. Moreover, we consider identical citizens, so that the propensity to evade is the same in both regions. The objective of the federal (central) and local governments is to raise tax revenue in the most efficient way (i.e. at the lowest cost). Since there is an information problem tax

\(^{(Cremer and Gahvari, 2000)}\) propose a two-state model in which fiscal competition in on the degree of effort put in tax auditing.
authorities need to enforce compliance through costly policies, such as monitoring and auditing activities. Therefore each tax authority decides the amount of resources to devote to the enforcement of the assigned tax base.

Information plays a crucial role for monitoring technology. We distinguish between direct and spillovers information. The direct information stems from direct observation of the characteristics of tax base. Information spillovers arise from the correlation between different tax bases (central and local). If we analyse one tax in isolation mobility would be the determinant characteristic for tax assignment. Each authority benefits from specific direct information depending on the mobility of tax base. The local authorities’ direct information is negatively correlated to the mobility of the tax base, in this case the proximity to citizens allows better knowledge of tax evasion. On the contrary the direct information of the central authority is positively correlated with mobility, evasion on mobile tax bases can be better discovered by a central authority.

We model this direct information structure as an iceberg cost on tax revenues. For simplicity we consider a dicotomic mobility: a tax base is either mobile or immobile. Thus the central authority which monitors an immobile tax base incurs in a cost \( C \geq 0 \), while the local authority which monitors a mobile tax base incurs in a cost \( c \geq 0 \). In particular the structure of such iceberg cost is described in table 1.

**Table 1. Structure of direct information costs**

<table>
<thead>
<tr>
<th></th>
<th>mobile</th>
<th>immobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>central</td>
<td>0</td>
<td>( C )</td>
</tr>
<tr>
<td>local</td>
<td>( c )</td>
<td>0</td>
</tr>
</tbody>
</table>

When we consider the whole tax system, possibly composed by more taxes, then an information spillover between tax bases arises. When tax bases are correlated, information gathered from one tax base monitoring activity can be useful for the others\(^3\).

2.1. **Monitoring technology.** We assumed that when audited the true level of tax base is disclosed. The auditing activity has a cost, which reflects the relationship between resources devoted to auditing and probability of being audited. We define the probability of being audited as an increasing concave function of the resources allocated. For instance for a local tax the probability of auditing \( p_i \) \( i = 1, 2 \) depends on resources devoted for monitoring \( r_i \) \( i = 1, 2 \), in this way \( \frac{\partial p_i(r_i)}{\partial r_i} > 0 \) and \( \frac{\partial^2 p_i(r_i)}{\partial r_i^2} \leq 0 \).

\(^3\)Aggiungere tabella su correlazione tra basi imponibili e relativi commenti.
The effect of resources on the probability of being audited depends on the level of
government which performs the task and on information spillover\(^4\). For simplicity we
assume that the direction of the spillover is from local to central taxes.

The auditing function of national taxes in region \(i\) is

\[
P_i = P_i(R_i, p_i)
\]

with

\[
\frac{\partial P_i}{\partial R_i} = \gamma > 0; \quad \frac{\partial^2 P_i}{\partial R_i^2} = \gamma' = 0; \quad \frac{\partial P_i}{\partial p_i} = \epsilon
\]

Thus we assume the marginal productivity of resources on monitoring (\(\gamma\)) positive and
constant and it does not depend on the monitoring activity on the other taxes, \(\epsilon\) is the
marginal spillover which we assume constant as well.

The auditing function of local taxes in region \(i\) is

\[
p_i = p_i(r_i)
\]

with

\[
\frac{\partial p_i}{\partial r_i} = \lambda > 0; \quad \frac{\partial^2 p_i}{\partial r_i^2} = \lambda' = 0
\]

The marginal productivity of resources on monitoring (\(\lambda\)) is positive and constant\(^5\).

The reduced form of equations 2.1 and 2.2 has the following Jacobian matrix \(J\):

\[
J = \begin{pmatrix}
\gamma & 0 & \epsilon\lambda & 0 \\
0 & \gamma & 0 & \epsilon\lambda \\
0 & 0 & \lambda & 0 \\
0 & 0 & 0 & \lambda
\end{pmatrix}
\]

The elements in the diagonal represent the marginal productivity of resources. Spillover
effects with respect to local resources are: \(\frac{\partial P_i}{\partial r_i} = \epsilon\lambda\).

2.2. **Auditing authority and tax revenue.** All levels of government implement the
auditing policy in order to maximise the (net) tax revenue i.e. the total amount of tax
revenue collected minus the amount of resources devoted to the auditing activity. The tax

\(^4\)Since direct information is modelled as an iceberg cost on tax revenue, it does not influence monitoring.

\(^5\)Decreasing marginal productivity determines similar results.
revenue collected depends on the citizens’ decisions on the amount of tax base to report. Given the probability of being audited (P and p, for central and local tax auditing\(^6\)), the tax rate (T and t) and the penalty (S and s), each citizen optimally decides the tax base to report \((X^*(P, T, S)\) and \(x^*(p, t, s)\)). The net tax revenue equals the sums of optimal reports of citizens plus the penalty that they pay when monitored minus the resources relative to monitoring.

\[
G_i = (1 - D) \sum_\iota [X^*(P_i, T, S) + P_i \cdot (S + T)(Y - X^*_i(P_i, T, S))] - R_i
\]

\[
g_i = (1 - d) \sum_\iota [x^*(p_i, t_i, s) + p_i \cdot (s + t_i)(y - x^*_i(p_i, t_i, s))] - r_i
\]

where \(\iota\) indicates the citizens, \(Y\) and \(y\) are the tax base of the central and local tax, respectively. \(D\) is the iceberg cost payed by central authority, it equals 0 if the tax monitored is mobile, it equals \(C\) when the tax is immobile. \(d\) is the iceberg cost payed by local authority, it equals 0 if the tax is immobile, it equals \(c\) when it is mobile. As in Allingham and Sandmo (1972) both an increase in the penalty (S and s) and in the probability of auditing (P and p) increases the reported tax bases; the effect of tax rate is ambiguous since it depends on the attitude towards risk. It is possible to prove that when the elasticity of the report \((X\) and \(x\)) with respect to the probability of auditing is positive and less than one, the tax revenue increases less than proportionally with \(P\) (and \(p\)) \((\frac{\partial G}{\partial P} > 0; \frac{\partial^2 G}{\partial P^2} < 0\) and \(\frac{\partial g}{\partial p} > 0; \frac{\partial^2 g}{\partial p^2} < 0\)). Moreover, if \(S + T < 1\) (and \(s + t < 1\)) the tax revenue increases with penalty \((\frac{\partial G}{\partial S} > 0; \frac{\partial^2 g}{\partial S^2} > 0\).

This set up is consistent with tax revenue functions of the following type

\[
G(P_i) = A(P_i, \Omega)H(T, S)(1 - D) - R_i
\]

\[
g(p_i) = a(p_i, \omega)h(t_i, s)(1 - d) - r_i
\]

where the partial derivatives are\(^7:\)

\[
\frac{\partial A}{\partial P_i} = A_P > 0; A_{PP} < 0; A_\Omega > 0; a_P > 0; a_{pp} < 0; a_\omega > 0; H_S > 0; h_s > 0
\]

\(^6\)Capital letters refer to central, small to local.

\(^7\)The complete partial derivatives with respect to parameter are:

\[
g_\omega = a_\omega h(t_i, s)(1 - d) > 0
\]

\[
g_s = ah_s(1 - d) > 0
\]

\[
g_d = -aH(t, s) < 0
\]
Ω and ω represent the average aversion towards risk or the propensity to declare central and local tax base, respectively. We remark that \( A(P_i, \Omega)H(T, S) \) represents the central tax gross revenue, given the citizens’ optimal choice (the same for the local tax).

3. Equilibrium

The equilibrium of the game depends on the assignment of the enforcement function between the local and central tax authority. We consider three possible institutional scenarios,

(A) decentralised (separate): perfect correspondence between tax assignment and enforcement powers

(B) centralised auditing: in this case is the central authority which enforces both taxes

(Γ) localised: the local tax authority enforces both the local and the central tax in its region

The strategy space of the tax authority changes according to the institutional scenario. We are going to derive a SPE for each of the institutional scenario and then we compare the results in terms of net revenue.

3.1. Separate enforcement. The national government activity concerns the revenue on the national tax only. The two regions are concerned with enforcing their own taxes at the local level\(^8\). The central and the local tax authorities move simultaneously and independently.

The two local tax authorities behave in the same way,

\[
\max_{r_i} g_i(r_i) = a(p_i, \omega)h(t_i, s)(1 - d_A) - r_i \tag{3.1}
\]

\[
\frac{dg_i}{dr_i} = g_r = a_p[p_i(r_i), \omega] \lambda \cdot h(t, s)(1 - d_A) - 1 = 0 \tag{3.2}
\]

The local authorities choose \( r_i^* \) which solves equation 3.2. Note that \( r_i^* \) is a dominant strategy for local authorities.

\(^8\)A critical feature is that the exchange of information between different levels of government: if there is no communication then there is no information spillover, while with a perfect exchange of information the spillover is present but the equilibrium choice of the local tax authority does not interlise this externality.
Analogously, the national government maximises its net total revenue choosing the resource for auditing the national tax in both regions

$$\max_{R_1, R_2} G(R_1, R_2) = G_1 + G_2 = \sum_i A(P_i, \Omega)H(T, S)(1 - D_A) - R_i \quad (3.3)$$

$$\frac{\partial G}{\partial R_i} = G_R = A_P [P_i(R_i, r_i), \Omega] \gamma \cdot H(T, S)(1 - D_A) - 1 = 0 \quad (3.4)$$

The equilibrium of the game is the strategy profile \( \{r^e_i, R^e_i\} \) which satisfies the system of equation given by equations 3.2 and 3.4. Since the local authorities have a dominant strategy \( r^e_i = r^*_i \).

We conduct some exercise of comparative statics on the equilibrium strategies.

**Proposition 3.1.**

1. The amount of equilibrium resource devoted to local tax enforcement decreases with the information direct cost \( d \) and increases with \( \omega \), the sensitivity of tax payers to the probability of being audited, and with penalty \( s \).

2. There is a crowding-out of the central resources, an increase in the equilibrium level of local resources reduces the resources at the central level \( \frac{dr_e}{dr_e} = -\frac{\lambda_e}{\gamma} \epsilon < 0 \).

3. The amount of equilibrium resource devoted to central tax enforcement decreases with the information direct cost \( D \) for central authority, but increases with information direct cost for local one (because of crowding out). It increases with \( \Omega \) (the sensitivity of tax payers to the probability of being audited by central authority), and with penalty \( S \) (on central tax evasion). It decreases with sensitivity on local auditing \( \omega \) and with local penalty \( s \).

4. There exists a value \( R^*_e(\gamma) \) for which, the optimal resources stated at central level increases with its marginal productivity \( \gamma \) when the resource is lower this value \( (R_e < R^*_e(\gamma)) \), for higher values \( (R_e > R^*_e(\gamma)) \). Thus it the optimal resources devoted to auditing are low an increase in its productivity leads to increase the resources, when many resources are devoted an increase in productivity, reduce the necessity of new resources. The same at local level.

Proof. \( g_{rr} = \lambda^2 a_{pp} \cdot h(t, s)(1 - d) < 0, g_{rw} = \lambda a_{p\omega} h(t, s)(1 - d) > 0 \) with \( a_{p\omega} > 0, g_{rs} = \lambda a_{phs}(1 - d) > 0, g_{rd} = -\lambda a_{ph}(t, s) < 0 \); the same for central tax revenue \( \frac{dr_e}{ds} = -\frac{g_{rs}}{g_{rr}} > 0, \frac{dr_e}{ds} = -\frac{g_{rd}}{g_{rr}} < 0 \); the same for central tax revenue. Morevoer \( G_{Rr} = \gamma \lambda A_P h(T, S)(1 - D) \epsilon < 0 \), thus \( \frac{dR_e}{dr_e} = -\frac{G_{Re}}{G_{RR}} = -\frac{\Delta_e}{\gamma} \epsilon < 0, \frac{dR_e}{ds} = \frac{dR_e}{dr_e} \frac{dr_e}{ds} < 0. \) \( \square \)
Proposition 3.2. An increase in the equilibrium resources of the local authority \((r_e)\) increases the tax revenue at central government, because of information spillover \(\left(\frac{dG(R_e, r_e)}{d r_e} = \epsilon \frac{\lambda}{\gamma} > 0\right)\). The tax revenue in the equilibrium strategies \((R_e, r_e)\) increases with the sensitivity of taxpayers to auditing \((\Omega \text{ and } \omega)\) and with penalty \((S \text{ and } s)\), it decreases with direct information costs \((D \text{ and } d)\). Tax revenues increases with marginal productivity and with marginal spillover.

Proof. The derivative of tax revenue in equilibrium strategies with respect to parameters equals the partial derivative of tax revenues. Moreover

\[
\frac{dG(R_e, r_e)}{dr_e} = G_P \cdot \epsilon = \lambda A_P H(T, S)(1 - D)\epsilon = \epsilon \frac{\lambda}{\gamma} > 0
\]

□

3.2. Centralised and localised enforcement. Now we assume that the central tax authority has the power to enforce both central and local taxes.

\[
\max_{\{R_i, r_i\}} W = G_1 + G_2 + g_1 + g_2 = \\
= \sum_i [A(P_i, \Omega)H(T, S)(1 - D_B) - R_i] + \\
+ \sum_i [a(P_i, \omega)h(t_i, s)(1 - D_B) - r_i]
\]

\[\text{(3.5)}\]

\[
\frac{dW}{dR_i} = A_P[P_i(R_i, r_i), \Omega] \gamma \cdot H(T, S)(1 - D_B) - 1 = 0
\]

\[\text{(3.6)}\]

\[
\frac{dW}{dr_i} = g_r + G_r = \\
= a_p[p_i(r_i), \omega] \lambda \cdot h(t, s) (1 - D_B) - 1 + \\
+ A_P[P_i(R_i, r_i), \Omega] \cdot H(T, S)(1 - D_B) \lambda \epsilon = 0
\]

\[\text{(3.7)}\]

The condition for the equilibrium can be written as

\[
A_P[P_i(R_i, r_i), \Omega] \gamma \cdot H(T, S)(1 - D_B) = 1
\]

\[\text{(3.8)}\]

\[
a_p[p_i(r_i), \omega] \lambda \cdot h(t, s)(1 - D_B) + \frac{\lambda \epsilon}{\gamma} = 1
\]

\[\text{(3.9)}\]

Note that condition 3.9 does not depend on \(R_i\), therefore the equilibrium level of local resources is the solution of 3.9.
Proposition 3.3. If the costs associated with the direct information flow does not change between the two scenarios, the equilibrium local resources are greater or equal than the equilibrium level in case of a separate enforcement, \( r^B_i \geq r^A_i \).

Corollary 3.4. Given propositions 3.3 and 3.1, for the crowding-out of central resources, the equilibrium level of central resources is lower or equal than in the separate scenario, \( R^B_i \leq R^A_i \).

Proposition 3.5. The tax revenue in the equilibrium strategies \((R_e, r_e)\) replays the same results of 3.1. Moreover the equilibrium strategy \(r_e\) for local authority increases with marginal spillover \((\epsilon)\) and decreases with marginal productivity at central level.

In a localised scenario the local authority maximised

\[
\max_{\{R_i, r_i\}} W_i = G_i + g_i
\]  

(3.10)

In this case if the parameter coincides, the solution of equation 3.10 coincides with the equation 3.5 solution. Note that when information spillover equals 0, \(ceteris paribus\), the equilibrium strategies in each scenario coincide and the centralised revenue coincides with the sum of revenues in separate and in localised ones.

4. Optimal Tax (and enforcement) Assignment

In this section we use our model to analyse the trade-off between direct information and spillovers when taxes are assigned according to mobility. The tax assignment theory suggest taxes should be assigned to central and local level in order to minimise inefficiencies due to mobility. Therefore mobile taxes are assigned at the central level, while immobile taxes are assigned to the local level.

This implies that in the separate scenario direct information costs are equal to zero \((D_A = d_A = 0)\). In the centralised scenario the enforcement of the local tax has a direct information cost \((D_B = 0, d_b = c)\), while in the localised scenario is the enforcement of the central tax that bears a direct information cost \((D_\Gamma = C, d_\Gamma = 0)\).

The comparison between the tax revenue in case of separate and centralised scenarios depends on the level of \(c\), the direct cost the central authorities incurs because of lack of information on the local tax base, and on \(\epsilon\) the level of spillovers between the local and the central tax base. Given the level of the spillover, the higher is \(c\) the smaller is the revenue in the centralised scenario with respect to the separate case. On the other hand, given \(c\)
the higher is $\epsilon$ the greater is the revenue of the centralised scenario with respect to the separate case. Therefore, the optimal enforcement assignment suggest to centralise when spillovers are high and direct information costs are lower.

We remark that when the spillover does not exists, *ceteris paribus* the centralised tax revenue equals the sum of separate one. Thus we can totally differentiate the centralised tax revenue in equilibrium strategies.

$$dG(R_1, R_2, r_1, r_2) = (G_1D + G_2D)dD + (g_1d + g_2d)dd + (G_1 + G_2)d\epsilon$$

since it is easy to demonstrate that with the assumption on the auditing technology $G_i = \frac{\lambda^2}{\gamma} \epsilon + \frac{\lambda}{\gamma} r_i^c$, let we start from initial condition $D = 0, d = 0, \epsilon = 0$ which coincides with separate scenario. If we assume that the parameters changes from separate to centralised scenarios are small enough, we can argue that the centralised scenario produces a greater tax revenue if

$$- \sum_i (a(p_i(r_i^c))h(t_i, s)) c + \frac{\lambda}{\gamma} \sum_i r_i^c \epsilon > 0$$

thus

$$\frac{\lambda}{\gamma} \epsilon > \frac{\sum_i (a(p_i(r_i^c))h(t_i, s))}{\sum_i r_i^c} c$$

Note that the left hand side represent the marginal tax revenue due to spillover, while the right hand side represent the iceberg cost for a unit of resource employed in auditing at local level.

**Proposition 4.1.** *Ceteris paribus, the higher is the direct cost of information at central level $c$ for monitoring an immobile tax base, the higher is the average gross local revenue at local level with respect to resource employed $\frac{\sum_i (a(p_i(r_i^c))h(t_i, s))}{\sum_i r_i^c}$, and the higher is the national marginal productivity of auditing with respect to local one, the higher should be the information spillover necessary to centralise the tax enforcement.*

The comparison among separate, localised and centralised follows the same idea. In particular a localised scenario is preferred to separate one if

$$\frac{\lambda}{\gamma} \epsilon > \sum_i (A(P_i(r_i^c))H(T, S)) \frac{\sum_i R_i^c}{\sum_i r_i^c} C$$

while a centralised tax auditing is preferred to localised one if

$$\sum_i (a(p_i(r_i^c))h(t_i, s)) c < \sum_i (A(P_i(r_i^c))H(T, S)) C$$
5. Conclusion
References


