Società italiana di economia pubblica

XXII CONFERENZA

NUOVE FRONTIERE DELL'INTERVENTO PUBBLICO
IN UN MONDO DI INTERDIPENDENZA

Pavia, Università, 20-21 settembre 2010

COMMUNICATING VESSELS

FEDERICO REVELLI

Communicating vessels

Federico Revelli*

May, 2010

Abstract

This paper models the local tax mix determination process in the presence of state-wide tax limitations - the decentralized government finance archetype - and shows how excess sensitivity of local public spending to grants (the so-called "flypaper effect") arises in the endogenously generated constrained tax mix. By means of a panel data switching regression approach that allows for fixed effects and endogenous selection, the paper exploits the clustering of Italian Provinces at the corners produced by upper and lower limitations on provincial tax revenue sources, and provides evidence of considerable "cap-generated" excess sensitivity.

JEL classification: C23; C25; H72.
Key words: flypaper effect; excess sensitivity; tax mix; switching regression; endogenous selection.

*Department of Economics, University of Torino, Via Po 53, 10124 Torino (Italy); e-mail: federico.revelli@unito.it.

1 Introduction

The overall size as well as the tax revenue bundle of the local public sector in multi-tiered structures of government are the outcomes of the decentralized decision-making process subject to the fiscal rules set by central (state) governments. As documented by Anderson [2] and Wolman et al. [34] for the US, and by Joumard and Kongsrud [21] and Sutherland et al. [33] for the OECD countries, top-down tax and expenditure limitations (TELs) are frequently so tight and pervasive as to jeopardize the very principle of local fiscal autonomy.[1]

This paper aims at investigating how (exogenously set) revenue raising limitation rules shape local governments' budget constraints, focusing in particular on the kinks that are typically generated by tax floors and caps, and at evaluating their effects on the determination of the local tax mix and on the response of local public expenditures to grants.

As far as the latter issue is concerned, a vast literature (most recently reviewed by Inman [19]) has investigated and sought to explain the anomalously high response of local spending to grants relative to the response to private income - the so-called flypaper effect by which money from central government "sticks where it hits."[2]

Two broad kinds of explanations of the flypaper effect have been offered in the literature (Hines and Thaler [17]). The first has to do with a variety of specification and estimation errors that applied researchers would have kept making for decades. Those errors range from mistakenly treating matching grants as if they were lump-sum to the omission of important variables - such as unobserved population characteristics or spatial lags of other governments' policies - that are simultaneously correlated with grants and local public expenditures. The second explanation relies on the argument that the political representation

[1] According to Nechyba [27], though, state command on local fiscal choices (in terms of income tax-funded grants and state-imposed caps on local property tax rates) arises in equilibrium as an optimal outside enforcement when a collusive agreement to simultaneously introduce local income taxes is not self-enforcing.

[2] According to Inman [19], over 3,500 research papers exist documenting and seeking to explain the flypaper effect. Payne [28] offers an insightful wide-ranging review of the more recent research into the mirror phenomenon of crowd-out.

process is substantially richer than the one postulated by the standard neoclassical model: asymmetric information, loss aversion, voters' misperceptions and fiscal illusion, separate mental accounting, special interest groups, and citizens' inability to write complete contracts with their elected officials would be responsible for the lack of fungibility between public and private uses of money, and would cause the observed large flypaper effect.

This paper models for the first time the local tax mix determination process in the presence of state-wide tax limitations - the decentralized government finance archetype - and shows how excess sensitivity of local public spending to grants arises in the endogenously generated constrained tax mix. In particular, the paper shows that the effect of private community income on public spending should be expected to be tiny or nil in the presence of binding limitations on all local tax revenue sources, while grants should be predicted to have a large - actually, a one-for-one - impact on local expenditures. Interestingly, a binding cap on just one of the available own revenue sources is enough to generate some form of flypaper effect, in the sense of an excess sensitivity of local spending to grants, and the above result holds when either upper or lower tax limitations are in place. Finally, since excess sensitivity of local public spending should be predicted to arise and generally tends to manifest itself both when grants increase and when they decrease, the "flypaper effect" label seems an inappropriate or even misleading one.

While the existing literature seems to have almost universally overlooked the potential impact of tax and expenditure limitation systems on the sensitivity of local public spending to exogenous variations in grants, two recent papers have brought the fiscal limitations issue into the empirical investigation of the flypaper effect. Lutz [25] conjectures that previous evidence of a flypaper effect might have arisen from state constraints preventing local governments from selecting their preferred bundle of public goods, and provides evidence of equivalence between grants and income from a school finance reform in New Hampshire - "one of only five states with no state-imposed limitations on the taxing or spending power of local governments" (Lutz [25], p. 317). Brooks and Phillips [5] use data on the US Community Development Block Grant (CDBG) program and argue that restrictive fiscal institutions - in terms of state TELs - may systematically force city governments to underprovide local public goods and therefore increase the stimulative effect of federal grants on city spending. However, since they neither observe the municipal tax bundle nor whether a revenue raising constraint is binding in any given city, they have to rely on a state-level index of fiscal constraints and ignore altogether both the municipal choice as to own revenue source diversification and, more importantly, the issue of endogenous selection of a city government into the fiscally constrained status.[3]

This paper concludes with an empirical application to Italian Provincial governments' panel data. An attractive feature of Italian Provinces is that their own tax revenue sources (a tax on vehicle registrations, a tax on electricity consumption for business uses, and a waste management surcharge) are subject to strict and frequently binding upper as well as lower tax rate limitations. The empirical analysis exploits the clustering of provincial authorities at the corners produced by those tax limitation rules, and, based on a panel data switching regression approach that allows for fixed effects and endogenous selection into the constrained regime, offers evidence of "cap-generated" excess sensitivity of local public spending to grants.

The paper is organized as follows. Section **2** sets up a simple model for the analysis of the local tax mix in the presence of right and left constrained tax instruments. Sections **3** develops the model's empirical implications and outlines the econometric strategy, while section **4** tackles the issue of endogenous selection. Finally, section **5** reports and discusses the estimation results on the Italian Provinces' panel data, and section **6** concludes.

[3]Interestingly, Brooks and Phillips [5] find excess sensitivity of spending to grants in a period of dramatic retrenchement, while they comprehensibly find limited evidence of an effect of overall state-level tax limitations on municipal governments' response to the collapse in CDGB grants.

2 Communicating vessels

Consider the two vessels in figure 1. Say that vessel v_{pn} represents consumption of private goods out of community n private income i_n ($n = 1, ..., N$), and vessel v_{gn} represents consumption of local public services. The structure depicted in (1.a) amounts to a perfect tax centralization arrangement, where expenditures on local public services are entirely funded by central government grants g_n, and nothing ensures that the allocation of resources to private and public uses reflects the preferences of the local community or that the marginal benefit from private consumption equals the marginal benefit from public consumption.

In the central picture (1.b), the two vessels are allowed to be communicating via local tax revenues. In order for the local public goods to be provided optimally, and given that the marginal rate of transformation between private and public goods is constant and equal to one, the marginal utility in the two vessels has to be equalized. Just like communicating vessels, where the force of gravity requires hydrostatic pressure to be balanced out in the two vessels regardless of their relative sizes, the welfare optimization forces make resources flow from v_{pn} to v_{gn} at the tax rate $\tau_n = \frac{t_n}{i_n}$. Once the equilibrium level is attained in the two vessels, whether additional resources are poured into v_{pn} or into v_{gn}, the same allocation of private and public consumption will result by the law of communicating vessels.

In the lower picture (1.c), local jurisdiction n is subject to a tax rate cap equal to $h = \frac{t_{hn}}{i_n}$, with the cap binding if $t_{hn} < t_n$. The Samuelson condition for optimal public good provision will not be satisfied if the tax cap is binding, meaning that more resources ought to flow from v_{pn} to v_{gn} in order to equate the "pressure" in the two vessels. An additional unit of private income will raise the consumption level at rate $1-h$ in v_{pn}, and at the rate h in v_{gn}. If additional grants are poured into (pumped out of) v_{gn}, the level will rise (fall) in v_{gn} only. The flypaper effect, so to say.

2.1 The one-tax case

According to the standard local public finance framework that is conventionally employed to analyze the decentralized tax-expenditure decision process (Inman [19]), the welfare of jurisdiction n ($n = 1, ..., N$) can be expressed as a quasi-concave function exhibiting decreasing marginal benefits from local public expenditures as well as increasing marginal cost of raising own revenues:

$$W_n = V(z_n; \mathbf{x}_n) - C(\tau_n b_n; i_n) \tag{1}$$

In general terms, z_n equals local public spending, local tax revenues are raised by setting a flat tax rate (τ_n) on a local tax base (b_n), $\mathbf{x}_n$ is a vector of community characteristics reflecting preferences for local public services, and i_n represents some meaningful measure of community income. Assume further that local authorities abide to a balanced budget rule:

$$z_n = g_n + \tau_n b_n \tag{2}$$

where g_n equals lump-sum grants from central government.[4] Maximization of (1) subject to (2) leads jurisdiction n to select the optimal tax rate-spending pair (τ_n^*, z_n^*) as a function of the assumed exogenous variables g_n, i_n, and b_n.[5] An exogenous increase (decrease) in grants would provoke an increase (decrease) in spending by a certain proportion of the grant itself, depending on the shape of the welfare function. Actually, a change in i_n by the same amount as the change in g_n should have an identical effect on z_n: when this does not happen, and in particular if a change in g_n turns out in practice to provoke a much larger reaction in z_n than a change in i_n does, a flypaper effect is said to exist (Hines and Thaler [17]).

Consider now the introduction of a tax rate cap such that $\tau_n \leq h_n$, and assume that $\tau_n^* > h_n$, meaning that local government n is at a corner solution. The constrained optimization problem is depicted in figure 2. Say that, in the

[4] It is ususally convenient to interpret all monetary variables in (2) as measured in per capita terms, thus implying that publicly provided services entering the welfare function (1) are private (rival) in nature.

[5] Tax base endogeneity is allowed for in the next section.

absence of the cap and with grant g_n and income i_n, authority n is at point 0. After capping, the constrained tax rate-spending pair is $(h_n, g_n + h_n b_n)$ at point 1 (a convex kink of the budget constraint). For given grants, an increase in i_n to $i_n^{\#}$ moves government n to point 2, with a zero impact on public spending (or little impact if i_n and b_n are positively correlated). On the other hand, figure 3 shows that, given the right censoring in local tax revenues, an increase in grants from g_n to $g_n^{\#}$ leads to a one-for-one increase in z_n (point 3). In the above circumstances, the flypaper effect is the result of capping. In fact, since excess sensitivity of local public spending should be predicted to arise (as figure 3 suggests) and generally tends to manifest itself both when grants increase and when they decrease (Stine [32], Hines and Thaler [17], Gamkhar and Oates [11]), the "flypaper effect" label seems an inappropriate or even misleading one.[6]

2.2 The multiple-tax case

Consider now the case of local government n relying on $M \geq 2$ distinct own tax revenue sources as well as on central government lump-sum grants. Dropping the n subscript to save on notation, and denoting by τ^m the flat rate set on tax base m (b^m), the budget constraint is:

$$z = g + \boldsymbol{\tau}'\mathbf{b} = g + \sum_{m=1}^{M} \tau^m b^m \tag{3}$$

Allowing for heterogeneous marginal costs of raising revenues from the potentially endogenous M tax sources (Hettich and Winer [15]), the welfare function can be expressed as:

$$W(\boldsymbol{\tau}; \mathbf{q}) = V(g + \boldsymbol{\tau}'\mathbf{b}; \mathbf{x}) - C(\tau^1 b^1, ..., \tau^M b^M; i) \tag{4}$$

where: $\mathbf{q}' = [g \; i \; \mathbf{x}']$. Letting $\epsilon^m \equiv \frac{\tau^m}{b^m}\frac{\partial b^m}{\partial \tau^m}$ be the own tax rate elasticity of tax base m, the first order conditions for maximization of (4) require equalization of

[6] Interestingly, the fiscal limitation approach can also explain the somewhat sparse evidence of "excess smoothness" of local public spending in response to grant cuts - what is improperly termed the asymmetric flypaper effect (in fact, the flypaper effect metaphor seems to be asymmetric by definition). In the presence of centrally mandated services, local public spending should be expected to exhibit little or no sensitivity to grants (excess smoothness) in periods of retrenchment.

the marginal contribution to welfare of spending an additional unit of own tax revenue on local public services $\left(v(g+\boldsymbol{\tau}'\mathbf{b};\mathbf{x}) = \frac{\partial V}{\partial \tau^m}\frac{1}{b^m(1+\epsilon^m)}\right)$ to the marginal costs of raising revenues across all tax bases $\left(c^m = \frac{\partial C}{\partial \tau^m}\frac{1}{b^m(1+\epsilon^m)},\ m=1,...,M\right)$, resulting in a vector of optimal tax rates and expenditure level:

$$\boldsymbol{\tau}^*(\mathbf{q})' = \left[\tau^{1*}(\mathbf{q}) \ ... \ \tau^{M*}(\mathbf{q})\right] \tag{5}$$

$$z^*(\mathbf{q}) = g + \sum_{m=1}^{M} \tau^{m*}(\mathbf{q})b^m(\mathbf{q}) \tag{6}$$

Figure 4 shows a graphical representation of the optimal tax mix determination problem under the simplifying assumption that the marginal cost of raising revenues from tax m is independent of τ^j for $j \neq m$ (Hettich and Winer [16]). We assume here that $M=2$, and let the marginal cost and benefit functions be linear for graphical convenience.

Given grants g, the optimal tax rates and spending vector is $\left[\tau^{1*}\ \tau^{2*}\ z^*\right]$. Similarly to the one-tax case, an exogenous increase in grants, say from g to $g^{\#}$, is expected to bring about an increase in spending as well as a decrease in reliance on both own tax sources $\left[\tau^{1\#}\ \tau^{2\#}\ z^{\#}\right]$, with the direction and intensity of the tax mix adjustment depending on the slope of the marginal cost functions.[7]

Consider now the consequences of central government imposing the following tax rate limitations ($m = 1, ..., M$):

$$0 \leq l^m \leq \tau^m \leq h^m \tag{7}$$

Given the welfare function (4) and the constraints (7), and letting $\boldsymbol{\lambda}' = \left[\lambda^{h1} \ ... \ \lambda^{hM} \ \lambda^{l1} \ ... \ \lambda^{lM}\right] \geq 0$ be the vector of Lagrange multipliers, we can write the Lagrangian function as:

$$L(\boldsymbol{\tau},\boldsymbol{\lambda}) = W(\boldsymbol{\tau};\mathbf{q}) + \sum_{m=1}^{M} \lambda^{hm}(h^m - \tau^m) + \sum_{m=1}^{M} \lambda^{lm}(-l^m + \tau^m) \tag{8}$$

[7] When grants go from g to $g^{\#}$, the v function shifts left by $(g^{\#}-g)$. It is easy to show that an identical increase in community's private income shifts the marginal cost function c to the right by $(g^{\#}-g)$, and should therefore be expected to have the same impact on local public spending as the grant increase.

The necessary Kuhn-Tucker conditions for the tax rate vector to be an optimum are:[8]

$$\nabla_{\boldsymbol{\tau}} L(\boldsymbol{\tau}, \boldsymbol{\lambda}) = 0 \qquad \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{\tau}, \boldsymbol{\lambda}) \geq 0 \qquad \boldsymbol{\lambda} \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{\tau}, \boldsymbol{\lambda}) = 0 \tag{9}$$

or, for $m = 1, ..., M$:

$$\frac{\partial L(\boldsymbol{\tau}, \boldsymbol{\lambda})}{\partial \tau^m} = \frac{\partial W(\boldsymbol{\tau}; \mathbf{q})}{\partial \tau^m} - \lambda^{hm} + \lambda^{lm} = 0 \tag{10}$$

$$\frac{\partial L(\boldsymbol{\tau}, \boldsymbol{\lambda})}{\partial \lambda^{hm}} = h^m - \tau^m \geq 0 \qquad \lambda^{hm}(h^m - \tau^m) = 0 \tag{11}$$

$$\frac{\partial L(\boldsymbol{\tau}, \boldsymbol{\lambda})}{\partial \lambda^{lm}} = -l^m + \tau^m \geq 0 \qquad \lambda^{lm}(-l^m + \tau^m) = 0 \tag{12}$$

The Kuhn-Tucker conditions (10)-(11)-(12) show that when $\tau^m = h^m$, the gradient of the welfare function is positive $\left(\frac{\partial W(\boldsymbol{\tau};\mathbf{q})}{\partial \tau^m} = \lambda^{hm}\right)$, meaning that the unconstrained optimal tax rate lies to the right of h^m. Similarly, if $\tau^m = l^m$, the gradient of the welfare function is negative $\left(\frac{\partial W(\boldsymbol{\tau};\mathbf{q})}{\partial \tau^m} = -\lambda^{lm}\right)$, so that the unconstrained optimal tax rate lies to the left of l^m.

Figures 5 and 6 offer a graphical representation of the constrained optimization process and illustrate how corner solutions arise.[9] In particular, figure 5 focuses on the following two cases:

- Constrained tax mix RR: both tax rates are right censored, meaning that an authority's optimal tax rates both lie to the right of the upper limits ($\tau^{1*} > h^1; \tau^{2*} > h^2;\ \lambda^{h1}, \lambda^{h2} > 0; \lambda^{l1} = \lambda^{l2} = 0$).

- Constrained tax mix LL: both tax rates are left censored ($\tau^{1*} < l^1; \tau^{2*} < l^2;\ \lambda^{l1}, \lambda^{l2} > 0; \lambda^{h1} = \lambda^{h2} = 0$).

On the other hand, figure 6 illustrates the following three cases:

- Constrained tax mix UR: one of the tax rates is uncensored and the other is right censored ($l^1 < \tau^{1*} < h^1; \tau^{2*} > h^2;\ \lambda^{h2} > 0; \lambda^{h1} = \lambda^{l1} = \lambda^{l2} = 0$).

[8] Provided that (4) is concave, the Kuhn-Tucker conditions are sufficient.

[9] Interestingly, all of the tax mix corners depicted in figures 5 and 6 turn out to be relevant in the empirical application in section **5**.

- Constrained tax mix LU: one of the tax rates is left censored and the other is uncensored ($\tau^{1*} < l^1; l^2 < \tau^{2*} < h^2;\ \lambda^{l1} > 0; \lambda^{h1} = \lambda^{h2} = \lambda^{l2} = 0$).
- Constrained tax mix LR: one of the tax rates is left censored and the other is right censored ($\tau^{1*} < l^1; \tau^{2*} > h^2;\ \lambda^{l1}, \lambda^{h2} > 0; \lambda^{h1} = \lambda^{l2} = 0$).

Let us ask again what effect an exogenous change in grants would have on local public spending. When both revenues sources are constrained - cases RR and LL in figure 5, and LR in figure 6 - any change in grants necessarily translates into an identical change in spending levels, with no variation in τ^1 and τ^2. Local public expenditure displays an excess sensitivity to grants.

When only one of the two fiscal instruments is at a corner - cases UR and LU in figure 6 - the change in spending is typically smaller than the change in grants. However, the binding constraint on one of the available local tax revenues makes the total marginal cost of raising revenues steeper than it would be in the absence of constraints. Since local authorities can rely on fewer tax instruments, the cost of raising revenues is more sensitive to changes in the budget requirement: as a result of the increase in grants, the impact on public expenditures is larger than would occur if both tax rates could freely adjust. Of course, excess sensitivity of spending will be observed both when grants increase and when they decrease.

Finally, as long as changes in exogenous variables - including grants - do not provoke a segment jump, lower-capped and upper-capped authorities should exhibit the same reaction of spending to grants. The above results can be summarized as follows.

- In a fully constrained tax mix, local public spending exhibits little or no sensitivity to private income changes; on the other hand, local public expenditures respond to changes in grants on a one-for-one basis.
- As long as changes in exogenous sources of revenue do not provoke a segment jump, upper capped authorities ($\tau^m = h^m, \forall m$), lower capped authorities ($\tau^m = l^m, \forall m$), and lower-upper capped authorities ($\tau^m =$

$l^m, m = 1, ..., \overline{m}; \tau^m = h^m, m = \overline{m}+1, ..., M$) exhibit the same sensitivity of public spending to grants.

- In a partially constrained tax mix where $\widetilde{m}$ tax limits are binding ($1 \leq \widetilde{m} < M$) and $M - \widetilde{m}$ are not: a) the total marginal cost of raising revenues is flatter and the sensitivity of local public spending to grants is smaller than it is in a fully constrained tax mix; b) the total marginal cost of raising revenues is steeper and the sensitivity of local public spending to grants is larger than it is in the absence of binding tax limitations.

3 Empirical implications and econometric approach

The theoretical model sketched in section **2** above prompts the estimation of the sensitivity of local public expenditures to changes in exogenous revenue sources, while allowing for heterogeneous responses depending on the degree to which local governments face "financing constraints." Put it this way, the empirical investigation framework of the flypaper effect phenomenon bears a striking similarity with two well developed lines of research.

The first concerns the inquiry into the role of financing and liquidity constraints in explaining the elasticity of investment to cash-flow in Q models of the firm (Bond and Meghir [3], Fazzari et al., [10], Hu and Schiantarelli [18], Kaplan and Zingales [22], Cummins et al. [8]).[10] The second relates to the borrowing constraint interpretation of the excess sensitivity of private consumption to disposable income in permanent income/life cycle frameworks (Runkle [30], Zeldes [37], Jappelli et al. [20]).[11]

In the empirical investment and consumption literatures, the conventional

[10] In their flypaper effect review, Hines and Thaler [17] mentioned the liquidity constraint explanation of flypaper-like effects in the private sector. However, they did not consider the possibility that local tax and expenditure limitations might be the root cause of the flypaper effect.

[11] Borge and Tovmo [4] test whether liquidity constraints imposed by balanced-budget rules affect the intertemporal spending behavior of Norwegian local governments, and find that departures from rational forward-looking public consumption smoothing can in part be explained by financing constraints.

approach consists in splitting the sample according to an *a priori* index of financing/liquidity constraint (typically related to the dividend payout or liquid assets to capital stock ratio for firms, and to the asset-income ratio for consumers), and compare the "switching regression" estimates of the sensitivity of investment (consumption) to cash flow (income) for the constrained and unconstrained subsamples (Fazzari et al. [10], Kaplan and Zingales [22], and Runkle [30]).

Similarly, in order to test whether the local public spending response to changes in exogenous sources of revenue is affected by the tax limitation regime a local government is subject to, a time-invariant selection criterion can be employed and local authorities assigned to either of two subsamples based on whether they are consistently constrained (or not constrained) during the whole period of observation ($t = 1, ..., T$):

$$z_{nt} = \mathbf{q}'_{nt}\boldsymbol{\beta}^1 + \zeta_n^1 + \eta_{nt}^1 \quad \text{if} \quad K_n = 1 \tag{13}$$

$$z_{nt} = \mathbf{q}'_{nt}\boldsymbol{\beta}^0 + \zeta_n^0 + \eta_{nt}^0 \quad \text{if} \quad K_n = 0 \tag{14}$$

where: $\mathbf{q}'_{nt} = [g_{nt}\ i_{nt}\ \mathbf{x}'_{nt}]$, $\boldsymbol{\beta}^{1\prime} = \left[\beta_g^1\ \beta_i^1\ \boldsymbol{\beta}_x^{1\prime}\right]$, $\boldsymbol{\beta}^{0\prime} = \left[\beta_g^0\ \beta_i^0\ \boldsymbol{\beta}_x^{0\prime}\right]$, and ζ_n^1 and ζ_n^0 are fixed jurisdiction effects that might be correlated with $\mathbf{q}_{nt}$. The switching indicator K_n is defined as:

$$K_n = \begin{cases} 1 & \text{if} \quad \tau_{nt}^m \in \{l^m, h^m\}, \forall m, t \\ 0 & \quad\quad \nexists t : \tau_{nt}^m \in \{l^m, h^m\}, \forall m \end{cases} \tag{15}$$

According to the sample separation criterion (15), parameter β_g^0 in equation (14) captures the response of spending to grants by authorities that are not structurally constrained (i.e., authorities that never have all constraints binding in any of the T observation years), while parameter β_g^1 in equation (13) measures the response by authorities that are structurally capped, in the sense that tax limitations are binding for all revenue sources and in all years.

One potential problem with the above approach is sample selection bias, in the sense that the splitting variable K_n might be correlated with spending z_{nt}. However, since selection effects can only occur through correlation between K_n and the time-invariant authority-specific effects ζ_n^1 and ζ_n^0, any selection bias is cancelled by differencing them away in equations (13)-(14). Consequently, a linear panel data fixed effects estimator can be applied to the two subsamples (Charlier et al. [7]).

A disadvantage of the separation rule (15), though, consists in the fact that it implies "freezing" the sample and renouncing to using information on governments that switch from one regime to the other over the period of observation (Hu and Schiantarelli [18]).[12]

An alternative empirical approach - based, among the others, on Bond and Meghir [3], Jappelli et al. [20], Zeldes [37] and Cummins et al. [8] - consists in allowing for a time-varying constraint status as in (16) below:

$$K_{nt} = \begin{cases} 1 & \text{if} \quad \tau_{nt}^m \in \{l^m, h^m\}, \forall m \\ 0 & \quad\quad \exists m : \tau_{nt}^m \notin \{l^m, h^m\} \end{cases} \tag{16}$$

giving rise to the switching regression model:

$$z_{nt} = \mathbf{q}'_{nt}\boldsymbol{\beta}^1 + \zeta_n^1 + \eta_{nt}^1 \quad \text{if} \quad K_{nt} = 1 \tag{17}$$

$$z_{nt} = \mathbf{q}'_{nt}\boldsymbol{\beta}^0 + \zeta_n^0 + \eta_{nt}^0 \quad \text{if} \quad K_{nt} = 0 \tag{18}$$

Clearly, though, whether an authority is at a tax corner solution might in principle be determined endogenously. This would occur in the presence of unobserved shocks to expenditures pushing local authorities towards the corners.

Given the sample separation rule (16), the endogenous selection problem is somewhat mitigated here by the fact that the constrained regime for which $K_{nt} = 1$ includes high spenders hitting the upper tax bounds ($\tau_{nt}^m = h^m, \forall m$: constrained tax mix RR), low spenders hitting the lower tax bounds ($\tau_{nt}^m = l^m, \forall m$: constrained tax mix LL) and authorities hitting lower and upper bounds on different tax rates at the same time ($m = 1, ..., \overline{m} : \tau_{nt}^m = l^m$; $m = \overline{m} + 1, ..., M : \tau_{nt}^m = h^m$: constrained tax mix LR). As a result, it is unclear *a priori* whether and how would K_{nt} be correlated with unobservable shocks to z_{nt}. However, the endogenous selection issue is developed further in the next section.

[12] This implies in our case that we could only use information on about half of the Italian Provinces.

4 Endogenous selection

In order to explore the endogenous selection issue in greater depth, the selection process (16) needs to be explicitly given a stochastic structure. We do so by following the Wooldridge [35] two-stage procedure for fixed effects panel data, with a selection equation being consistently estimated in the first stage, and the main spending equation being estimated in the second stage after correcting for selection bias.[13] The Wooldridge [35] approach relies on estimation in levels, and has the great advantage relative to econometric approaches based on pairwise differencing on units for which $K_{nt} = K_{ns} = 1, t \neq s$ (Kyriazidou [24]) of not requiring any exclusion restrictions in the main equation. In fact, since the constrained optimization model in section **2** predicts the level of spending z_{nt} and the capping regime K_{nt} to be determined simultaneously as a function of the vector of exogenous variables $\mathbf{q}_{nt}$, exclusion restrictions in the z_{nt} equation would be logically inconsistent.[14]

Let us focus on selection into the fully constrained regime $K_{nt} = 1$. The fixed effects estimator of equation (17) on the unbalanced panel, with authority n being observed for $T_n = \sum_{t=1}^{T} K_{nt}$ years in the $K_{nt} = 1$ regime, is (Wooldridge [35]):

$$\widehat{\boldsymbol{\beta}}^1 = \left(\sum_{n=1}^{N} \sum_{t=1}^{T} K_{nt} \widetilde{\mathbf{q}}_{nt} \widetilde{\mathbf{q}}'_{nt} \right)^{-1} \left(\sum_{n=1}^{N} \sum_{t=1}^{T} K_{nt} \widetilde{\mathbf{q}}_{nt} \widetilde{z}_{nt} \right) \tag{19}$$

$$\widetilde{\mathbf{q}}_{nt} = \mathbf{q}_{nt} - \frac{1}{T_n} \sum_{s=1}^{T} K_{ns} \mathbf{q}_{ns} \tag{20}$$

$$\widetilde{z}_{nt} = z_{nt} - \frac{1}{T_n} \sum_{s=1}^{T} K_{ns} z_{ns} \tag{21}$$

A sufficient condition for (19) on the selected subsample $K_{nt} = 1$ to be consistent is that the selection process be strictly exogenous conditional on ζ_n^1

[13] Dustmann and Rochina-Barrachina [9] give a simple illustration and an application of the Wooldridge [35] and related procedures.

[14] In addition, the Kyriazidou [24] pairwise differencing estimator has the data-shrinking shortcoming of using only those observations for which the linear index in the selection equation is approximately equal in periods t and s: $\mathbf{q}'_{nt} \widehat{\boldsymbol{\delta}} \cong \mathbf{q}'_{ns} \widehat{\boldsymbol{\delta}}$, $t \neq s$ (see equation (27) below).

and $\mathbf{q}_n$:

$$E(\eta^1_{nt} \left| \zeta^1_n, \mathbf{q}_n, \mathbf{K}_n \right.) = E(\eta^1_{nt} \left| \zeta^1_n, \mathbf{q}_{n1}, ..., \mathbf{q}_{nT}, K_{n1}, ..., K_{nT} \right.) = 0 \qquad (22)$$

As clearly shown by equation (16), K_{nt} depends in a structural way on M distinct tax rate realizations, where the process underlying the determination of each tax rate τ^m_{nt} $(m = 1, ..., M)$ can be expressed in a latent variable form as:

$$\tau^m_{nt} = \begin{cases} h^m & \tau^{m*}_{nt} \geq h^m \\ \tau^{m*}_{nt} \quad \text{if} & l^m < \tau^{m*}_{nt} < h^m \\ l^m & \tau^{m*}_{nt} \leq l^m \end{cases} \qquad (23)$$

τ^{m*}_{nt} in (23) is the partially unobserved "optimal" tax rate of authority n in year t, whose non-stochastic component descends directly from the model sketched in section **2**:

$$\tau^{m*}_{nt} = \mathbf{q}'_{nt}\boldsymbol{\gamma}^m + \xi^m_n + a^m_{nt} \qquad (24)$$

where: $\boldsymbol{\gamma}^{m\prime} = \left[\gamma^m_g \ \gamma^m_i \ \boldsymbol{\gamma}^{m\prime}_x\right]$, and ξ^m_n is a Province and revenue-source specific time-invariant effect that might be correlated with $\mathbf{q}_{nt}$. In fact, correlation between fixed effect and regressors in (24) can be accommodated by assuming - as in Mundlak [26] - a linear relationship between ξ^m_n and the time averages of $\mathbf{q}_{nt}$:

$$\tau^{m*}_{nt} = \mathbf{q}'_{nt}\boldsymbol{\gamma}^m + \overline{\mathbf{q}}'_n\boldsymbol{\theta}^m + u^m_{nt} \qquad (25)$$

where: $\xi^m_n = \overline{\mathbf{q}}'_n\boldsymbol{\theta}^m + j^m_n$, $\overline{\mathbf{q}}_n = \frac{1}{T}\sum_{t=1}^T \mathbf{q}_{nt}$, and $u^m_{nt} \equiv (j^m_n + a^m_{nt}) \sim N(0, \sigma^2_{ut})$ is independent of $\overline{\mathbf{q}}_n$.

For $K_{nt} = 1$, let $\overline{m}$ be the number of binding lower constraints $(\tau^m_{nt} = l^m)$ and $M - \overline{m}$ the number of binding upper constraints $(\tau^m_{nt} = h^m)$, with $0 \leq \overline{m} \leq M$. According to (16), (23) and (25), and letting $\phi^M(\mathbf{u}_{nt})$ be the multivariate distribution of the vector of error terms from the M tax rate equations, the probability that authority n is fully constrained can be expressed as:

$$P(K_{nt} = 1) = \underbrace{\int_{-\infty}^{l^1 - \mathbf{q}'_{nt}\boldsymbol{\gamma}^1 - \overline{\mathbf{q}}'_n\boldsymbol{\theta}^1} ... \int_{-\infty}^{l^{\overline{m}} - \mathbf{q}'_{nt}\boldsymbol{\gamma}^{\overline{m}} - \overline{\mathbf{q}}'_n\boldsymbol{\theta}^{\overline{m}}}}_{\tau^m_{nt} = l^m} \qquad (26)$$

$$\underbrace{\int_{h^{\overline{m}+1} - \mathbf{q}'_{nt}\boldsymbol{\gamma}^{\overline{m}+1} - \overline{\mathbf{q}}'_n\boldsymbol{\theta}^{\overline{m}+1}}^{+\infty} ... \int_{h^M - \mathbf{q}'_{nt}\boldsymbol{\gamma}^M - \overline{\mathbf{q}}'_n\boldsymbol{\theta}^M}^{+\infty}}_{\tau^m_{nt} = h^m} \phi^M(\mathbf{u}_{nt})\, du^1_{nt}...du^M_{nt}$$

Estimation of (26) is complicated by the need to evaluate an M-dimensional integral.[15] For reasons of tractability, we therefore focus on the reduced form of the observed selection outcome K_{nt}:

$$K_{nt} = 1\left[K^*_{nt} = \mathbf{q}'_{nt}\boldsymbol{\delta} + \mu_n + \varepsilon_{nt} > 0\right] \tag{27}$$

meaning that $K_{nt} = 1$ if $K^*_{nt} > 0$, where K^*_{nt} is a latent variable, $\varepsilon_{nt} \sim N(0, \sigma^2_\varepsilon)$, and where μ_n is a time-invariant Province-specific effect that can be allowed to be correlated with the explanatory variables in a linear way:

$$K_{nt} = 1\left[K^*_{nt} = \mathbf{q}'_{nt}\boldsymbol{\delta} + \overline{\mathbf{q}}'_n\boldsymbol{\alpha} + v_{nt} > 0\right] \tag{28}$$

with: $\mu_n = \overline{\mathbf{q}}'_n\boldsymbol{\alpha} + \kappa_n$, $\overline{\mathbf{q}}_n = \frac{1}{T}\sum_{t=1}^{T}\mathbf{q}_{nt}$, and $v_{nt} \equiv (\kappa_n + \varepsilon_{nt})$ is normally distributed and is independent of $\overline{\mathbf{q}}_n$.[16]

Wooldridge [35] suggests the following procedure. Assume:

$$E(\eta^1_{nt}\,|\mathbf{q}_n, v_{nt}) = E(\eta^1_{nt}\,|\mathbf{q}_{n1}, ..., \mathbf{q}_{nT}, v_{nt}) = E(\eta^1_{nt}\,|v_{nt}) = \rho_{\eta t}v_{nt} \tag{29}$$

$$E(\zeta^1_n\,|\mathbf{q}_n, v_{nt}) = E(\zeta^1_n\,|\mathbf{q}_{n1}, ..., \mathbf{q}_{nT}, v_{nt}) = \overline{\mathbf{q}}'_n\boldsymbol{\psi} + \rho_{\zeta t}v_{nt} \tag{30}$$

According to (29), η^1_{nt} is mean independent of $\mathbf{q}_n$ conditional on v_{nt}, and its expectation is linear in v_{nt}, while (30) is a linearity assumption on the conditional expectation of the fixed effect ζ^1_n. No distributional assumptions are imposed on η^1_{nt} and ζ^1_n. Consequently, upon defining $\rho_t \equiv \rho_{\eta t} + \rho_{\zeta t}$, equation (17) can be written as:

$$E(z_{nt}\,|\mathbf{q}_n, v_{nt}, K_{nt} = 1) = \mathbf{q}'_{nt}\boldsymbol{\beta}^1 + \overline{\mathbf{q}}'_n\boldsymbol{\psi} + \rho_t v_{nt} \tag{31}$$

The procedure consists in estimating (28) by Probit, obtaining the inverse Mills ratio, and replacing it for v_{nt} in (31):

$$z_{nt} = \mathbf{q}'_{nt}\boldsymbol{\beta}^1 + \overline{\mathbf{q}}'_n\boldsymbol{\psi} + \rho_t\widehat{\lambda}_{nt} + o_{nt} \tag{32}$$

[15] See Heien and Wessells [14], Shonkwiler and Yen [31], Golan et al. [12], and Yen and Lin [36].

[16] In performing the Wooldridge [35] procedure, the Mundlak [26] correction seems preferable in this context to the alternative Chamberlain [6] approach. The latter consists in expressing fixed effects as a linear combination of the explanatory variables from all time periods: $\mathbf{q}'_{n1}\boldsymbol{\alpha}_1 + ... + \mathbf{q}'_{nt}\boldsymbol{\alpha}_t + ...\mathbf{q}'_{nT}\boldsymbol{\alpha}_T$. Assuming instead, in the Mundlak [26] spirit, that the fixed effect depends only on the time average of $\mathbf{q}_{nt}$ greatly conserves on parameters, at the cost of imposing the parameter restriction: $\boldsymbol{\alpha}_1 = ... = \boldsymbol{\alpha}_t = ... = \boldsymbol{\alpha}_T = \frac{1}{T}\boldsymbol{\alpha}$ in equation (28).

$$\widehat{\lambda}_{nt} = \widehat{\lambda}(\mathbf{q}'_{nt}\widehat{\boldsymbol{\delta}} + \overline{\mathbf{q}}'_n\widehat{\boldsymbol{\alpha}}) = \frac{\phi\left(\mathbf{q}'_{nt}\widehat{\boldsymbol{\delta}} + \overline{\mathbf{q}}'_n\widehat{\boldsymbol{\alpha}}\right)}{\Phi\left(\mathbf{q}'_{nt}\widehat{\boldsymbol{\delta}} + \overline{\mathbf{q}}'_n\widehat{\boldsymbol{\alpha}}\right)} \tag{33}$$

Wooldridge [35] proves that pooled OLS on (32) leads to consistent estimates of the parameter vector of interest $\boldsymbol{\beta}^1$.

5 Empirical implementation

5.1 Local tax limitations in Italy

The impact of tax limitations on the sensitivity of local public spending to grants is tested on panel data for the Italian Provinces. The Italian system of local government is organized as a three-tier structure, with the 103 Provinces constituting the intermediate level of government between the regional (20 Regions) and the municipal (over 8,000 municipalities) ones. Provinces have responsibility for intermunicipal road construction and maintenance, local transportation systems, secondary education schools, waste management and environmental protection. Provincial expenditures rose considerably in recent years, mostly due to the devolution of functions from the national and regional governments. In fact, average per capita spending increased by about 25% in real terms between 2000 and 2007.

Over $\frac{3}{4}$ of total current provincial spending is funded by grants from upper levels of government (State and Regions), with the proportion of grant-funded expenditures remaining roughly constant through the 2000-2007 period. State grants are for the most part general and formula-based. They rely on the definition of a standardized spending level for each Province built on exogenous needs indicators falling into three broad areas (age structure of the resident population; geomorphological complexion; socioeconomic deprivation), as well as of a fiscal capacity index capturing the ability of each Province to raise own and shared revenues.[17] In particular, Provinces are divided into four demographic bands, and average service cost indices for a number of mandated provincial functions and average tax bases are periodically computed (usually every three

[17] Shared revenues crowd-out grants one-for-one, and are therefore subsumed into the latter.

years) for each band. Expenditures on non-mandated provincial services do not enter the grant distribution scheme and must be entirely funded by own revenues. On the other hand, Regional grants typically finance specific functions that were devolved to Provinces during the decentralization process of the late 1990s.

As a result of the above institutional arrangement, State and regional grants can to a large extent be considered exogenous with respect to own funding decisions by provincial governments. In particular, given the infrequent central assessment of spending needs and fiscal capacity, changes in provincial socioeconomic conditions are not promptly reflected into State grant adjustments.[18] Moreover, the fact that State grants are based on a Province's needs and fiscal capacity indices relative to its demographic band mean should alleviate the potential problem of grant endogeneity arising from shocks moving grants and local expenditures in the same direction.[19]

The rest of current spending is funded by three own tax revenue sources: the vehicle registration tax, the electricity consumption tax, and the waste management tax. The vehicle registration tax represents over 50% of total own tax revenues. All brand new vehicles - as well as used vehicles in case of change of ownership - are liable to the payment of the tax the first time they are registered in the provincial archive under a given owner's name. The total tax due is made of a lump-sum amount plus a variable component that is related to the size, power and destination of the vehicle. As shown in table 1, central government establishes a lower and an upper bound on the vehicle tax parameters that Provinces can set, with the upper bound corresponding to a 20% higher tax burden (raised to 30% in 2007) than the one corresponding to the lower bound. Consequently, the decision of each Province basically consists in determining autonomously the surcharge τ^v. The electricity consumption tax is applied by

[18] Gordon [13] exploits the infrequent updating of poverty data used in the US federal education grants to school districts (Title I). However, since Title I grants also depend on state education spending, she computes a purely Census-determined grant change measure and uses it as an instrument for actual Title I revenue change - a step that seems unnecessary in our simpler context.

[19] A similar point is made in Brooks and Phillips [5].

Provinces on business uses of electricity. As shown in table 1, Provinces set a tax rate τ^e between a minimum of 9.3 and a maximum of 11.4 Euro cents per kW. Electricity tax revenues correspond to above $\frac{1}{3}$ of total own tax revenues. Finally, the waste management tax is a surcharge applied by Provinces on the waste collection bill charged by the municipalities located in the province on all households and businesses. Table 1 shows that the surcharge τ^w must lie between 1% and 5% of the municipal levy. Revenues from the waste management tax amount to about 10% of total provincial own tax revenues.

Table 2 reports the number of authorities setting tax rates at the lower and upper limits respectively, while table 3 rates the authority-year observations based on how severely they are affected by the tax rate limitations.[20] More than half of the observations in the dataset (416 out of 720) correspond to fully bound instances, with all available tax sources being set at left or right corners, while in only 9 observations none of the constraints is binding. In about 40% of the observations, either one or two tax limitations are binding.

5.2 Time-invariant splitting indicator

The sample is first split based on a time-invariant indicator K_n that equals 1 if Province m is constrained on all own tax revenue sources for the entire period of observation, and equals 0 if the authority never has all constraints binding. This accords with the splitting criterion (15). By doing so, we rule out the effect of yearly changes in grants on the capping status, and focus on the difference in the response of local public expenditures to grant changes. Moreover, since selection effects occur through correlation between K_n and the time-invariant authority-specific effects ζ_n^1 and ζ_n^0 in (13)-(14), any selection bias is cancelled by applying the fixed effects estimator (19) on the two subsamples.

In our dataset, $K_n = 1$ for 24 provincial authorities, and $K_n = 0$ for 20 authorities in the 2001-2006 period, the rest of the observations being discarded (to be used later on) because of changing regime during the period. This leaves

[20]The data refer to the 90 Provinces (out of 103) for which all information from 2000 to 2007 is available.

us with 264 observations.[21] Of the 24 structurally capped authorities, 17 were at the upper bounds on all three own tax rates for the entire period, 5 were hitting two upper bounds and one lower bound, one Province was at one upper and two lower bounds, and one Province was consistently at the three lower bounds. On the other hand, the authorities in the $K_n = 0$ regime have one to two constraints binding.

We first estimate the switching regression model (13)-(14)-(15) as a single equation, with K_n working as a switcher, thus allowing us to test the difference between the $\boldsymbol{\beta}^0$ and $\boldsymbol{\beta}^1$ coefficient vectors:

$$z_{nt} = \mathbf{q}'_{nt}\boldsymbol{\beta}^0 + K_n \times \mathbf{q}'_{nt}(\boldsymbol{\beta}^1 - \boldsymbol{\beta}^0) + \zeta_n^0 + \eta_{nt}^0 + K_{nt} \times ((\zeta_n^1 + \eta_{nt}^1) - (\zeta_n^0 + \eta_{nt}^0)) \quad (34)$$

where z_{nt} equals current spending per capita in real terms, and the vector of explanatory variables $\mathbf{q}'_{nt} = [g_{nt}\ i_{nt}\ \mathbf{x}'_{nt}]$ includes:

g_{nt}: per capita grants = all current financial transfers from upper levels of government (State and Regions), including the fixed shares of national tax revenues devolved to Provinces (national personal income tax and national motor-vehicle insurance tax);

i_{nt}: per capita consumption of electricity for domestic uses as a proxy for private income;

$\mathbf{x}_{nt}$: population size to control for economies of scale in service provision; demographic composition of the resident population (share of the population aged 0 to 4 years and aged over 65 years); a binary election year indicator to allow for opportunistic policy manoeuvring prior to elections;[22] a binary government ideology indicator to capture a partisan cleavage in spending policy between right-wing and left-wing governments.[23]

[21] In defining the time-invariant K_n dummy variable, it seems sensible to exclude the last year in the sample because the vehicle tax cap relaxation that occurred in 2007 (from 20% to 30%) actually changed in an abrupt way the structually capped status of a non-negligible number of Provinces. This would considerably shrink the number of authorities in the structurally capped regime. Similarly, in order to preserve the size of the $K_{nt} = 1$ sample, the first year (2000) is excluded since several Provinces became consistently capped from the year 2001 on.

[22] Provincial elections take place every five years with direct election of the President of the Province. The provincial election schedule is asymmetric, meaning that Provinces hold elections at different points in time.

[23] In particular, we use a right-wing control dummy. In most instances, the President of the

The fixed effects estimation results of equation (34) are reported in table 4, while table 5 reports the separate estimation results of equations (13) and (14) for the two subsamples. All equations include year dummies. Descriptive statistics and data sources for all variables are reported in Appendix A.

Interestingly, all authorities exhibit what would be termed a flypaper effect according to conventional criteria in the literature. The results in table 4 show that the grant effect is large and highly significant, while the coefficient on the income proxy is hardly different from zero. In fact, the $K_n = 0$ subsample is far from being unconstrained in practice, given that those authorities are capped on one or two tax rates along the period. However, structurally capped authorities' expenditures react to grants to a significantly larger extent, actually on a one-for-one basis. The estimate of the effect of grants on spending is around 0.7 for the moderately constrained subsample (columns (4.2), (4.4) and (4.6) in table 4; columns (5.2), (5.4) and (5.6) in table 5), while the coefficient estimate virtually equals 1 for structurally bound Provinces (columns (5.1), (5.3) and (5.5) in table 5).[24] The results are robust to the introduction of various controls, none of which, though, due to their comprehensibly limited variation over the relatively short time period and the overwhelming role of grants, contributes much to further explaining the pattern of spending.

One might wonder at this point whether the grant coefficient estimate is in fact inflated by spurious correlation between local expenditure and grants due to omitted variables driving both. However, a grant coefficient estimate of around 1 in the fully constrained sample is hardly surprising. Given that other sources of revenue (including balances) constitute a negligible average share of total provincial revenues, equation (6) represents a reasonable approximation to the actual budgeting choice, meaning that $\frac{\partial z^*(\mathbf{q})}{\partial g} \cong 1$ if all tax rates are frozen at their (upper or lower) limits. As for the moderately constrained sample, an endogeneity bias would most likely play against the point we are making here,

Province is backed by a well-defined right-wing or left-wing coalition.

[24] When estimating a constant elasticity specification (results not reported in full, but available upon request), the grant and income elasticities equal 0.8 and 0.1 in the $K_{nt} = 1$ sample, and around 0.6 and 0.5 in the $K_{nt} = 0$ sample.

in the sense of driving up the estimate of β_g^0 and narrowing the gap between the two subsamples.

5.3 Endogenous time-varying selection

In order to implement the Wooldridge [35] procedure and allow for endogenous selection into the $K_{nt} = 1$ regime and fixed Province effects, we now need to focus on the authorities that are observed to be switching between regimes over time. This allows us to proceed as discussed in section **4**, and estimate the Probit selection equation (28) in the first stage, with K_{nt} as defined in equation (16).

After excluding Provinces that are consistently constrained or unconstrained over the entire time period, as well as those observed in the fully constrained $K_{nt} = 1$ regime for less than two years, we end up with a balanced panel of 43 "switching" Provinces over the eight years 2000-2007. Importantly, since we aim at estimating the response of authority n's spending to grants in year t provided that authority n stays on the same portion of its budget constraint, we require the fully constrained tax mix of authority n in year t to be identical as in year $t-1$ in order for that observation to be selected into the $K_{nt} = 1$ regime (Bond and Meghir [3]).

Table 6 reports the benchmark fixed effects estimation results of the switching regression model (16)-(17)-(18) under assumption (22) of exogenous selection into the fully capped regime. It is remarkable that local authorities' expenditures exhibit the expected excess sensitivity when fully constrained $\left(\beta_g^1 \cong 1\right)$, while the sensitivity of spending to grants is significantly lower $\left(\beta_g^0 < 0.8\right)$ and the sensitivity to income is higher $\left(\beta_i^0 > \beta_i^1 \cong 0\right)$ when the same authorities are moderately constrained.

Table 7 reports the two-stage estimation results allowing for endogenous selection and fixed authority effects both in the selection equation and in the main equation. Probit estimation of the first-stage binary selection equation (28) - columns (7.1)-(7.2) - reveals that grants have no systematic impact on selection into the $K_{nt} = 1$ regime. This is to be expected due to the constrained regime heterogeneity including high spending authorities hitting the upper tax bounds and low spending authorities hitting the lower tax bounds: in fact, almost 40% of the observations in the selected $K_{nt} = 1$ sample hit at least one lower tax limit, making it difficult to predict the impact of the variables in the $\mathbf{q}_{nt}$ vector on the reduced form selection index. As far as the stochastic component of equation (28) is concerned, the second stage estimation results suggest that the selection process can be considered exogenous with respect to the local public spending pattern described by equation (17): the Wooldridge [35] variable addition test reported at the bottom of table 7 is far from statistical significance.[25] Finally, columns (7.3)-(7.4) reveal that performing the Wooldridge [35] correction discussed in section **4** - equation (32) - has a negligible impact on the estimation results, including in particular the excess sensitivity of local public spending to grants.

6 Concluding remarks

By explicitly recognizing and incorporating the left and right corners that are typically produced by state-wide limitations on local tax rates, this paper has modelled the local tax policy determination process and shown how excess sensitivity of local public spending to grants arises in the endogenously generated constrained tax mix.

In particular, the paper has shown that the effect of private income on public spending should be expected to be tiny or nil in the presence of binding limitations on all local tax revenue sources, while grants should be predicted to have a large - actually, a one-for-one - impact on local expenditures. Interestingly, the above result holds when either upper or lower tax limitations are in place, and the analysis shows that a binding limitation on just one of the available own revenue sources is enough to generate some form of flypaper effect, in the

[25] The test relies on fixed effects estimation of the spending equation for the unbalanced panel of $K_{nt} = 1$, after adding the inverse Mills' ratio from the first stage Probit estimation on the balanced panel. The null hypothesis of the test is: $E(\eta^1_{nt} \, |\zeta^1_n, \mathbf{q}_n, \mathbf{v}_n) = 0$ against the alternative: $E(\eta^1_{nt} \, |\zeta^1_n, \mathbf{q}_n, \mathbf{v}_n) = \rho v_{nt}$, while no hypothesis is made on how ζ^1_n depends on $\mathbf{q}_n, \mathbf{v}_n$ (Wooldridge [35]).

sense of an excess sensitivity of local spending to grants. In fact, since excess sensitivity of local public spending should be predicted to arise and generally tends to manifest itself both when grants increase and when they decrease, the "flypaper effect" label seems an inappropriate or even misleading one.

By using panel data on the Italian Provinces over the years 2000s, the paper has exploited the clustering of provincial authorities at the corners generated by central government lower and upper tax limitation rules to estimate the sensitivity of local public expenditures to grants. The empirical evidence arising from a switching regression approach that allows for fixed effects and endogenous selection into the constrained sample consistently suggests that the response of local spending to grants is significantly higher for fully capped authorities than for authorities that can manoeuvre at least one tax instrument. While not dismissing the role of alternative explanations of the excess sensitivity of local public spending, it seems that the intensity of tax and expenditure limitations ought not to be ignored when investigating the local tax mix determination process and the reaction of local authorities to central government grant policy.

This paper represents just the first step into the investigation of the decentralized tax and spending determination process within a formal framework that fully encompasses the frequently encountered top-down fiscal limitations on local decision makers, and much remains to be unveiled about how tax and expenditure limitations arise, why they differ so widely across states and countries, and in the end to what extent the kinky behavior of local decision-makers is in fact their optimal - or possibly unique - response when deliberately driven into a corner.

References

[1] Amemiya, T., Multivariate regression and simultaneous equation models when the dependent variables are truncated normal, Econometrica 42 (1974) 999-1012.

[2] Anderson, N., Property tax limitations: An interpretative review, National Tax Journal 59 (2006) 685-694.

[3] Bond, S., Meghir, C., Dynamic investment models and the firms' financial policy, Review of Economic Studies 61 (1994) 197-222.

[4] Borge, L-E., Tovmo, P., Myopic or constrained by balanced-budget rules? The intertemporal spending behavior of Norwegian local governments, FinanzArchiv: Public Finance Analysis, 65 (2009) 200-219.

[5] Brooks, L., Phillips, J., An institutional explanation for the stickiness of federal grants, Journal of Law, Economics and Organization (2010) *forthcoming.*

[6] Chamberlain, G., Multivariate regression models for panel data, Journal of Econometrics 18 (1982) 5-46.

[7] Charlier, E., Melenberg, B.,Van Soest, A., An analysis of housing expenditure using semiparametric models and panel data, Journal of Econometrics 101 (2001) 71-107.

[8] Cummins, J., Hassett, K., Oliner, S., Investment behavior, observable expectations, and internal funds, American Economic Review 96 (2006) 796-810.

[9] Dustmann, C., Rochina-Barrachina, M.E., Selection correction in panel data models: An application to the estimation of females' wage equations, Econometrics Journal 10 (2007) 263-293.

[10] Fazzari, S., Hubbard, R., Petersen, B., Financing constraints and corporate investment, Brookings Papers on Economic Activity 1 (1988) 141-195.

[11] Gamkhar, S., Oates, W., Asymmetries in the response to increases and decreases in intergovernmental grants: Some empirical findings, National Tax Journal 49 (1996) 501-512.

[12] Golan, A., Perloff, J., Shen, E., Estimating a demand system with non-negativity constraints: Mexican meat demand, Review of Economics and Statistics 83 (2001) 541-550.

[13] Gordon, N., Do federal grants boost school spending? Evidence from Title I, Journal of Public Economics 88 (2004) 1771-1792.

[14] Heien, D., Wessells, C., Demand systems estimation with microdata: A censored regression approach, Journal of Business and Economic Statistics 8 (1990) 365-371.

[15] Hettich, W., Winer, S., Economic and political foundations of tax structure, American Economic Review 78 (1988) 701-712.

[16] Hettich, W., Winer, S., A positive model of tax structure, Journal of Public Economics 24 (1984) 67-87.

[17] Hines, J., Thaler, R., Anomalies. The flypaper effect, Journal of Economic Perspectives 9 (1995) 217-226.

[18] Hu, X., Schiantarelli, F., Investment and capital market imperfections: A switching regression approach using US firm panel data, Review of Economics and Statistics 80 (1998) 466-479.

[19] Inman, R., Flypaper effect, in: S. Durlauf and L. Blume (Eds.) The New Palgrave Dictionary of Economics Online, Palgrave Macmillan, 2009, DOI:10.1057/9780230226203.1901.

[20] Jappelli, T., Pischke, J-S., Souledes, N., Testing for liquidity constraints in Euler equations with complementary data sources, Review of Economics and Statistics 80 (1998) 251-262.

[21] Joumard, I., Kongsrud, M., Fiscal relations across government levels, OECD Economics Department Working Paper 375 (2003).

[22] Kaplan, S., Zingales, L., Do investment-cash flow sensitivities provide useful measures of financing constraints? Quarterly Journal of Economics 112 (1997) 169-215.

[23] Kenny, L., Winer, S., Tax systems in the world: An empirical investigation into the importance of tax bases, administration costs, scale and political regime, International Tax and Public Finance 13 (2006) 181-215.

[24] Kyriazidou, E., Estimation of a panel data sample selection model, Econometrica 65 (1997) 1335-1364.

[25] Lutz, B., Taxation with representation: Intergovernmental grants in a plebiscite democracy, Review of Economics and Statistics 92 (2010) 316-332.

[26] Mundlak, Y., On the pooling of time series and cross section data, Econometrica 46 (1978) 69-85.

[27] Nechyba, T., Local property and state income taxes: The role of interjurisdictional competition and collusion, Journal of Political Economy 105 (1997) 351-384.

[28] Payne, A., Does government funding change behavior? An empirical analysis of crowd-out, Tax Policy and the Economy 23 (2009) 159-184.

[29] Pudney, S., Modelling Individual Choice. The Econometrics of Corners, Kinks and Holes. Blackwell, Oxford, 1991.

[30] Runkle, D., Liquidity constraints and the permanent income hypothesis: Evidence from panel data, Journal of Monetary Economics 27 (1991) 73-98.

[31] Shonkwiler, J., Yen, S., Two-step estimation of a censored system of equations, American Journal of Agricultural Economics 81 (1999) 972-982.

[32] Stine, W., Is local government revenue response to federal aid symmetrical? Evidence from Pennsylvania county governments in an era of retrenchment, National Tax Journal 47 (1994) 799-816.

[33] Sutherland, D., Price, R., Joumard, I., Sub-central government fiscal rules, OECD Economic Studies 41 (2005) 141-181.

[34] Wolman, H., McManmon, R., Bell, M., Brunori, D., Comparing local government autonomy across states, Working Paper No. 35/2008, George Washington Institute for Public Policy, Washington.

[35] Wooldridge, J., Selection corrections for panel data models under conditional mean independence assumptions, Journal of Econometrics 68 (1995) 115-132.

[36] Yen, S., Lin, B., A sample selection approach to censored demand systems, American Journal of Agricultural Economics 88 (2006) 742-749.

[37] Zeldes, S., Consumption and liquidity constraints: An empirical investigation, Journal of Political Economy 97 (1989) 305-346.

Appendix A: Data description

Table A1 Variables used in the analysis: descriptive statistics

	obs.	mean	s.d.	min	max
Vehicle registration surtax rate (%)	720	17.7	7.2	0	30
Electricity consumption tax rate (€ cents per kW)	720	10.4	1.0	9.3	11.4
Waste management surtax rate (%)	720	4.5	1.1	1	5
Electricity (domestic consumption per capita, kW)	720	1089.7	113.3	771.5	1484.2
Population (,000)	720	567.9	631.6	89.0	4061.5
Aged 0-4 share	720	4.4	0.6	3.0	6.3
Aged 65+ share	720	20.4	3.1	12.0	27.5
Real current spending per capita (€)	720	146.1	46.1	56.9	291.9
Real grants per capita (€)	720	118.2	44.7	36.6	249.0
Election year (%)	720	15.4			
Right-wing control (%)	720	33.6			

Table A2 Variables used in the analysis: data sources

	Data source
Vehicle registration tax rate	Automobile Club Italy - Quattroruote
Electricity consumption tax rate	Italian Government, Ministry of Finance
Electricity (domestic consumption)	Terna Rete Elettrica Nazionale - S.p.A.
Waste management tax rate	Italian Government, Home Office
Current spending	Italian Government, Home Office
Grants	Italian Government, Home Office
Election year	Italian Government, Home Office
Right-wing control	Italian Government, Home Office
Population & demographics	National Statistics Institute

Table 1 Lower and upper tax limitation rules

		2000-6	2007
Vehicle registration tax τ^v	l^v	0	0
(% surcharge on national rate)	h^v	20	30
Electricity consumption tax τ^e	l^e	9.3	9.3
(Euro cents per kW)	h^e	11.4	11.4
Waste management tax τ^w	l^w	1	1
(% surcharge on municipal levy)	h^w	5	5

Table 2 Number of authorities ($N = 90$) at lower (l) and upper (h) limits

		2000	2001	2002	2003	2004	2005	2006	2007
τ^v	l^v	25	15	9	7	7	5	4	3
	h^v	55	65	71	72	73	77	79	43
τ^e	l^e	66	54	43	37	34	27	18	15
	h^e	16	29	39	45	47	52	59	64
τ^w	l^w	3	2	2	3	3	2	3	3
	h^w	66	64	66	65	65	66	66	68

Table 3 Tax limitation intensity

	Fully constrained												total
	(h,h,h)	(h,h,l)			(h,l,l)			(l,l,l)					
τ^v	h^v	h^v	h^v	l^v	h^v	l^v	l^v	l^v					
τ^e	h^e	h^e	l^e	h^e	l^e	h^e	l^e	l^e					
τ^w	h^w	l^w	h^w	h^w	l^w	l^w	h^w	l^w					
obs.	238	6	130	0	3	0	28	11					416
	Moderately constrained												
	(h,h)			(h,l)						(l,l)			
τ^v	h^v	h^v	τ^{v*}	h^v	l^v	h^v	l^v	τ^{v*}	τ^{v*}	l^v	l^v	τ^{v*}	
τ^e	h^e	τ^{e*}	h^e	l^e	h^e	τ^{e*}	τ^{e*}	h^e	l^e	l^e	τ^{e*}	l^e	
τ^w	τ^{w*}	h^w	h^w	τ^{w*}	τ^{w*}	l^w	h^w	l^w	h^w	τ^{w*}	l^w	l^w	
obs.	61	60	27	36	3	0	0	1	39	32	0	0	
	(h)			(l)									
τ^v	h^v	τ^{v*}	τ^{v*}	l^v	τ^{v*}	τ^{v*}							
τ^e	τ^{e*}	h^e	τ^{e*}	τ^{e*}	l^e	τ^{e*}							
τ^w	τ^{w*}	τ^{w*}	h^w	τ^{w*}	τ^{w*}	l^w							
obs.	1	15	4	1	15	0							295
	Not constrained												
τ^v	τ^{v*}												
τ^e	τ^{e*}												
τ^w	τ^{w*}												
obs.	9												9
													720

Table 4 Time-invariant splitting criterion (K_n)

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)
grants (g_{nt})	$0.844^{\circledast}$	$0.722^{\circledast}$	$0.840^{\circledast}$	$0.716^{\circledast}$	$0.842^{\circledast}$	$0.731^{\circledast}$
	(0.035)	(0.046)	(0.034)	(0.046)	(0.035)	(0.049)
income (i_{nt})			0.040^{*}	0.055^{*}	0.033	0.042
			(0.021)	(0.025)	(0.023)	(0.030)
population					-0.091	-0.142
					(0.081)	(0.124)
age 0-4 share					-2.147	-5.750
					(5.986)	(7.569)
age 65+ share					-3.789	-6.076
					(3.340)	(5.217)
election					0.300	-0.315
					(1.356)	(1.784)
right-wing					2.176	6.767
					(2.881)	(6.662)
$K_n \times$ grants (g_{nt})		$0.253^{\circledast}$		$0.258^{\circledast}$		$0.259^{\circledast}$
		(0.069)		(0.069)		(0.073)
$K_n \times$ income (i_{nt})				-0.047		-0.027
				(0.045)		(0.050)
$K_n \times$ population						0.094
						(0.163)
$K_n \times$ age 0-4 share						15.520
						(12.217)
$K_n \times$ age 65+ share						10.066
						(7.047)
$K_n \times$ election						0.898
						(2.718)
$K_n \times$ right-wing						-5.677
						(7.393)
observations	264	264	264	264	264	264
authorities	44	44	44	44	44	44

Notes: Dependent variable: real current spending per capita. Fixed Province and year effects included; year effects interacted with the switching indicator K_n in columns (4.2) and (4.4). K_n defined in equation (15). Standard errors in parentheses. $^{\circledast}$: p-value < 0.01; *: p-value < 0.10.

Table 5 Time-invariant splitting criterion: separate equations

	(5.1)	(5.2)	(5.3)	(5.4)	(5.5)	(5.6)
	K_n= 1	K_n= 0	K_n= 1	K_n= 0	K_n= 1	K_n= 0
grants (g_{nt})	0.975$^{\circledast}$	0.722$^{\circledast}$	0.974$^{\circledast}$	0.716$^{\circledast}$	0.989$^{\circledast}$	0.731$^{\circledast}$
	(0.048)	(0.050)	(0.048)	(0.050)	(0.051)	(0.053)
income (i_{nt})			0.008	0.055*	0.015	0.042
			(0.035)	(0.026)	(0.037)	(0.032)
population					-0.048	-0.142
					(0.099)	(0.133)
age 0-4 share					9.770	-5.750
					(8.986)	(8.115)
age 65+ share					3.991	-6.076
					(4.439)	(5.593)
election					0.583	-0.315
					(1.921)	(1.913)
right-wing					1.090	6.767
					(3.003)	(7.143)
observations	144	120	144	120	144	120
authorities	24	20	24	20	24	20

Notes: Dependent variable: real current spending per capita. Fixed Province and year effects included. K_n defined in equation (15). Standard errors in parentheses. $^{\circledast}$: p-value $<$ 0.01; *: p-value $<$ 0.10.

Table 6 Time-varying splitting criterion: switching Provinces

	(6.1)	(6.2)	(6.3)	(6.4)	(6.5)	(6.6)
	K_{nt}= 1	K_{nt}= 0	K_{nt}= 1	K_{nt}= 0	K_{nt}= 1	K_{nt}= 0
grants (g_{nt})	0.986$^{\circledast}$	0.792$^{\circledast}$	0.986$^{\circledast}$	0.785$^{\circledast}$	0.979$^{\circledast}$	0.793$^{\circledast}$
	(0.023)	(0.056)	(0.023)	(0.056)	(0.025)	(0.057)
income (i_{nt})			0.004	0.042	0.001	0.047
			(0.025)	(0.033)	(0.026)	(0.050)
population					0.044	0.026
					(0.053)	(0.056)
age 0-4 share					-3.987	-15.957*
					(5.353)	(8.896)
age 65+ share					0.255	-2.318
					(2.405)	(3.810)
election					1.126	0.160
					(1.353)	(2.527)
right-wing					0.993	7.239
					(2.461)	(5.747)
observations	230	114	230	114	230	114
authorities			43			

Notes: Dependent variable: real current spending per capita. Fixed Province and year effects included. K_{nt} defined in equation (16). Standard errors in parentheses. $^{\circledast}$: p-value $<$ 0.01; *: p-value $<$ 0.10.

Table 7 Wooldridge two-stage approach ($K_{nt} = 1$)

	(7.1)	(7.2)	(7.3)	(7.4)
	First stage (K_{nt})		Second stage (z_{nt})	
	Probit (balanced)		Wooldridge correction	
	q_{nt}	$\overline{q}_n$	q_{nt}	$\overline{q}_n$
grants (g_{nt})	-0.003	0.001	0.969$^{\circledast}$	-0.039
	(0.004)	(0.005)	(0.045)	(0.050)
income (i_{nt})	-0.002	0.004	0.009	0.015
	(0.004)	(0.004)	(0.044)	(0.047)
population	-0.010	0.010	0.090	-0.089
	(0.007)	(0.007)	(0.099)	(0.100)
age 0-4 share	-4.086$^{\circledast}$	4.168$^{\circledast}$	6.408	-3.221
	(0.816)	(0.872)	(22.653)	(21.597)
age 65+ share	-0.916*	0.921*	2.707	-0.363
	(0.389)	(0.395)	(6.127)	(5.925)
election	0.292	-1.533	3.137	-14.817
	(0.267)	(1.619)	(2.832)	(11.938)
right-wing	-0.626	0.317	-1.431	-0.677
	(0.525)	(0.594)	(5.519)	(5.069)
λ_{n2000}			-11.337 (10.410)	
λ_{n2001}			-5.202 (12.476)	
λ_{n2002}			6.908 (13.097)	
λ_{n2003}			19.600 (21.033)	
λ_{n2004}			24.571 (20.813)	
λ_{n2005}			-16.283 (16.846)	
λ_{n2006}			-6.494 (14.890)	
λ_{n2007}			-3.103 (9.673)	
observations	344		230	
authorities	43			
Wald test $\boldsymbol{\rho} = 0$ (p value)			0.99 (0.44)	
Wooldridge test λ_{nt} (p value)			-0.631 (0.95)	

Notes: Year effects included in all columns. Column (7.3): deviations from group means. K_{nt} defined in equation (16). Standard errors in parentheses. $^{\circledast}$: p-value $<$ 0.01; *: p-value $<$ 0.10.

Figure 1: Communicating vessels

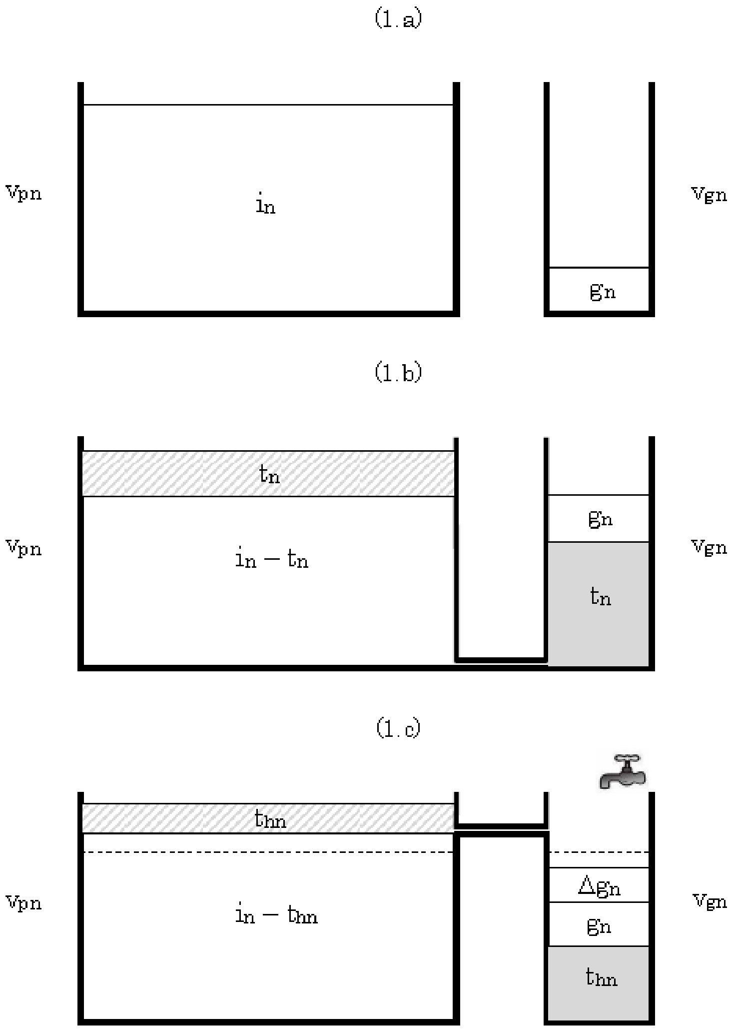

Figure 2: One tax: income change

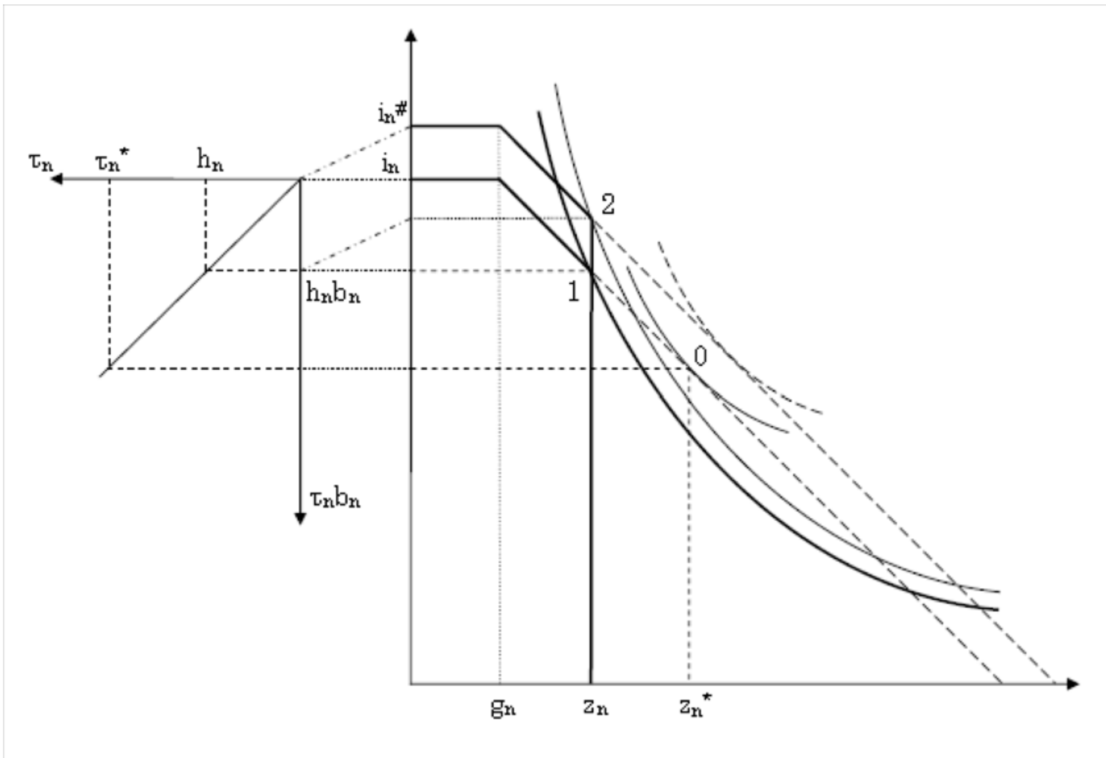

Figure 3: One tax: grant change

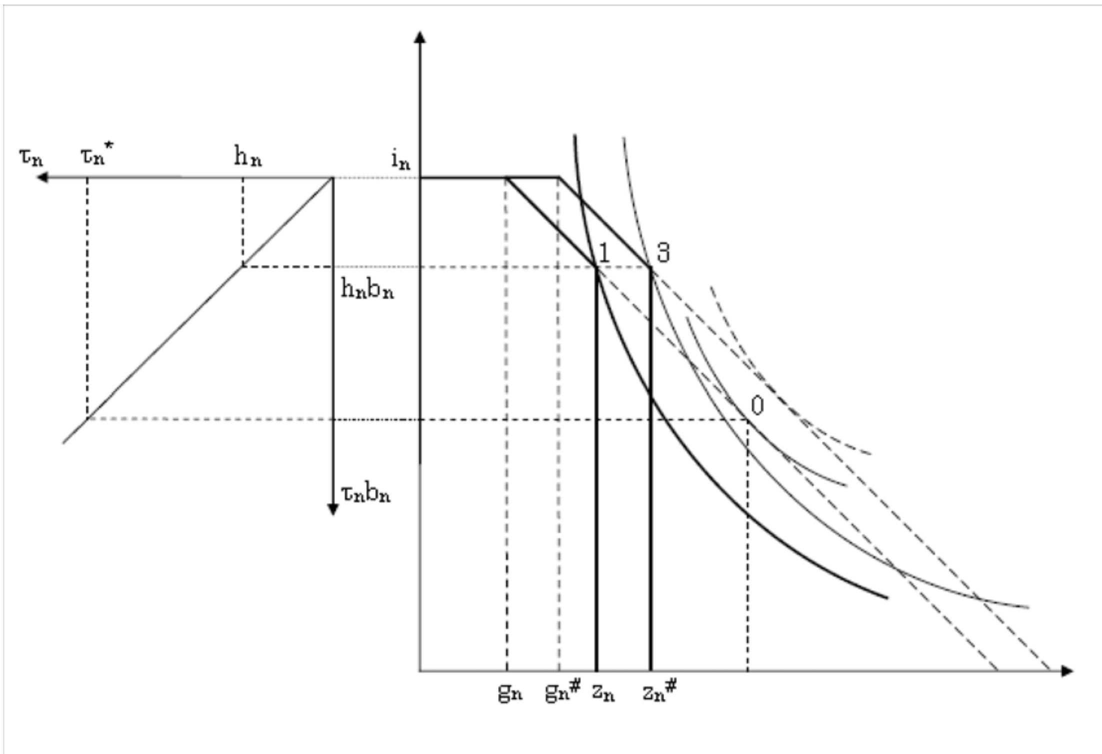

Figure 4: Two taxes: no limitations

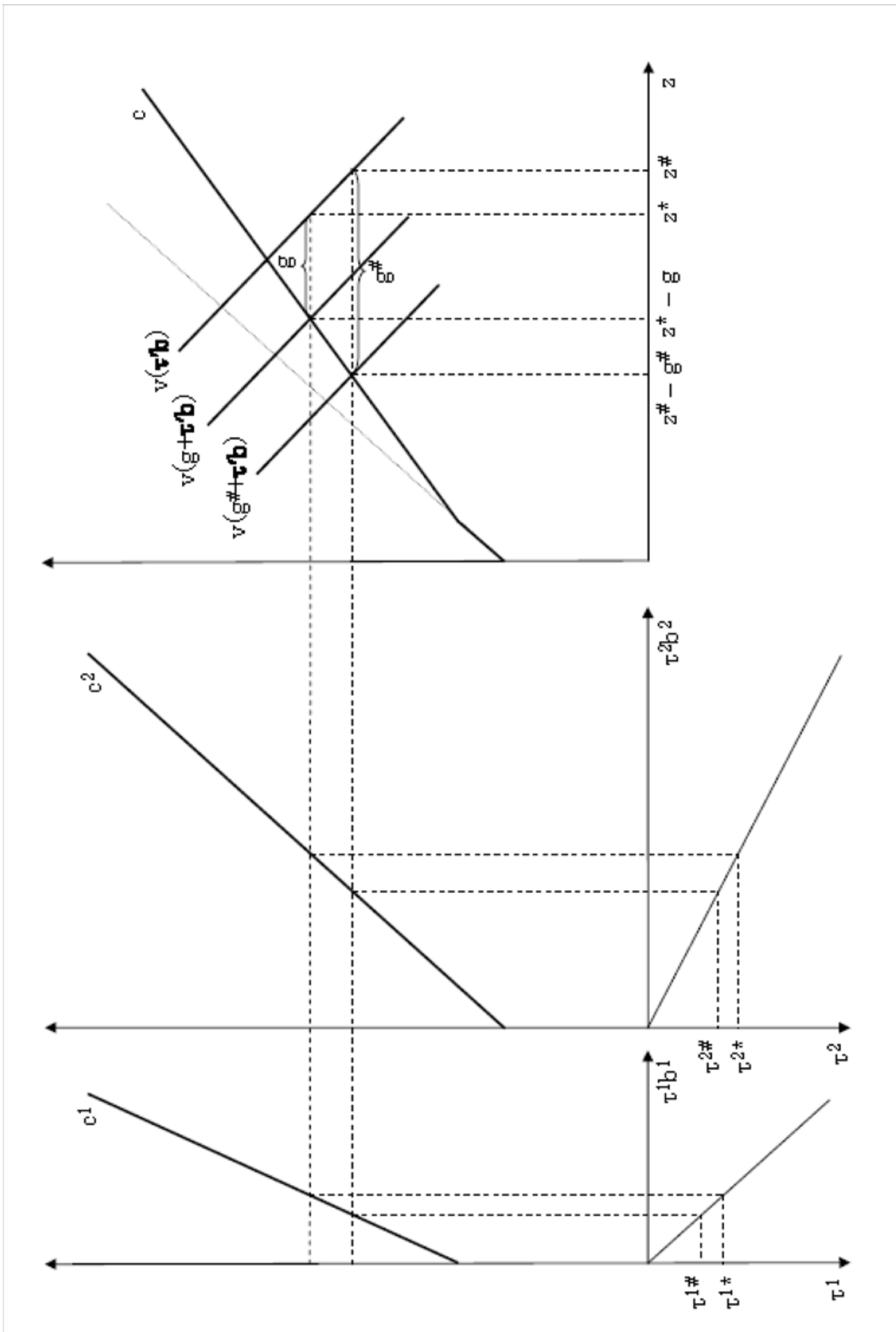

Figure 5: Two taxes: LL and RR cases

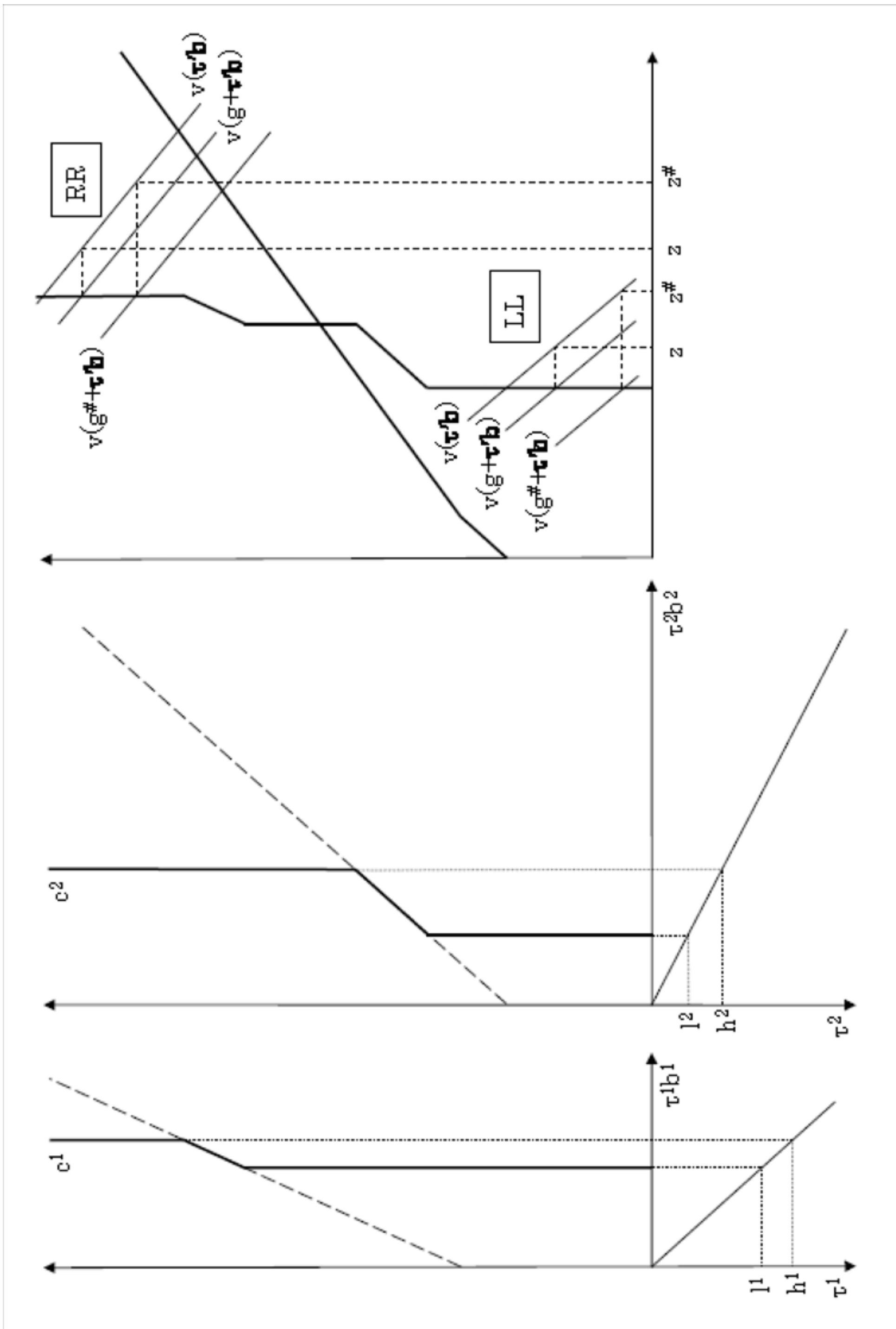

Figure 6: Two taxes: LU, LR and UR cases