FinanzArchiv

Public Finance Analysis

3

Volume 68
September
2012

Mohr Siebeck

FinanzArchiv /Public Finance Analysis

is published quaterly by

Mohr Siebeck, P.O. Box 2040, 72010 Tübingen, Germany.

FinanzArchiv is printed on acid-free and permanent/book paper ♾
meeting all library standards including ANSI/NISO Z39.48-1988.

Subscriptions/Full Text Online

New orders, renewals, claims, and all other correspondence relating to subscriptions should be addressed to the Journals Marketing Manager at the publisher's address above (fax +49 7071 51104 or e-mail to **journals@mohr.de**).
Please quote the name of the journal. – Subscription rate is € 329.– per volume for institutions and € 169.– for individuals plus postage (airmail rates on request). Cloth covers for binding are € 16.50 per volume. For back volumes ask for availability and quotation.
Free access to the full text online is included in a subscription. We ask institutions with more than 20.000 users to obtain a price quote directly from the publisher. Contact: sandra.witt @mohr.de.
In order to set up Online access for institutions / libraries, please go to: http://www.ingentaconnect.com/register/institutional. In order to set up Online access for private persons, please go to: http://www.ingentaconnect.com/register/personal.

Transfer of Rights

When an article is accepted for publication, the exclusive copyright is granted to Mohr Siebeck for publication in a print and an electronic version. Further information on this and the rights retained by the author can be found at **www.mohr.de/fa**. No one may reproduce or distribute the entire journal or parts of it in a print or an electronic version without the publisher's permission. Please contact **rights@mohr.de.**

Internet

For information on all Mohr Siebeck books, journals and services log onto **www.mohr.de.** – **FA** has its own page with up-to-date tables of content, abstracts, sample articles, and with detailed instructions for authors: **www.mohr.de/fa.**

Advertising

For details contact Tilman **Gaebler**,
P.O. Box 113, 72403 Bisingen, Germany, Phone +49 7476 3405, Fax +49 7476 3406,
tilman.gaebler@t-online.de.

© 2012 by Mohr Siebeck, Tübingen, Germany.

This journal and the individual contributions contained in it are protected by the copyright of Mohr Siebeck and may not be reproduced in any form (beyond that permitted by copyright law) without the publisher's written permission. This applies particularly to reproductions, translations, micro-films as well as storage and processing in electronic systems.
Ask **jill.sopper@mohr.de** by e-mail or the Permissions Manager at the publisher's address by mail. – In the USA, permissions may be cleared and payments made through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, Phone (978) 750 8400, Fax (978) 750 47744, www.copyright.com. Other countries may have a local reprographic rights agency for payments.

ISSN 0015-2218

Instructions to authors/Submissions

Manuscripts should be submitted online via the homepage of FinanzArchiv/Public Finance Analysis (**www.mohr.de/en/journals/economics/finanzarchiv-fa/manuscripts.html**). Submission and verification comprise five easy steps and allow to upload a cover letter as well as the manuscript in pdf format. There is no submission fee. Submission of the paper will be held to imply that it contains original unpublished work and is not being submitted for publication elsewhere. In the case of inadequate internet access you can send a pdf version of the manuscript as an e-mail attachment directly to one of the three editors.

Professor Christian **Keuschnigg**
University of St. Gallen, FGN-HSG
Varnbuelstrasse 19
CH-9000 St. Gallen
christian.keuschnigg@unisg.ch

Professor Jenny **Ligthart**
Tilburg University
PO Box 90153
5000 LE Tilburg/Netherlands
j.ligthart@uvt.nl

Professor Alfons J. **Weichenrieder**
Managing Editor
University of Frankfurt (RuW 45)
Grüneburgplatz 1
60323 Frankfurt (Main)/Germany
a.weichenrieder@em.uni-frankfurt.de

Manuscripts must be in English, American spelling is preferred. Manuscripts should be typed double-spaced with wide margins.

There is no submission fee. Submission of a paper will be held to imply that it contains original unpublished work and is not being submitted for publication elsewhere. The editors are not responsible for damage or loss of papers submitted.

The **first page** of the manuscript should contain the following information:

1. title of the paper
2. name(s) of author(s)
3. date of submission
4. abstract (not exceeding 100 words)
5. keywords (up to five)
6. Journal of Economic Literature classification

The **last page** of the manuscript should include the full address (including email-address) of the corresponding author.

All submitted manuscripts, which qualify for publication, are peer reviewed by two referees. The final version of papers accepted for publication must be sent to the managing editor as a pdf-file and as a Winword or LaTex file in order to allow for electronic formatting and language editing. Figures and tables should be added at the end of the paper on separate sheets of paper.

Authors are kindly asked to comply with the following **format instructions**.

Acknowledgements can be given in a first footnote, marked by an asterisk.

Sections and subsections should be numbered consecutively in Arabic decimals, e.g.,

1. Introduction
2. The Model
2.1. The Household Decision
2.2. The Production Sector, etc.

Footnotes should be numbered consecutively in Arabic numerals.

Figures and tables should be numbered consecutively in Arabic numerals. A short heading in italics should summarize the contents of each figure and table, e.g.,

Figure 3
Earned Income Tax Credit for Different Household Types

Displayed **formulae** should be numbered consecutively throughout the manuscript (including the appendix) as (1), (2), etc. on the right-hand side of the page.

The **References** section should be the final section and contain all publications cited in the text or in footnotes, arranged alphabetically by the author's last name. Entries by the same author should be arranged by the year of publication, if necessary by adding lowercase letters, e.g., 1984a, 1984b, etc. References should appear as follows:

Periodicals:
Head, J. G. (1988), On Merit Wants, FinanzArchiv 46, 1–37.
Ballard, C. L., Shoven, J. B., and Whalley, J. (1985), General Equilibrium Computations of the Marginal Welfare Cost of Taxes in the United States, American Economic Review 75, 128–138.

Monographs:
Atkinson, A. B., and Stiglitz, J. E. (1980), Lectures on Public Economics, McGraw-Hill, London.
Stiglitz, J. E. (2000), Economics of the Public Sector, 3rd edition, W. W.Norton, New York, N.Y.

Contributions to handbooks, proceedings, etc.:
Sandmo, A. (1985), The Effects of Taxation on Savings and Risk Taking, in: Auerbach, A. J., and Feldstein, M. (Eds.), Handbook of Public Economics, vol. 1, North-Holland, Amsterdam, 265– 311.

Working Papers:
Eggert, W., and Kolmar, M. (2002), Contests with Size Effects, EPRU Working Paper 2002–04, Economic Policy Research Unit, University of Copenhagen.

Reference to publications in the text or in footnotes should be restricted to last name(s) of the author(s), year of publication, and, if necessary, page number(s), e.g.,
….as stated by Head (1988, p. 15)
….as stated by Head (1988, pp. 15–17).
….in another paper (Ballard, Shoven, and Whalley, 1985).
….(see Auerbach, 1985).
References to more than one article by the same author should be separated by commas or semicolons, e.g.,
….by Wildasin (1991, 1998)
….by Wildasin (1991, p. 15; 1998, p. 397)
The instructions to authors can be downloaded from www.mohr.de/finanzarchiv.

The Nexus of Corporate Income Taxation and Multinational Activity

Johannes Becker and Clemens Fuest*

Received 21 September 2009; in revised form 2 February 2012; accepted 7 February 2012

If a firm operates abroad, the host and the residence country have to decide how to divide the taxing rights among them. Firstly, the host country has to determine whether or not the firm has to file for income taxation at source. Secondly, the income has to be split between the two jurisdictions for tax purposes. These two decisions determine the extent of source- and residence-based taxation. We build a two-country model with costs of tax administration and compliance in order to analyze these two decision margins. We show that the globally optimal solution may imply a mix of source-based and residence-based taxation. Decentralized policies may attain the global optimum if specific transfer pricing rules are applied.

Keywords: corporate taxation, nexus threshold, multinational firms

JEL classification: H 25, F 23

1. Introduction

The Commentary on Article 7 of the OECD Model Tax Convention states:

When an enterprise of a Contracting State carries on business in the other Contracting State, the authorities of that second State have to ask themselves two questions before they levy tax on the profits of the enterprise: the first question is whether the enterprise has a permanent establishment in their country; if the answer is in the affirmative the second question is what, if any, are the profits on which that permanent establishment should pay tax. (Paragraph 1)

The answers to these two questions draw a dividing line between source- and residence-based taxation at two margins. Firstly, the government has to decide whether or not the activity of a particular firm has a nexus to the tax system of the host country. If so, the firm has to file for income taxation in

* Paper prepared for the conference on 'The Role of Firms in Tax Systems', April 2009 in Ann Arbor. We thank Alfons Weichenrieder and two anonymous referees as well as Mike Devereux, Wolfram Richter and participants at workshops and presentations in Berlin, Dortmund, Oxford and Vienna for valuable comments. The usual disclaimer applies. We gratefully acknowledge financial support from the ESRC (Grant No RES -060-25-0033).

FinanzArchiv 68 (2012), 231–251 doi: 10.1628/001522112X653822
ISSN 0015-2218 © 2012 Mohr Siebeck

this country. In the following, we refer to this margin as the *extensive margin*. Secondly, given that a multinational firm has to file for income taxation in more than one country, it has to be decided how the firm's overall income is divided between the source and the residence country. In the following, we refer to this margin as the *intensive margin*.

These margins are economically important because they determine the allocation of taxing rights and, hence, tax revenue, across countries. Moreover, they affect the nature and the magnitude of possible distortions of firm behavior caused by the tax system. While it seems plausible that governments agree on some sharing rule for the multinational firm's tax base at the intensive margin, the question arises why countries should decide not to tax certain firms at all. One answer to this question is based on the existence of tax administration and compliance cost: It may simply be too expensive to tax certain firms operating abroad.[1] The country of residence may deliberately abstain from taxing affiliates abroad if this saves administration costs or – if the government cares about firm income – compliance costs.

In this paper, we develop a model with tax administration and compliance costs, where the extent of residence- and source-based taxation along the extensive and the intensive margin is determined endogenously. We consider a two-country model where firms have a headquarters in one jurisdiction and an affiliate in the other jurisdiction. Firms differ in the distribution of real economic activity between the headquarters and the affiliate. Countries that host affiliates of foreign firms decide under which circumstances they want to tax these affiliates (extensive margin) and, if so, how to divide profits for purposes of taxation (intensive margin).

Our analysis yields the following results: Firstly, we determine conditions under which a mix of residence- and source-based taxation is optimal. The literature on the taxation of multinational firms usually takes the extent of residence- and source-based taxation as given. The question of how a mix of these principles may arise endogenously has received little attention. We argue that administration and compliance costs may favor residence-based taxation of foreign affiliates that are small in terms of real economic activity, but source-based taxation of larger affiliates.

Secondly, we show that if the source country is allowed to set the nexus threshold unilaterally, it may choose to tax more or fewer firms than in the optimum. The transfer-pricing scheme is crucial in aligning national with global efficiency. Then, decentralized threshold setting may lead to the global optimum.

1 Alternative motivations for choosing a mix of these tax principles are discussed in section 5.

Thirdly, starting from the global optimum, no country has an incentive to unilaterally deviate in the direction of double taxation. However, we show that residence countries have an incentive to exempt some of their affiliates from tax and, thus, to allow for "white income."

Our analysis is related to various strands of the literature on international taxation. These are reviewed in the next section, where we also give a brief sketch of legal rules on nexus thresholds and transfer pricing between countries and U.S. states. Sections 3 and 4 present the model and the results. Section 5 discusses the policy implications of our results and concludes.

2. Previous Literature and Legal Rules

In this section, we relate the central points of our paper to previous literature and provide a brief overview of jurisdiction relevant for nexus thresholds and transfer pricing.

2.1. Source- or Residence-Based Taxation?

If all income is taxed where the ultimate investor resides, taxation leaves unaffected the choice of *where* to invest. This result is known as capital export neutrality. In contrast, if all income is taxed at source, taxation does not distort the choice of *who* invests (capital import neutrality). As is well known, it follows directly from the production-efficiency theorem (Diamond and Mirrlees, 1971) that capital export neutrality is superior in its efficiency properties in a world where production factors are mobile across borders. The superiority of residence-based taxation has become one of the most powerful theorems in the literature on international taxation. The fact that source-based taxes are nevertheless frequently used is generally explained by difficulties in enforcing residence-based taxes, the backstop function of some source-based taxes, and the existence of pure profits (see Mintz, 1996).

More recently, this line of argument has been questioned by Desai and Hines (2003, 2004). Using the concept of ownership neutrality, established by Devereux (1990), the authors argue that the efficiency properties of source-based taxation dominate those of residence-based taxation in a world where investment mainly takes the form of reallocation of ownership rights. According to the authors, countries sticking to the tax credit system (which, under some assumptions, leads to capital export neutrality) should rather switch to exemption, i.e., source-based taxation.[2]

2 For a formal analysis of this issue, see Becker and Fuest (2010). In Becker and Fuest (2011), we analyze the efficiency consequences of both tax principles with international investment taking the form of mergers and acquisitions. We show that, under certain

Given the large number of contributions that highlight the differences between the two taxation principles, it is surprising that only few papers are concerned with the question of what makes governments choose a mix of them. As indicated in the introductory section, such a mix occurs if the two governments under consideration choose to share taxing rights at the extensive or at the intensive margin or both. In fact, real-world governments do so at both margins. An obvious reason for this may be that it is considered as a matter of fairness that a host country should participate in the tax revenue from income that is partly generated within its borders. This *entitlement approach* is different from the efficiency approach on which we focus in this paper. Precisely, we ask three questions. Firstly, apart from fairness, is there an efficiency argument for sharing taxing rights? Secondly, if so, what are the optimal sharing rules? Thirdly, how can the optimum be attained? Efficiency issues may arise at the extensive and the intensive margin. The related literature and the relevant tax law will be discussed in the following.

2.2. The Division of Taxing Rights and Nexus Rules

As mentioned above, there are only few papers dealing with the allocation of taxing rights across jurisdictions. In national tax law and double-taxation agreements (DTAs), the right of a jurisdiction to tax a firm is associated with the notion of a "fixed place of business" or "permanent establishment." For example, article 7.1 of the DTA between Germany and the United States says: "The business profits of an enterprise of a Contracting State shall be taxable only in that State unless the enterprise carries on business in the other Contracting State through a permanent establishment situated therein." Article 5.1 defines that "the term 'permanent establishment' means a fixed place of business through which the business of an enterprise is wholly or partly carried on." Paragraphs 2 to 5 include further definitions according to which permanence is assumed if various – easily observable – conditions are satisfied. These conditions have a physical dimension in that the place of activity has to be a fixed structure, a time dimension in that the activity has to last more than twelve months, and an organizational dimension in that someone who has the right to conclude contracts represents a sufficient condition for permanence.

In the absence of DTAs, most countries adopt rules similar to those expressed in the DTAs. However, as Davies (2004) reports, "[e]ven within treaties, there is significant variation in definitions.[...] The idea of permanence is [...] open for debate." A study by the United Nations (1998) reports substantial variation in the definition of "permanent establishment," which

conditions, source-based taxation does not distort international M&A activity, whereas residence-based taxes effectively do so.

demonstrates that governments have some discretion about determining nexus thresholds.

§ 381 of U.S. trade law[3] specifies under which circumstances U.S. states have the power to levy income taxes on interstate activity. According to the law, no state shall have the power to levy income taxes if the only business activity is "the solicitation of orders by such person, or his representative, in such State for sales of tangible personal property, which orders are sent outside the State for approval or rejection, and, if approved, are filled by shipment or delivery from a point outside the State." This de minimis rule has been relevant in the Supreme Court decisions on Illinois versus Bellas Hess, North Dakota versus Quill[4], and Wisconsin versus Wrigley[5]. In all of these three cases, U.S. states levied income taxes and the firms argued that they were below the nexus threshold. To support this, they presented evidence on quantitative indicators of economic activity.[6]

When electronic commerce became important some ten years ago, the issues of interstate commerce and nexus were raised again. With substantially reduced cost of transaction and transport, physical presence is not crucial for trade and business anymore. The implications for taxation are discussed by McLure (2000, 2002): Among other things, he argues that the legal institute of "permanent establishment" should be complemented (see McLure, 2000): "[l]iability of income tax should be subject to a de minimus test; because of the compliance costs involved, it would not make sense to levy income tax on all vendors that have an economic presence, no matter how small their sales" (p. 13). Therefore, he suggests a de minimis test on a gross basis.

2.3. The Allocation of Profits and Transfer Prices

Most of the literature on multinational investment and taxation assumes that there are some naturally given locational profits. These "true" levels of locational profits are assumed to be those that would emerge in the absence of taxation or tax differentials, and tax authorities seek methods to gain information about "true" profits, e.g., arm's-length pricing; see Devereux and

3 http://www.law.cornell.edu/uscode/15/381.html.

4 http://www2.law.cornell.edu/supct/html/91-0194.ZO.html.

5 http://www.law.cornell.edu/supct/html/91-119.ZO.html.

6 In the Quill case, the State of North Dakota argued that Quill had the status of a retailer, of which the definition includes "three or more advertisements within a 12 month period." In the Wrigley case, the court's decision includes a thorough quantification of activity: "The sales or 'field' representatives in the Milwaukee region, each of whom was assigned his own territory, resided in Wisconsin. They were provided with company cars, but not with offices. They were also furnished a stock of gum (with an average wholesale value of about $1000), a supply of display racks, and promotional literature. These materials were kept at home, except that one salesman, whose apartment was too small, rented storage space at about $25 per month, for which he was reimbursed by Wrigley."

Keuschnigg (2009). However, sometimes headquarters and affiliate production are complementary, so that locational profits are hard to determine. As the OECD (2006) Report on the Attribution of Profits to Permanent Establishments puts it: "The hypothesis by which a PE is treated as a functionally distinct and separate enterprise is a mere fiction necessary for purposes of determining the business profits of this part of the enterprise under Article 7." (p. 13)

As mentioned above, it seems a matter of fairness and international equity that a sufficiently high level of activity entitles countries to levy taxes on the related income. It is noteworthy, though, that from an efficiency standpoint this is less clear. To illustrate this, assume that tax rates are the same in both locations. Then, a change in the location of profits leaves firm income unaffected, provided that there are no effects on compliance costs. The only effect is a redistribution of government funds across countries. In such a situation, there is *a priori* no efficiency-guided recommendation for the allocation of profits across locations. However, in this paper we provide an efficiency argument for the distribution of profits, which connects the decisions at the intensive and the extensive margin.

How does the allocation of profits look in real-world tax systems? The most important technique is the arm's-length principle, according to which the tax authority attributes to an entity within its borders profits that this entity "would have earned at arm's length if it were a legally distinct and separate enterprise performing the same or similar functions under the same or similar conditions" (OECD 2006, p. 12). Arm's-length pricing is mainly analyzed in the literature as a means to prevent firms from engaging in tax-induced profit shifting. Indeed, there exist a large number of studies providing evidence that firms use various accounting techniques to shift profits from high-tax to low-tax jurisdictions; see, e.g., Hines and Rice (1994) and Huizinga and Laeven (2008). When the conflicting interests of firms and tax authorities are settled by courts, it becomes clear that governments effectively define some kind of transfer price that limits the extent to which firms may shift profits.[7] For example, the British–American enterprise GlaxoSmithKline was forced in 2006 by the U.S. Internal Revenue Service (IRS) to recalculate its U.S. tax liabilities and pay an additional $3.4 billion; see Nutt (2006). In detail, the payments of royalties were limited as well as the purchasing prices for some necessary input products. Thus, the IRS considered the input provided by the U.K. headquarters (and other affiliates abroad) as less than indicated by GlaxoSmithKline.

7 Clearly, transfer pricing also raises conflicts of interest between tax authorities of different countries. The strategic transfer-pricing incentives for countries are studied in Mansori and Weichenrieder (2001).

2.4. The Cost of Tax Administration and Compliance

This paper is part of a recent literature that seeks to bridge the gap between puristic optimal-tax-policy models and actual real-world tax systems. A main finding in this context is that firms play a central role in modern tax systems; see Bird (2002). An obvious explanation for this is that tax collection is subject to some economies of scale: since tax subjects gather in firms, it is efficient to collect taxes there.[8] In the presence of tax administration and compliance cost, the tax-collecting authority may want to exempt some taxpayers from tax, if the expected revenue falls short of the cost involved. Keen and Mintz (2004) derive an optimal threshold for the value-added tax. Dharmapala, Slemrod, and Wilson (2011) consider optimal output taxes in the presence of administrative costs and build a model that replicates the "missing middle," i.e., the missing of firms of intermediate size, an empirical phenomenon often observed in developing countries. Our paper builds on these contributions, although there are some crucial differences.

The cost of tax compliance and administration is economically important. The European Commission (2004) estimates that large firms on average bear compliance costs of more than 1.4 million euros, or 1.9% of their total tax payments.[9] For small firms, average compliance costs are estimated to equal 200,000 euros, or 30.9% of their tax payments. Thus, compliance costs are effectively regressive. These estimations are in line with the results found by Slemrod and Blumenthal (1996) for large U.S. firms.[10] Most importantly for the purpose of this study, the European Commission (2004) finds that compliance costs increase by more than 400% when the firm is active in more than one jurisdiction. The fact that multinational activity is accompanied by a strong increase in tax compliance costs is often interpreted as the main driving force behind the proposals for a common consolidated tax base or home-state taxation; see European Commission (2001). There is less evidence for tax administration costs. Some studies suggest that administration costs are somewhat lower than compliance costs (see Slemrod and Blumen-

8 In such a setting, Kopczuk and Slemrod (2006) show that the production-efficiency theorem derived in the seminal paper by Diamond and Mirrlees (1971) breaks down if tax-specific administration costs are considered. Slemrod (2008) demonstrates that another textbook theorem is put into question if administrative costs are taken into account: the irrelevance of who remits taxes.

9 Large firms are those with more than 250 employees; small firms, those with fewer.

10 Evans (2003) summarizes a survey on studies measuring compliance and tax administration costs as follows: "Compliance costs are highly significant for the main central government taxes [...]. They are high however measured – whether in absolute money terms or relative to tax yield, GDP or administrative costs. For example, the studies suggest that compliance costs of such taxes are typically anywhere between 2% and 10% of the revenue yield from those taxes; up to 2.5% of GDP; and usually a multiple (of between two and six) of administrative costs."

thal, 1996, and Evans, 2003), but especially estimates of differences in these costs between auditing a purely national firm and a multinational firm are not available.[11]

3. The Model Setup

Consider a world with two countries each of which is populated by a representative household and a continuum of multinational firms.

3.1. Households

The representative household receives utility from a private consumption good c with a price of unity and a publicly provided good h according to the utility function

$$u = c + \lambda h\,, \tag{1}$$

where $\lambda > 1$ indicates the marginal utility of public-good consumption.

The only income source is dividends from firms owned by the representative household. There is no direct tax on household income.

3.2. Firms

Each multinational consists of two entities, a headquarters and an affiliate in the other country. Each firm's output is sold at the affiliate level. For presentational ease, we will refer to the headquarters location as the residence country and the affiliate location as the source country. Of course, both countries in our model are simultaneously residence country and source country. Accordingly, when we consider the welfare of a country we have to take account of the welfare resulting from being a residence country (we will call this the residence country's welfare) and the welfare from being a source country (the source country's welfare).

Firms differ in their exogenously given organizational structure $\sigma \in [0, 1]$, which summarizes the distribution of the firm's real economic activity (employees, fixed assets) across the residence country and the source country. To keep notation simple we assume a uniform distribution of firms across σ and normalize the density function to one. Thus, each multinational firm in our model consists of a headquarters with an activity level of $1 - \sigma$ located in the residence country and an affiliate with activity level σ located in the source country. Pretax profits for all firms are given by Q.

Assume that there is no market outside the firm for the services provided by the headquarters. Then, there is no objective way of pricing these services

11 See also Slemrod and Venkatesh (2002).

in order to divide Q into a part generated domestically and another part generated abroad. From the viewpoint of the firm, there is no need for this attribution of profits to headquarters and affiliate locations. However, if the firm has to file for income taxation in both jurisdictions, such a division becomes necessary. Then, a transfer price p applies, which effectively attributes an income of $Q - p$ to the affiliate and an income of p to the headquarters. The transfer price is set by the government and cannot be manipulated by the firm.

The multinational firm pays corporate income taxes at the headquarters and, potentially, at the affiliate level. Taxation involves a compliance cost C. For notational convenience, we normalize the compliance of residence-based taxes levied on headquarters profits to zero. The taxation of affiliate profits gives rise to a compliance cost of $C^r(\sigma)$ if affiliate profits are taxed on a residence basis (superscript r), and $C^s(\sigma)$ if they are taxed at source (superscript s). We assume that compliance costs increase in the level of real economic activity: $C^{r\prime}, C^{s\prime} \geq 0$ with $C^{r\prime\prime}, C^{s\prime\prime} \lesseqgtr 0$.

An important assumption is that we allow compliance costs to differ between residence- and source-based taxation. Dealing with two administrations may be more difficult than dealing just with one. If this effect dominates, the compliance cost of source-based taxation would be higher than that of residence-based taxation. But if the affiliate has its own tax department, it may also find it easier to deal with local tax authorities and tax laws.[12] Finally, we assume that it does not matter for the level of compliance costs whether a high or a low transfer price is chosen to divide the tax base between the residence and the source country. This seems justified because transfer-price variations are unlikely to change the administrative burden of complying with the tax law.

3.3. Government

Governments provide the public good, which is financed by corporate taxes. Corporate tax rates are exogenously given by t^r and t^s in the residence country and the source country, respectively.[13]

12 One might argue that economic activity of a firm in a country always involves dealing with the tax authorities of that country, irrespective of whether or not a firm ends up filing for income taxation. This would imply that compliance costs in our model would represent the incremental costs that arise if firms, after screening by tax authorities, do have to file for income taxation.

13 This assumption is mainly made for keeping the analysis tractable. However, it can be justified as follows. Nexus rules concern relatively small levels of activity, maybe too small to affect general corporate tax policies. In addition, nexus rules may differ by industry, which makes it even less probable that there is a substantial interaction with corporate tax rate setting.

We assume that all firms have to file for income taxation in their residence country. However, they may or may not be obliged to do so in the source country, i.e., in technical terms, they may or may not have nexus to source-based taxation (*extensive margin*). If the firm does not have nexus, all income is taxed by the residence country. If the firm has nexus and thus has to file for income taxation in the source country, the firm's profits are divided into an affiliate profit and a headquarters profit (*intensive margin*).

Consider firstly the intensive margin. As noted above, the only reason for defining a transfer price is that the firm is taxed by two jurisdictions. This raises the question of who determines the transfer price. In the absence of coordination, the two jurisdictions may set their own transfer prices $\bar{p}^r(\sigma)$ and $\bar{p}^s(\sigma)$ for a firm with structure σ. Note that, if $\bar{p}^r(\sigma)$ and $\bar{p}^s(\sigma)$ differ, double taxation $[\bar{p}^r(\sigma) > \bar{p}^s(\sigma)]$ or white income $[\bar{p}^r(\sigma) < \bar{p}^s(\sigma)]$ may occur.

The extensive margin is determined as follows. Both countries may choose to tax all affiliates or none, or to relate this choice to the affiliate activity level σ. In the last case, both governments choose thresholds for the affiliate level of economic activity, denoted by $\bar{\sigma}^r$ and $\bar{\sigma}^s$. For any firm characterized by $\sigma \geq \bar{\sigma}^r$, the residence country exempts the affiliate's income from taxation.[14] Accordingly, if $\sigma \geq \bar{\sigma}^s$ ($\sigma < \bar{\sigma}^s$), the affiliate's profit is subject to (exempt from) taxation in the source country. As in the case of the intensive margin, double taxation ($\bar{\sigma}^r > \bar{\sigma}^s$) or white income ($\bar{\sigma}^r < \bar{\sigma}^s$) may occur.

In correspondence to the firms' compliance cost, corporate taxation involves an administration cost A, which is assumed to equal zero for the taxation of residence-based income (i.e., headquarters income). Again, administration costs differ between residence- and source-based taxation of affiliate income: $A^r(\sigma)$ and $A^s(\sigma)$, with $A^{r\prime}, A^{s\prime} \geq 0$ and $A^{r\prime\prime}, A^{s\prime\prime} \lesseqgtr 0$. On the one hand, the residence country may have some administrative advantage in taxing the affiliate, compared to the source country, since it taxes the headquarters anyway. On the other hand, due to mere geographical distance, language barriers, reduced enforcement rights, and many other reasons, taxing economic activity abroad is more difficult than taxing domestic activity. For this reason, source-based taxation may be administratively less costly than residence-based taxation in some cases, in particular as the magnitude of the affiliate's activity σ increases. Finally, we assume again that administration costs are independent of the level of transfer prices. Thus, in our model, changes at the extensive margin matter for administration and compliance costs, whereas changes at the intensive margin do not.

14 It is not necessarily the case that countries will want to tax firms above a given nexus threshold if they discriminate on the basis of σ. We will come back to this issue below.

3.4. Firm Profits and Tax Revenue under Coordinated Thresholds and Transfer Prices

In what follows, we assume that residence and source country agree upon one nexus threshold $\bar{\sigma}$ and a transfer-pricing scheme $\bar{p}(\sigma)$, which maps transfer prices upon firm structure σ. Deviations from the coordinated policy parameter and uncoordinated transfer pricing and nexus rules are analyzed in section 5. In the following, we will omit the argument and write $\bar{p}$ instead of $\bar{p}(\sigma)$, C^r instead of $C^r(\sigma)$, etc., unless misunderstanding might arise.

A multinational firm with a structure σ has after-tax profits Π of

$$\Pi(\sigma) = \begin{cases} (Q - C^r)(1 - t^r) & \text{if } \sigma < \bar{\sigma}\,, \\ \bar{p}(1 - t^r) + (Q - \bar{p} - C^s)(1 - t^s) & \text{if } \sigma \geq \bar{\sigma}\,. \end{cases} \tag{2}$$

As long as the level of economic activity of the affiliate, σ, is below the nexus threshold $\bar{\sigma}$, the firm's entire income is taxed by the residence country, and the firm faces no compliance costs related to source country taxation. Firms with affiliates above the nexus threshold, in contrast, do have to file for income taxation in the source country and apply the transfer-pricing rule $\bar{p}(\sigma)$ set by the tax authorities to divide their income for purposes of taxation into affiliate profits $Q - \bar{p} - C^s$ and headquarters profits $\bar{p}$.

Accordingly, the tax revenue of the residence country net of administration cost is given by

$$T^r(\sigma) = \begin{cases} t^r(Q - C^r) - A^r & \text{if } \sigma < \bar{\sigma}\,, \\ t^r\bar{p} & \text{if } \sigma \geq \bar{\sigma}\,. \end{cases} \tag{3}$$

and the tax revenue of the source country net of administration cost by

$$T^s(\sigma) = \begin{cases} 0 & \text{if } \sigma < \bar{\sigma}\,, \\ t^s(Q - \bar{p} - C^s) - A^s & \text{if } \sigma \geq \bar{\sigma}\,. \end{cases} \tag{4}$$

4. Welfare and Optimal Policy Choices under Coordinated Thresholds and Transfer Prices

In our model, both countries are simultaneously the source and the residence countries of multinational firms. Because of this symmetry, we can focus on the welfare resulting from being a residence location and from being a source location. Residence countries have firm profit income and tax revenue as sources of welfare, and source countries have only tax revenue (from firms having nexus to corporate income taxation).

4.1. Welfare

We assume that governments are benevolent and define welfare W as the representative household's utility given by $u = c + \lambda h \equiv W$ with $c = \Pi$ and $h = T^r + T^s$. Accordingly, the welfare from being a residence country equals $W^r = \Pi + \lambda^r T^r$, and the welfare from being a source country $W^s = \lambda^s T^s$. We differentiate between λ^r and λ^s for the purpose of comparison between the two countries.

Under coordinated thresholds and transfer prices, the welfare of the residence country is given by

$$W^r = \int_0^{\bar{\sigma}} [(Q - C^r)(1 - t^r) + \lambda^r (t^r (Q - C^r) - A^r)] d\sigma + \int_{\bar{\sigma}}^1 [\bar{p}(1 - t^r) + (Q - \bar{p} - C^s)(1 - t^s) + \lambda^r t^r \bar{p}] d\sigma . \quad (5)$$

The first term on the right-hand side is the income and tax revenue generated by firms whose affiliates operate below the nexus threshold, so that they do not file for income taxation abroad. The second term is the profit and tax revenue of firms operating above that threshold.

The welfare of the source country equals

$$W^s = \int_{\bar{\sigma}}^1 \lambda^s [t^s (Q - \bar{p} - C^s) - A^s] d\sigma . \quad (6)$$

4.2. Globally Optimal Policies

What do globally optimal policies look like? Global welfare is given by $W^g = W^r + W^s$. Consider firstly the globally optimal transfer pricing scheme $\bar{p}(\sigma)$ for firms with nexus in both countries. For this purpose, it is helpful to assume that $\bar{p}(\sigma)$ consists of a fixed part, $\hat{p}$, and a part depending on σ, denoted by $f(\sigma)$: $\bar{p}(\sigma) = \hat{p} + f(\sigma)$, where $\hat{p}$ may well be zero. Now, assume that a given transfer-pricing scheme $\bar{p}(\sigma)$ is marginally increased by a small amount at each point σ of the scheme. In other words, $\hat{p}$ is marginally increased. The global welfare effects of such an increase are given by

$$\frac{\partial W^g}{\partial \hat{p}} = \int_{\bar{\sigma}}^1 [\nu^r - \nu^s] d\sigma , \quad (7)$$

where $\nu^i = 1 + (\lambda^i - 1) t^i$ for $i \in \{r, s\}$ denotes the social valuation of profit income in the residence and the source country. A change in the transfer price scheme only redistributes taxing rights. The impact of this redistribution on global welfare depends on the tax rates and the marginal utility of public consumption in the two countries. Clearly, (7) does not contribute much to explaining the allocation of taxing rights in existing tax systems. In particular, from a global perspective, there is no particular reason in our model to divide

taxing rights according to the distribution of real economic activity across the two countries. We will return to this issue when we discuss decentralized policies.

Consider next the global welfare effect of setting the nexus threshold. The effect of a marginal change in $\bar{\sigma}$ on global welfare, given the transfer price $\bar{p}$, can be expressed as

$$\frac{\partial W^g}{\partial \bar{\sigma}} = \Omega(\bar{\sigma}) + S(\bar{\sigma}) - R(\bar{\sigma}), \tag{8}$$

where

$$\Omega(\bar{\sigma}) = (Q - \bar{p}(\bar{\sigma}))(\nu^r - \nu^s), \tag{9}$$

$$S(\bar{\sigma}) = \nu^s C^s(\bar{\sigma}) + \lambda^s A^s(\bar{\sigma}), \tag{10}$$

$$R(\bar{\sigma}) = \nu^r C^r(\bar{\sigma}) + \lambda^r A^r(\bar{\sigma}). \tag{11}$$

Ω captures the welfare effect of moving the right to tax a given base, $Q - \bar{p}(\bar{\sigma})$, from source- to residence-based taxation, which, under symmetry, equals zero. $S(\bar{\sigma})$ is the social cost of compliance and administration of a source-based tax on an affiliate of type $\bar{\sigma}$. Accordingly, $R(\bar{\sigma})$ is the social compliance and administration cost of taxing the same affiliate on a residence basis.

From a policy perspective, a key question is whether the coexistence of source- and residence-based taxation can be optimal from a global perspective. This would require the first-order condition in (8) to equal zero and the second-order condition to satisfy $\frac{\partial^2 W^g}{\partial(\bar{\sigma})^2} = -\bar{p}'(\bar{\sigma})(\nu^r - \nu^s) + S'(\bar{\sigma}) - R'(\bar{\sigma}) < 0$.

A possible interior optimum is illustrated by figure 1, where equal tax rates and preferences are assumed: $\nu^r = \nu^s$. Such an equilibrium would imply that, for low real economic activity at the affiliate level, residence-based taxation would be optimal, because the additional compliance and administration cost of filing for income taxation in the source country would be too high. But if real economic activity increases at the affiliate, taxation becomes more difficult. As mentioned in the preceding section, it may in particular become very costly for the tax authorities to audit the activity of an affiliate located abroad. This disadvantage may be so significant that it may be optimal from a global perspective to switch to source-based taxation.

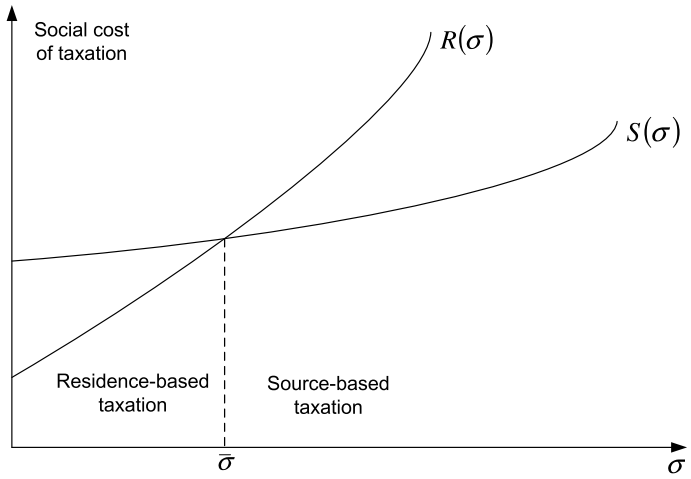
The above-presented results may be summarized as

Proposition 1 Global Optimality: Let $C^{r\prime}, C^{s\prime}, A^{r\prime}, A^{s\prime} \geq 0$, and suppose transfer prices do not affect tax administration or compliance costs. Then,

a. Transfer prices: a change in transfer prices only affects global welfare if countries differ in tax rates or preferences for the public good, and

b. Thresholds: (i) if $\frac{\partial W^g}{\partial \bar{\sigma}} > 0$ for all $\bar{\sigma}$, the globally optimal regime is pure residence-based taxation; (ii) if $\frac{\partial W^g}{\partial \bar{\sigma}} < 0$ for all $\bar{\sigma}$, the globally optimal regime is pure source-based taxation; (iii) if $\frac{\partial W^g}{\partial \bar{\sigma}} = 0$ and $\frac{\partial^2 W^g}{\partial(\bar{\sigma})^2} < 0$ hold, the optimal regime is a mix of source- and residence-based taxation.

Figure 1
Global Optimum



Our model thus provides a rationale of why existing tax systems foresee that affiliates of multinational firms will file for income taxation in their host country if their economic activity exceeds a certain threshold. The question arises whether the global optimum is stable. We therefore turn to the perspective of the individual countries.

4.3. Decentralized Policies

The global optimum is self-enforcing if individual choices – under the restriction that there are coordinated transfer prices and thresholds – are in line with such an outcome.

With respect to transfer prices, this actually will never be the case. As noted above, a change in transfer prices is simply a redistribution of taxing rights. If $\bar{p} \in [0, Q]$, the source country would always like to set $\bar{p} = 0$ for all σ, while the residence country would set $\bar{p} = Q$ for all σ. Note that this is true independent of the actual firm characteristic σ or the threshold level $\bar{\sigma}$, because transfer prices (by definition) do not affect administration and compliance costs.

In order to prevent the analysis of decentralized nexus threshold policies from depending on extreme values of $\bar{p}$ (0 or Q), we assume that there is a transfer-pricing scheme $\bar{p} = \bar{p}(\sigma)$ that both countries agree upon. Since we do not *a priori* impose restrictions on the functional form of $\bar{p}(\sigma)$, this assumption allows for the extreme values of $\bar{p}$, but also includes more intuitive transfer-pricing functions, where the transfer price $\bar{p}$ would decrease with

increasing amount of economic activity located in the source country, i.e., $\bar{p}'(\sigma) < 0$.

How do individual countries set the nexus thresholds, given a transfer pricing scheme? Assume that the source country has the right to set the threshold and the residence country accepts this.[15] Given this, the optimal threshold from a national point of view is set where the marginal firm's tax payments just cover the tax administration cost A^s:

$$\frac{\partial W^s}{\partial \bar{\sigma}} = -\lambda^s[t^s(Q - \bar{p}(\bar{\sigma}) - C^s(\bar{\sigma})) - A^s(\bar{\sigma})] = 0\,. \tag{12}$$

For (12) to give a welfare maximum, the second-order condition would have to satisfy $\frac{\partial^2 W^s}{\partial \bar{\sigma}^{s2}} < 0$, which is given by

$$\frac{\partial^2 W^s}{\partial \bar{\sigma}^2} = \lambda^s[t^s(\bar{p}' + C^{s\prime}) + A^{s\prime}] < 0\,. \tag{13}$$

Here, two general cases may occur. Firstly, if $\bar{p}' \geq -(C^{s\prime} + A^{s\prime}/t^s)$ for all σ (including the case of $\bar{p}' = 0$), no interior optimum is found. Then, the source country will either tax all affiliates operating within its borders or no affiliate at all (because tax administration costs exceed tax revenue for all affiliates). In this case, the global optimum described by $\frac{\partial W^g}{\partial \bar{\sigma}} = 0$ and $\frac{\partial^2 W^g}{\partial (\bar{\sigma})^2} < 0$ (see above) is not compatible with the source country's national interest.

Secondly, if (12) and (13) hold, which implies $\bar{p}' < -(C^{s\prime} + A^{s\prime}/t^s)$, an interior optimum exists. Then, the source country chooses a level of $\bar{\sigma}$ above which affiliates from abroad are taxed and below which such affiliates are exempt from tax. A negative first derivative, $\bar{p}' \leq 0$, means that the larger the fraction of activity at the affiliate and the smaller the fraction of activity at the headquarters, the smaller the fraction of income taxed at the headquarters level, which seems to be plausible as an assumption.

For the level of $\bar{\sigma}$ derived in (12) and (13) to be a global optimum, the externality on the residence country is required to be zero. The externality is given by

$$\frac{\partial W^r}{\partial \bar{\sigma}} = (Q - \bar{p})(t^s + (\lambda^r - 1)t^r) + C^s(1 - t^s) - \nu^r C^r - \lambda^r A^r\,. \tag{14}$$

The first term on the right-hand side of (14) is positive. It reflects that the residence country raises more revenue as the source country increases the threshold. If $t^r > t^s$, after-tax profits accruing to firm owners may decline, but the additional tax revenue compensates for this. In the absence of compliance and administration costs, the residence country would unambiguously benefit

15 This seems to be a plausible assumption, since, as Graetz and Oosterhuis (2001) state, "The OECD nations have all conceded that the country of source – the nation where the income is earned – enjoys the primary right to tax active business income, with the residence country – the nation where the business is incorporated or managed – retaining at most a residual right to tax such income."

from an increase in the threshold. But if there are such costs, things may be different. The change in compliance cost ($C^s(1-t^s)-\nu^r C^r$) may have a positive or negative effect. The increase in tax administration costs in response to an increase in the threshold has an unambiguously negative effect.

Thus, due to the existence of compliance and administration costs, it may be the case that an increase in the threshold set by the source country reduces the welfare of the residence country. This implies that, from a global perspective, there may be too much or too little source-based taxation.

As the externality depends crucially on the transfer-pricing scheme, $\bar{p}(\sigma)$ may be used to reduce the externality to zero. In fact, appropriate transfer pricing rules can ensure that uncoordinated threshold setting leads to globally optimal thresholds and that fiscal externalities of threshold setting are fully internalized. This is achieved if the right-hand side of (14) equals zero, which requires

$$\bar{p}^*(\sigma) = Q + \frac{C^s(\sigma)(1-t^s) - \nu^r C^r(\sigma) - \lambda^r A^r(\sigma)}{t^s + (\lambda^r - 1)t^r}. \tag{15}$$

The transfer-pricing scheme $\bar{p}^*(\sigma)$ ensures that the externality of the source country setting the threshold, (14), equals zero. Its slope is given by $\bar{p}^{*\prime} = \frac{C^{s\prime}(1-t^s)-\nu^r C^{r\prime}-\lambda^r A^{r\prime}}{t^s+(\lambda^r-1)t^r}$. Recall that the transfer price has to satisfy $\bar{p}' < -(C^{s\prime} + A^{s\prime}/t^s)$ for (12) to be a welfare maximum. Using the first derivative of (15) as well as (10) and (11), it can be shown that this condition holds if

$$S'(\bar{\sigma}) - R'(\bar{\sigma}) < \left(C^{s\prime} + \frac{A^{s\prime}}{t^s}\right)(\nu^s - \nu^r). \tag{16}$$

The right-hand side of the above equation is zero if tax rates and preferences λ are equal across countries. Then, existence of a global optimum requires $S'(\bar{\sigma}^s) - R'(\bar{\sigma}^s) < 0$. We summarize these results as follows:

Proposition 2 Decentralized policies with coordinated transfer pricing and uncoordinated threshold policies: Assume that a global optimum with $0 < \bar{\sigma} < 1$ exists.

a. The source country chooses a nexus threshold $0 < \bar{\sigma}^s < 1$ only if (12) and (13) hold. This requires the transfer price to depend negatively on σ: $\bar{p}' < -(C^{s\prime} + A^{s\prime}/t^s)$.

b. Depending on compliance and administration cost functions, the level of source-based taxation preferred by the source country may be higher or lower than globally optimal.

c. Choosing a specific transfer price scheme as in (15) may align national and global efficiency considerations. If countries agree on a transfer-pricing function as in (15) before thresholds are set, decentralized threshold setting leads to a global optimum.

5. Incentives to Deviate from Coordinated Thresholds and Transfer Prices

In the previous section, we have shown under which circumstances the residence country prefers a higher or lower (coordinated) threshold than the one chosen by the source country. In this section, we ask whether the country has the incentive to unilaterally deviate from the source country's choice of nexus threshold, i.e., to abandon the coordinated threshold and to set its own threshold. We analyze this under the assumption that deviation by one country does not affect the behavior of the other country.[16]

We will focus on threshold setting only because it is straightforward to show that countries always have incentives to deviate from coordinated transfer prices and to move to double taxation because this increases their tax revenue at no cost. The reason is that double taxation at the intensive margin does not affect compliance or administration costs and, by assumption, does not trigger any retaliation. Consequently, countries never have incentives to allow for white income.

5.1. Double Taxation

Consider firstly the issue of double taxation, with $\bar{\sigma}^r > \bar{\sigma}^s$. We assume that source-based taxes are at least deductible from the tax base in the residence country. In the presence of double taxation (dt), the residence country's welfare[17] is given by

$$\begin{aligned} W^r_{dt} = &+ \int_0^{\bar{\sigma}^s} [(Q - C^r)(1 - t^r) + \lambda^r(t^r(Q - C^r) - A^r)]d\sigma \\ &+ \int_{\bar{\sigma}^s}^{\bar{\sigma}^r} [\nu^r(\bar{p} - C^r + (Q - \bar{p} - C^s)(1 - t^s)) - \lambda^r A^r]d\sigma \\ &+ \int_{\bar{\sigma}^r}^{1} [\bar{p}\nu^r + (Q - \bar{p} - C^s)(1 - t^s)]d\sigma\,. \end{aligned} \tag{17}$$

The effect of a small change in the residence-based threshold $\bar{\sigma}^r$ on the source country is zero, whereas on the residence country it is given by

$$\frac{\partial W^r_{dt}}{\partial \bar{\sigma}^r} = t^r(1 - t^s)(\lambda^r - 1)(Q - \bar{p} - C^s) - \nu^r C^r - \lambda^r A^r\,. \tag{18}$$

Increasing the threshold $\bar{\sigma}^r$ above $\bar{\sigma}^s$ increases the tax revenue (first term) and the social cost of tax compliance and administration (second and third

16 It is well known that, with retaliation possibilities, cooperative equilibria of the type analyzed in the preceding sections may emerge.

17 The source country's welfare is – as before – equal to

$$W^s_{dt} = \int_{\bar{\sigma}^s}^{1} \lambda^s[t^s(Q - \bar{p} - C^s) - A^s]d\sigma\,.$$

terms). Thus, the effect of an increase in double taxation on the residence country's welfare is ambiguous. If, however, the threshold set by the source country is globally optimal, i.e., (14) equals zero, the above equation can be rewritten as

$$\frac{\partial W^r_{dt}}{\partial \bar{\sigma}^r} = -t^s \nu^r (Q - \bar{p}) - \nu^r C^s (1 - t^s) < 0\,. \tag{19}$$

Thus, we can state

Proposition 3 If the source-based threshold is at or above the globally optimal threshold level, the residence country has no interest in unilaterally deviating from the threshold set by the source country by moving to double taxation.

This can be explained as follows. The global optimum implies that, if the *coordinated* threshold is marginally increased, the additional tax revenue equals the additional administration and compliance costs. If however the threshold is *unilaterally* increased, the additional tax revenue is lower, because foreign taxes are deducted while additional administration and compliance costs remain the same. Therefore, the welfare effect is negative in sum.

5.2. White Income

Are there incentives for residence countries to deviate in the other direction, i.e., to withdraw from taxing affiliates although the source country does not tax these affiliates ($\bar{\sigma}^r < \bar{\sigma}^s$)? In the presence of white income (*wi*) the residence country's welfare is given by

$$\begin{aligned} W^r_{wi} = &\int_0^{\bar{\sigma}^r} [(Q - C^r)\nu^r - \lambda^r A^r]d\sigma + \int_{\bar{\sigma}^r}^{\bar{\sigma}^s} [\nu^r \bar{p} + (Q - \bar{p})]d\sigma \\ &+ \int_{\bar{\sigma}^s}^{1} [\nu^r \bar{p} + (Q - \bar{p} - C^s)(1 - t^s)]d\sigma\,. \end{aligned} \tag{20}$$

The welfare effect of a unilateral change in $\bar{\sigma}^r$ is

$$\frac{\partial W^r_{wi}}{\partial \bar{\sigma}^r} = t^r(\lambda^r - 1)(Q - \bar{p} - C^r) - C^r - \lambda^r A^r\,. \tag{21}$$

A reduction in $\bar{\sigma}^r$ reduces tax revenue (first term) but saves compliance and tax administration costs (second and third terms). Again, assuming that $\bar{\sigma}^s$ is optimally set from a global point of view, i.e., that (14) equals zero, the above equation can be rewritten as

$$\frac{\partial W^r_{wi}}{\partial \bar{\sigma}^r} = -(Q - \bar{p})t^s - C^s(1 - t^s) < 0\,. \tag{22}$$

We can thus state

Proposition 4 If the source-based threshold is equal to or above the globally optimal threshold level, the residence country has an interest in unilaterally deviating from the threshold set by the source country by moving to white income.

This result can be explained as follows. At the globally optimal nexus threshold, a further reduction of the source country threshold neither increases nor decreases the welfare of the residence country. It transfers tax revenue from the residence country to the source country, which is just compensated by savings in compliance and administration costs. A unilateral move to white income implies the same savings in compliance and administration costs, but at a lower cost: rather than transferring tax revenue to the source country, the unilateral decline in the residence-based threshold implies a tax cut for domestic firms. This explains why moving to white income is attractive.

6. Discussion and Conclusions

The analysis in this paper has shown that compliance and administration costs may provide an economic rationale for the fact that countries often choose a mix between source-based and residence-based taxation. Our model thus sheds additional light on the legal provisions of "permanent establishment" and locational profits for tax purposes and the fact that these institutions refer to the extent of economic activity in a given location. It is interesting to contrast our argument with existing approaches to explaining the observable mix of taxation principles.

As far as we know, there are mainly three approaches. The first explains the existence of source-based affiliate taxation by the mere inability of countries to coordinate their tax policies. We do not question the empirical relevance of this argument, but may add that, even if they could coordinate, they might find it desirable to choose a tax mix. The second approach is based on the theorem that source-based taxation is efficient if there are pure profits. While this is a valid argument under certain assumptions, it does not apply to our framework, for the following reason. By setting all transfer prices $\bar{p}(\sigma) = Q \ \forall\sigma$, and consequently allowing for deduction of all cost from the headquarters tax base (not modeled in our framework), it is possible to tax pure profits on a residence basis. The third approach focuses on prohibitively high costs of implementing or enforcing residence-based taxation. This may be the case if, for example, the affiliate has many shareholders, each of which would have to remit taxes if source-based taxation of the affiliate were ruled out. Although we do not model dispersed ownership, we would claim that

this approach refers to special cases, which our model implies by generally referring to *relative cost differences* between the two principles of taxation.

With respect to policy implications, our paper is probably most relevant for debates about the allocation of taxing rights and obligations between states of federations and within the European Union. One may note that compliance and administration costs play a key role in corporate tax coordination proposals in the European Union. For instance, the establishment of home-state taxation would imply that source-based taxation of small and medium-sized firms in the EU would virtually disappear. In the framework of our model, this would be equivalent to a coordinated increase of nexus thresholds. The objective of this proposal is to reduce compliance costs borne by small and medium-sized firms. Reducing compliance and administration costs is also the main objective of other proposals, including the initiative for a Common Consolidated Corporate Tax Base.

References

Becker, J., and Fuest, C. (2010), Taxing Foreign Profits with International Mergers and Acquisitions, International Economic Review 51, 171–186.

Becker, J., and Fuest, C. (2011), Tax Competition and International Mergers and Acquisitions: Source versus Residence Based Taxation, Journal of Public Economics 95, 28–40.

Bird, R. M. (2002), Why Tax Corporations?, Bulletin for International Fiscal Documentation 56, 194–203.

Davies, R. B. (2004), Tax Treaties and Foreign Direct Investment: Potential versus Performance, International Tax and Public Finance 11, 775–802.

Desai, M. A., and Hines, J. R. (2003), Evaluating International Tax Reform, National Tax Journal 56, 487–502.

Desai, M. A., and Hines, J. R. (2004), Old Rules and New Realities: Corporate Tax Policy in a Global Setting, National Tax Journal 57, 937–60.

Devereux, M. P. (1990), Capital Export Neutrality, Capital Import Neutrality, Capital Ownership Neutrality and All That, Unpublished Manuscript.

Devereux, M. P., and Keuschnigg, C. (2009), The Distorting Arm's Length Principle, Oxford University Centre for Business Taxation Working Paper Series No. 09/10.

Dharmapala, D., Slemrod, J., and Wilson, J. D. (2011), Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs, Journal of Public Economics 95, 1036–47.

Diamond, P. A., and Mirrlees, J. (1971), Optimal Taxation and Public Production, American Economic Review 61, 618–27.

European Commission (2001), Company Taxation in the Internal Market, Commission Staff Working Paper, Brussels.

European Commission (2004), European Tax Survey. Directorate-General Taxation and Customs Union Working Paper 3/2004.

Evans, C. (2003), Studying the Studies: An Overview of Recent Research into Taxation Operating Costs, eJournal of Tax Research 1, 64–92.

Graetz, M. J., and Oosterhuis, P. W. (2001), Structuring an Exemption System for Foreign Income of U.S. Corporations, National Tax Journal 54, 771–786.

Hines, J. R., and Rice, E. M. (1994), Fiscal Paradise: Foreign Tax Havens and American Business, Quarterly Journal of Economics 109, 149–182.

Huizinga, H., and Laeven, L. (2008), International Profit Shifting within European Multinationals, Journal of Public Economics 92, 1164–1182.

Keen, M., and Mintz, J. (2004), The Optimal Threshold for a Value-Added Tax, Journal of Public Economics 88, 559–576.

Kopczuk, W., and Slemrod, J. (2006), Putting Firms into Optimal Tax Theory, American Economic Review 96, 130–134.

Mansori, K. S., and Weichenrieder, A. (2001), Tax Competition and Transfer Pricing Disputes, FinanzArchiv/Public Finance Analysis 58, 1–11.

McLure, C. E. (2000), Alternatives to the Concept of Permanent Establishment, CESifo Forum 1, 10–16.

McLure, C. E. (2002), Thinking Straight about the Taxation of Electronic Commerce: Tax Principles, Compliance Problems, and Nexus, in: Poterba, J. M. (Eds.), Tax Policy and the Economy. Volume 16, MIT Press for the National Bureau of Economic Research, Cambridge, 115–40.

Mintz, J. (1996), Corporation Tax: A Survey, Fiscal Studies 16, 23–68.

Nutt, A. (2006), Glaxo, U.S. Settle Transfer Pricing Dispute, Tax Notes International 43, 956–956.

OECD (2008), Report on the Attribution of Profits to Permanent Establishments, Centre for Tax Policy and Administration, Brussels.

Slemrod, J. (2008), Does It Matter Who Writes the Check to the Government? The Economics of Tax Remittance, National Tax Journal 61, 251–275.

Slemrod, J., and Blumenthal, M. (1996), The Income Tax Compliance Cost of Big Business, Public Finance Quarterly 24, 411–438.

Slemrod, J., and Venkatesh, V. (2002), The Income Tax Compliance Cost of Large and Mid-Size Businesses, Ross School of Business Paper No. 914.

United Nations (1998), The United Nations Model Double Taxation Convention Between Developed and Developing Countries in Practice, in: International Cooperation in Tax Matters, United Nations, New York.

Clemens Fuest
University of Oxford
Saïd Business School
Centre for Business Taxation
Park End Street
Oxford OX1 HP
United Kingdom
clemens.fuest@sbs.ox.ac.uk

Reforming the Retirement Scheme: Flexible Retirement versus Legal Retirement Age

Juan A. Lacomba and Francisco M. Lagos*

Received 6 July 2011; in revised form 6 February 2012; accepted 7 February 2012

We compare a social security system where people can retire at an age of their own choice with one in which there is a legal retirement age elected through a majority voting process. We show that individuals prefer a legal retirement age higher than the one they would choose in the flexible scheme. In spite of this, we show that when the legal retirement age significantly limits the retirement age of high-wage workers, a flexible scheme would improve the financing of the pension system. Finally, we show that even when pension benefits are higher with a legal retirement age, a flexible system might be implemented.

Keywords: social security, flexible retirement, legal retirement age

JEL classification: H 55, J 26

1. Introduction

Reforms of social security systems are now among the main issues on most industrialized countries' economic policy agendas. It is widely held that, unless serious changes take place, the rise in the number of retirees relative to that of workers will threaten the viability of pay-as-you-go (PAYG) public pension systems in the long run. With the aim of eliminating these future financing problems, one of the main goals of pension reforms is to raise the average age of retirement of workers; see Blondal and Scarpetta (1998) or Gruber and Wise (1999).

In order to achieve this objective, one of the main economic policy measures is to allow greater flexibility in social security retirement rules (as, e.g., in Germany, Italy, or Sweden). Indeed, this measure is one of the policy conclusions of "Maintaining Prosperity in an Ageing Society," OECD (2000,

* We are especially grateful to Ignacio Ortuño Ortín for helpful and constructive discussions and suggestions during the elaboration of this paper. We also thank Elena Bárcena, Georges Casamatta, Subir Chattopadhyay, Ramõn Faulí, Francisco Marhuenda, and Shlomo Weber for valuable comments. Special thanks go to two anonymous referees for their useful comments. Financial support from the Spanish Ministry of Education and Science (SEJ2009-11117/ECON) and Junta de Andalucía (P07-SEJ-03261) is gratefully acknowledged.

FinanzArchiv 68 (2012), 252–268 doi: 10.1628/001522112X653831
ISSN 0015-2218 © 2012 Mohr Siebeck

p. 8): "...the most appropriate reform would be to allow people to retire at the age of their own choice and to adjust the pension level so that the pension system is neutral on average."

There is recent literature dealing with a flexible retirement age and social security. Casamatta et al. (2005) study the distortion caused by the continued activity of elderly workers in a setting with flexible retirement. They allow individuals to vote on the level of the payroll tax and provide sufficient conditions for the existence of a voting equilibrium. Conde Ruiz and Galasso (2003) analyze the effects of a simultaneous voting process on the contribution rate and on the decision to introduce or not an early retirement provision with an endogenous retirement age. Simonovits (2006) analyzes the optimal design of the pension system with flexible retirement, focusing on the importance of asymmetric information.[1]

The present article augments the existing literature on endogenous retirement decisions by explicitly examining the effectiveness of increasing the flexibility of the pension system.[2] Should the pensionable age be eliminated and greater flexibility in retirement decisions allowed? Similarly to Casamatta et al. (2005), we consider a two-period overlapping-generations (OLG) model where each individual in the first period works one unit of time and in the second period works for some time and then retires. Individuals live for two periods and differ according to productivity. We study two different PAYG social security programs with flat pension benefits. In the first setting people can retire at an age of their own choice. In the second one, there is a legal retirement age elected through a majority voting process.[3]

We analyze how incentives on retirement decisions depend on the retirement rules. We also compare both the financing of the pension system and the welfare of the population associated with the two retirement rules. The preferences with regard to these two opposite systems, flexibility versus legal retirement age, will depend on the effect on welfare. We interpret these preferences as a voting decision on the retirement rules.

We show that individuals prefer a legal retirement age higher than that they would choose in the flexible scheme. This result highlights the different incentives on prolonging the working period related to each retirement

1 Earlier literature mainly focused on the effect of the introduction of a pension system on the individual retirement decision (see among others Burbidge and Robb, 1980; Crawford and Lilien, 1981; or Sheshinski, 1978).

2 Michel and Pestieau (1999), in a model that allows for endogenous retirement, show that a mandatory early retirement may be socially desirable in the case of underaccumulation.

3 To date, there are few studies that have examined the role of the legal retirement age in the pension system. Lacomba and Lagos (2006) and (2007) study the problem of a direct vote on the legal retirement age and the effect of the aging of the population on the optimal legal retirement age.

scheme. In a flexible system individuals ignore the impact of their decisions on the social security budget constraint, as they only optimize their own retirement ages. But in a system with a legal retirement age that affects all the population, these indirect effects are taken into account and lead to higher preferred legal retirement ages.

In spite of the aforementioned, when there is sufficient dispersion in retirement ages, that is, when the legal retirement age significantly limits the retirement age of high-wage workers, a shift from a pension system with legal retirement age to a flexible scheme might enhance the financing of the pension system. This retirement dispersion mainly depends on the wage distribution and on the elasticity of the labor force.

Finally, we show that even if pension benefits were higher with a legal retirement age, a majority of the population, formed by low- and high-wage workers, might prefer the flexible system. Only the middle class would be in favor of a legal retirement age as an instrument to increase their pension benefits by forcing lowest-wage workers to work longer.

The paper is organized as follows. Section 2 develops the model. In Section 3 optimal retirement decisions are obtained. In Section 4 we compare the financing of the pension system and welfare levels under the two retirement schemes. In Section 5 a numerical example illustrates the results obtained in the previous sections. Section 6 summarizes the main results. The proofs appear in the appendix.

2. The Model

This model is similar to that of Casamatta et al. (2005). Individuals live for two periods. They are located between a minimum and a maximum wage level per unit of time (productivity), $[w_-, w^+]$, with mean $\bar{w}$ and median $w_m < \bar{w}$.

The intertemporal utility function is as follows:

$$U(c, d) = u(c) + \beta u(d)\,. \tag{1}$$

The utility function $u(.)$ is, as usual, increasing and concave: $u'(.) > 0, u''(.) < 0$; c and d are respectively the first- and second-period consumptions; and β is the time preference factor, which is equal to $1/(1 + r)$, r being the interest rate.

The two periods are of equal length, normalized to unity. Labor supply is assumed to be inelastic in the first period. In the second period, we have to distinguish between the two settings. In the first case, individuals choose their own retirement age by deciding the fraction of the second period they continue working, $R \in [0, 1]$; so R can be interpreted as an indicator of the individual retirement age. In the second case, individuals have to work the

fraction of the second period chosen through a majority voting process; this fraction can be interpreted as the indicator of the legal retirement age of the system.

It should be noted that the second-period consumption d includes the normal consumption minus the monetary disutility of working in this second period. We assume a particular specification for this disutility, $d = x - \gamma R^{\delta+1}/(\delta+1)$, where x is the normal consumption in the second period, and $\gamma > 0$ and $\delta \geqslant 1$ can be interpreted as intensity factors of the disutility of work.[4] So the first- and second-period consumptions for an individual of wage w are

$$c = w(1-\tau) - s\,, \tag{2}$$

$$x = s(1+r) + Rw(1-\tau) + (1-R)p\,, \tag{3}$$

where $s \geqslant 0$ is the amount of savings; $\tau \in [0, 1]$ is the social security contribution tax rate; and p is the constant stream of pension benefits per unit of time, collected from a PAYG social security system. Similarly to Casamatta et al. (2005), we consider a Beveridgean pension system where individuals contribute an amount proportional to their wages but the total pension benefit received does not differ across them. We also assume that the contribution rate is given. In this way, we focus our attention only on how the different retirement rules affect the financing of the system and the welfare levels of individuals. The optimal design of the social security parameters has already been analyzed in recent literature.[5] Our study can be considered as complementary to these, concentrating on another specific issue of the pension-system reform problem.

3. Retirement Decisions

In this section, we characterize retirement decisions (and savings) of old and young individuals under the two retirement schemes. We denote by $\left(R_F^*, R_L^*\right)$ the optimal retirement decisions, where R_F^* is the optimal individual retirement decision of young and old individuals under the flexible scheme, and R_L^* is the optimal legal retirement age of young and old individuals under the legal retirement scheme.

4 The quadratic specification used by Casamatta et al. (2005) is a particular case of the specification used here, where $\delta = 1$. It should also be noted that with this utility function income effects are disregarded: changes in the optimal retirement decision will only be caused by variations in the relative price of labor and consumption.

5 See among others Breyer (1994); Casamatta et al. (2000a), (2000b), and (2005); or Conde Ruiz and Galasso (2003, 2004).

3.1. Flexibility in Retirement Decision

Nowadays, there are many examples of countries with flexible retirement regimes (see Blondal and Scarpetta, 1998). In our model, under this pension scheme, individuals in the second period are allowed to retire at an age of their own choice, and pension benefits are paid out after leaving the labor force. Let R_F be the individual retirement age. The budget constraint of a feasible social security system must satisfy

$$\tau\left(N^y\int_{w_-}^{w^+} wf(w)dw + N^o\int_{w_-}^{w^+} R_F wf(w)dw\right) = N^o\int_{w_-}^{w^+}(1-R_F)pf(w)dw\,, \tag{4}$$

where N^o and $N^y = (1+n)N^o$ are respectively the numbers of old and young individuals, and n is the population growth rate. Pension benefits per unit of time are given by

$$p = \tau\left(\frac{(1+n)\bar{w}}{1-R_F} + \frac{\overline{p}_F}{1-R_F}\right), \tag{5}$$

where $\bar{w}$ denotes the mean wage and

$$\overline{p}_F = \int_{w_-}^{w^+} R_F wf(w)dw \tag{6}$$

satisfies the government budget constraint (4). Note that a Beveridgean system where the total pension received does not differ across individuals implies that under a flexible scheme the total pension

$$P = (1-R_F)p = \tau((1+n)\bar{w} + \overline{p}_F) \tag{7}$$

does not depend either on R_F or on w. In other words, although both R_F and p are different for individuals with different wages, the product $(1-R_F)p$ is assumed constant. Though retiring later has a cost in forgone p, this cost is canceled out by the increase in p derived from retiring later.

Now we turn to the analysis of the optimal retirement decisions. The old individuals' problem can be formally represented as

$$\max_{R_F} s_F(1+r) + R_F w(1-\tau) + (1-R_F)p - \frac{\gamma R_F^{\delta+1}}{\delta+1} \tag{8}$$

subject to

$$0 \leqslant R_F \leqslant 1\,.$$

From (8) we obtain the optimal retirement age, R_F^*:

$$R_F^* = \left(\frac{1-\tau}{\gamma}\right)^{1/\delta} w^{1/\delta}\,. \tag{9}$$

It is easy to check that the young individuals anticipate that they will choose her retirement age according to (9).[6] It is worth noting that (9) is obtained assuming that the individual considers her pension benefits as given (that is, $\partial \overline{p}_F / \partial R_F = 0$), thereby disregarding the effect of her retirement decision on the pension benefits via the "macro" constraint, (4). As Sheshinski (1978) states, this is a plausible assumption under competitive conditions with many individuals. Besides, note that although retiring later has a cost in forgone p, this cost is canceled out by the increase in p derived from retiring later. In other words, although p depends on R_F, the total pension received, P, does not depend on R_F.

Due to the positive substitution effect and the absence of the income effect, the retirement decision is positively related to the wage level. On the other hand, a larger contribution rate reduces the net wage and consequently leads individuals to retire earlier. Finally, a higher intensity factor of the disutility of work, δ, not only reduces optimal retirement ages but also diminishes the elasticity of the labor force.

On the other hand, we can substitute (9) into (6), and denoting

$$\xi(w) = \int_{w_-}^{w^+} w^{\frac{(\delta+1)}{\delta}} f(w) dw \,, \tag{10}$$

we can rewrite (6) as

$$\overline{p}_F = \left(\frac{1-\tau}{\gamma} \right)^{1/\delta} \xi(w) \,. \tag{11}$$

Equation (11) allows us to observe that pension benefits coming from old workers' contributions, $\overline{p}_F$, depend on the elasticity of the labor force.[7]

3.2. Legal Retirement Age

In some countries there are direct restrictions on work beyond the standard retirement age (in Portugal and Spain entitlement to pension benefits beyond the standard age is conditional on complete withdrawal from work), or, frequently, individuals have to leave their current jobs to receive their pensions (see Blondal and Scarpetta 1998 or Gruber and Wise 1999). So we can observe that the average retirement age in some OECD countries, such as the United Kingdom, Portugal, and Ireland, is very close to this standard retirement age.[8] Thus, in this setting we consider the legal retirement age

6 See Cassamatta et al. (2005) for a more exhaustive analysis of the optimal savings decisions in this setting.

7 Equations (10) and (11) are needed for the further analysis of the financing of the system and the welfare levels.

8 If there is a possibility of early access to pension benefits with some adjustment to the value of retirement benefits, the average retirement age is usually found between the

as the age at which workers are obliged to leave the labor force, that is, as a mandatory retirement.

In the following, we shall first derive the optimal legal retirement age of individuals. Then, we shall turn our attention to the majority voting process and obtain the elected legal retirement age.

Let R_L be the legal retirement age. A feasible social security system's budget constraint must now satisfy

$$\tau\left(N^y\int_{w_-}^{w^+} wf(w)dw + N^oR_L\int_{w_-}^{w^+} wf(w)dw\right) = N^o(1-R_L)\int_{w_-}^{w^+} pf(w)dw\,. \tag{12}$$

Under this scheme, the individual pension is

$$p = \tau\left(\frac{(1+n)\bar{w}}{1-R_L} + \frac{\overline{p}_L}{1-R_L}\right), \tag{13}$$

with

$$\overline{p}_L = R_L\bar{w}\,. \tag{14}$$

3.2.1. Optimal Retirement Decisions

The old individuals' problem can be formally represented as

$$\max_{R_L}\; s_L(1+r) + R_Lw(1-\tau) + (1-R_L)p - \frac{\gamma R_L^{\delta+1}}{\delta+1} \tag{15}$$

subject to

$$0 \leqslant R_L \leqslant 1.$$

From (15) we obtain the optimal legal retirement age

$$R_L^* = \left(\frac{w+\tau(\bar{w}-w)}{\gamma}\right)^{1/\delta}. \tag{16}$$

It is easy to check that the young individuals anticipate that they will choose their optimal legal retirement age according to (16).[9] Under this scheme the positive relationship between the wage and the preferred legal retirement age can again be explained by the substitution effect, which calls for a higher retirement age. A higher intensity factor of the disutility of work also reduces optimal legal retirement ages.[10]

However, unlike the flexible system, a higher τ delays the optimal legal retirement age for those individuals with wages lower than the mean wage.

age at which pensions can be accessed and the standard retirement age. See Blondal and Scarpetta (1998) or Samwick (1998).

9 See Lacomba and Lagos (2007) for a more exhaustive analysis of the optimal decisions in a setting with legal retirement age.

10 In this scheme, the total pensions received are $P = (1-R_L)p = \tau((1+n)\bar{w} + R_L\bar{w})$.

In order to explain this result, let us recall the double effect associated with changing the legal retirement age. These changes affect the working population's lifetime income in two ways: fixing the length of the working period and, in an indirect way, determining the pension benefits via the dependency ratio. For instance, a delay in the legal retirement age not only increases the working period but also increases the pension benefits by increasing the dependency ratio. Thus, the larger the pension benefits, the bigger the indirect effect on the lifetime income of a change in the legal retirement age. The reason is the larger relative weight of the pension benefits on the individuals' lifetime income. Therefore, the increase in net pension benefits of low-wage workers caused by a greater contribution rate augments the importance of these indirect effects, increasing the relative price of leisure, so that individuals relocate their demand from leisure to consumption and delay their retirement age.

3.2.2. The Voting Process on the Legal Retirement Age

It is easy to check that preferences are single-peaked with respect to the legal retirement age and thus a Condorcet winner exists. The majority voting process leads to a legal retirement age, R_L^e, that divides the population into two groups of equal size: those who prefer a retirement age above the elected age and those who prefer a retirement age below the elected one. Since optimal legal retirement ages are increasing in the wage, the elected one is the median wage individual's optimum legal retirement age

$$R_L^e = \left(\frac{w_m + \tau(\bar{w} - w_m)}{\gamma} \right)^{1/\delta}. \tag{17}$$

Since $\bar{w} > w_m$, we find that a larger contribution rate will lead to a higher elected legal retirement age and therefore to a longer working period. This result contrasts with the one obtained in the flexible system, where a larger contribution rate yields lower retirement ages.[11]

Comparing the retirement decisions obtained under the two retirement schemes, the following proposition can be stated.

Proposition 1 (i) $R_L^* > R_F^*$ for any $w \in [w_-, w^+]$.
(ii) $R_L^e > R_F^*$ for any $w \in (w_-, w_\mu)$ with $w_\mu > w_m$.

Proof. Relation (i) follows straightforwardly from (9) and (16).

11 If the mean wage were lower than the median wage, $\bar{w} < w_m$, the opposite result would be obtained. In this case, an increase in the contribution rate would decrease the net pension benefits of the median voter, reducing the importance of the indirect effects. This fact would lower the relative price of leisure, so that the median voter would relocate her demand from consumption to leisure and would advance the legal retirement age.

Regarding (ii), (17) and (9) can be respectively rewritten as

$$R_L^e = \left(\frac{w_m(1-\tau) + \tau \bar{w}}{\gamma} \right)^{1/\delta} \tag{18}$$

and

$$R_F^* = \left(\frac{w(1-\tau)}{\gamma} \right)^{1/\delta}. \tag{19}$$

Needless to say, $R_L^e > R_F^*$ for any $w \in [w_-, w_m]$. From (18) and (19) it can be derived that there exists a wage $w_\mu > w_m$ such that $R_F^*(w_\mu) = R_L^e$. ■

The first point of the proposition states that any individual would have her preferred legal retirement age higher than that under a flexible system. Moreover, the second point shows that more than 50% of the population would retire earlier than the legal retirement age, all else remaining unchanged, if the pension system shifted from a legal retirement age to a flexible scheme.[12]

This result crucially relies on the different incentives on retirement decisions embedded in each pension scheme. As mentioned above, in a majority vote on the legal retirement age, the effects on the aggregate constraint of the adjustment made in the ratio of workers to retirees when the legal retirement age is lowered or raised must be taken into account. And these effects play such an important role that they lead people to prefer higher legal retirement ages.

On the contrary, in a flexible scheme the individual ignores the impact of her decision on the aggregate constraint (and therefore on her pension benefits) and considers that her retirement decision only affects the length of her working period. This attitude yields individual retirement ages lower than the legal one. This result can be regarded as the *fiscal externality* from imposing the legal retirement age on everyone.[13]

4. Financing of the Pension System and Welfare Levels

In this section we study the financing of the pension system and the welfare levels associated with each retirement scheme. In order to do so, we define S_F

12 It would be possible to find individuals who would prefer a higher legal retirement age than the current one, but they would retire even earlier than if the pension system shifted from a legal retirement age to a flexible scheme.

13 The results would not change if the utility function gave rise to an income effect. Although the income effect induces people to opt for a lower retirement age (the more income, the more demand for leisure), the positive relationship between the wage and the preferred retirement age would still occur both in the flexible system and in the legal retirement system, just assuming that the substitution effect is larger than the income effect. This assumption is usual in the related literature on this topic (see among others Crawford and Lilien, 1981, or Sheshinski, 1978).

and S_L as the amounts of money collected with flexible retirement and with a legal retirement age, respectively,

$$S_F \equiv \tau \left(N^y \int_{w_-}^{w^+} wf(w)dw + N^o \int_{w_-}^{w^+} R_F^* wf(w)dw \right), \tag{20}$$

$$S_L \equiv \tau \left(N^y \int_{w_-}^{w^+} wf(w)dw + N^o R_L^e \int_{w_-}^{w^+} wf(w)dw \right). \tag{21}$$

Using (10) and substituting (9) and (17) into (20) and (21), S_F and S_L can be rewritten as

$$S_F \equiv \tau \left(N^y \bar{w} + N^o \left(\frac{1-\tau}{\gamma} \right)^{\frac{1}{\delta}} \xi(w) \right) \tag{22}$$

and

$$S_L \equiv \tau \left(N^y \bar{w} + N^o \left(\frac{w_m + \tau(\bar{w} - w_m)}{\gamma} \right)^{1/\delta} \bar{w} \right). \tag{23}$$

The following proposition highlights the main results derived from (22) and (23).

Proposition 2 (i) The higher the contribution rate, the larger S_L.

(ii) The more dispersed the wage distribution or the more elastic the labor force, the larger S_F.

Proof. Statement (i) follows straightforwardly from (23). Statement (ii) follows straightforwardly from (22). ■

Raising contribution rates and delaying the retirement age are among the main reforms for eliminating the future financial problems of pension systems. The first point of the proposition tells us that in a system with legal retirement age these two measures may complement each other to achieve an increase in the amount of money collected (an increase in the contribution rate would facilitate the delay of the legal retirement age). On the contrary, in a flexible system it seems more difficult to implement these two measures together. As can be observed in (22), an increase in the contribution rate has a negative indirect effect on S_F (apart from the obvious positive direct effect). A higher τ would reduce the incentives to prolong the working period, worsening the financing of the system.

On the other hand, the second point states that both the wage distribution and the elasticity of the labor force are crucial in the financing of a flexible retirement system. The higher the wage dispersion or the more elastic the labor supply, the more likely that the financing of the system will be enhanced

by shifting to a flexible scheme. The intuition is the following. A higher elasticity of the labor force leads lower-wage individuals to retire earlier and higher-wage individual to retire later. And the increase in the financing of the pension system derived from the delay in the retirement decisions of higher-wage individuals is larger than the reduction in the financing derived from the lower retirement ages of lower-wage individuals.

We shall now compare the roles the retirement schemes play in determining the welfare of the population. We consider that the individual's welfare is determined by her utility level. Notice that the preferences of population with regard to retirement rules will depend on how their welfare levels are affected by the different retirement rules. Thus, these preferences can be interpreted as a voting decision on changing the retirement scheme. In this manner, we can also examine whether a flexible retirement system would be implemented or not. The results are characterized in the next proposition.

Proposition 3 (i) If $S_F \geqslant S_L$, the whole population will prefer a pension system with flexible retirement.

(ii) If $S_F < S_L$, a legal retirement system may be majority-preferred to a flexible system.

Proof. For (i), see appendix. For (ii), see the numerical example below. ■

The first point of the proposition is obvious. The effect of retirement rules on welfare has two different aspects: on the one hand, the effect on pension benefits, and on the other hand, the effect on retirement decisions. Needless to say, if a shift from a legal-retirement-age scheme to a flexible one enhances the financing of the system, all individuals will have, firstly, larger pension benefits, and secondly, the option of retiring at the age of their own choice, which unambiguously will improve their welfare levels.

A pension system with a legal retirement age yields higher welfare levels only if the related pension benefits are sufficiently large to compensate for the forced retirement. If they are large enough, as the second point of proposition states, a majority formed by individuals with wages around the median one could support a legal retirement system. This group could be considered as the middle class of the population (it would also include individuals with wages above the mean wage). The intuition is the following. Their optimal retirement ages are similar to the legal one, and therefore the forced retirement would not be very harmful, so a flexible system would bring about lower pension benefits only for them.

This result gives an intuition as to why social security in most countries has been related to a standard age of entitlement to public pensions instead of allowing total flexibility in the retirement decision. The legal retirement age might have been used by a vast middle class as a tool for improving

their pension benefits. The underlying idea is the following. By fixing a determined age at which workers are eligible for benefits, low-wage workers were forced to work longer. In this way, these workers had more income, which implies less redistribution from the richest workers to them, resulting in larger pension benefits for the middle class.

On the other hand, if pension benefits are not sufficiently large, the disutility derived from the forced retirement in the legal retirement system could lead a majority formed by a coalition made up of the tails of the wage distribution (low- and high-wage workers) to prefer the flexible scheme even with lower pension benefits – the low-wage group in order to be able to retire earlier without penalties, and the rich group in order to retire later.[14]

5. Numerical Illustrations

In this numerical example we shall illustrate the effect of the two different retirement schemes on the financing of the pension system and on individual welfare levels. In order to do so, we express the intertemporal utility function (1) of individuals in the following way. Let $V\left(R_F^*, w\right)$ and $V\left(R_L^e, w\right)$ be the indirect utility functions under flexible retirement and under legal retirement age, respectively. It is easy to check that $V\left(R_F^*(w), w\right) > (<) V\left(R_L^e, w\right)$ if $v\left(R_F^*(w), w\right) > (<) v\left(R_L^e, w\right)$, with

$$v\left(R_F^*(w), w\right) = R_F^*(w)\left(w(1-\tau) - \frac{\gamma\left(R_F^*(w)\right)^{\delta}}{\delta+1}\right) + \tau\overline{p}_F \tag{24}$$

and

$$v\left(R_L^e, w\right) = R_L^e\left(w(1-\tau) - \frac{\gamma\left(R_L^e\right)^{\delta}}{\delta+1}\right) + \tau R_L^e \bar{w}\,. \tag{25}$$

Thus, to obtain the results we use the following specifications. We consider two different distributions. In both of them, wages are distributed on $\left[w_-, w^+\right]$ with $w_- = 300$ and $w^+ = 16000$. They have the same mean wage, $\bar{w} = 3067.79$, but different median wages: $w_m = 2469.14$ in the first distribution and $w_m = 2674.22$ in the second one.[15] The first columns of tables 1 and 2 describe the contribution rates used. We consider three possibilities: $\tau = 0.25$, $\tau = 0.30$, and $\tau = 0.35$. The second, third, and fourth columns of the tables are related to

14 This result, a coalition made up of the tails of the income distribution, can also be seen in Epple and Romano (1996).

15 Data from the first distribution have been obtained from an income distribution of Spain (as an approximation of the wage distribution) estimated with the Dagum triparametric model. Annual data in thousands of pesetas. Year 1996. The second one is a different Dagum distribution skewed to the right.

the first wage distribution. The second and third columns contain the wages of the individuals indifferent between the two schemes (indifferent individuals with low wage, w_{lo}, and with high wage, w_{hi}, respectively). These individuals have the same welfare level under the two retirement schemes. The fourth columns display the percentage of individuals that enjoy an increased welfare level with the system with legal retirement age. They are located between the two previous wages. The fifth, sixth, and seventh columns contain the same information as the three previous ones, but related to the second wage distribution.

To complete the picture, we consider two different elasticities of the labor force. Table 1 shows the results related to the labor force LF_1, where, for $\tau = 0.25$, the range of optimal retirement ages under the flexible scheme is $R_F^*(w) \in [0.26; 0.72]$, and the legal retirement ages for each wage distribution are $R_L^e = 0.5$ and $R_L^e = 0.51$ respectively.

Table 1
Indifferent Wages and In-Between Percentages for LF_1

	$w_m = 2469.14$			$w_m = 2674.22$		
	w_{lo}	w_{hi}	%	w_{lo}	w_{hi}	%
$\tau = 0.25$	–	–	–	2985.8	4443.9	26.4
$\tau = 0.30$	–	–	–	2807.8	5265.5	37.8
$\tau = 0.35$	3232.1	5060.9	19.5	2642.6	6194.0	47.5

Table 2 contains the results for a more inelastic labor force, LF_2. In that case, the range of optimal retirement ages under flexibility is less dispersed, $R_F^*(w) \in [0.45; 0.53]$ for $\tau = 0.25$, and now the legal retirement age for both wage distributions is $R_L^e = 0.5$.[16,17]

We shall start with the analysis of table 1. The results illustrate the main intuitions suggested in the previous sections. For the first wage distribution, we obtain that a flexible-retirement pension system would always be preferred by the majority of the population. For $\tau = 0.25$ and $\tau = 0.30$ we observe that the whole population would be in favor of a change from a legal retirement age to a flexible retirement system. This implies that this change would lead

16 The different labor-force elasticities are generated considering $\delta = 4$ for LF_1 and $\delta = 24$ for LF_2. On the other hand, each δ is related to a different γ in order to get the same R_L^e for $\tau = 0.25$.

17 $\xi(w, \delta)$ has the following values for each wage distribution. For LF_1: $\xi(w, 4) = 24201.77$ for the first Dagum distribution, and $\xi(w, 4) = 23500.3$ for the second one. For LF_2: $\xi(w, 24) = 4321.07$ for the first distribution, and $\xi(w, 24) = 4303.35$ for the second one.

Table 2
Indifferent Wages and In-Between Percentages for LF_2

	$w_m = 2469.14$			$w_m = 2674.22$		
	w_{lo}	w_{hi}	%	w_{lo}	w_{hi}	%
$\tau = 0.25$	–	–	–	2891.0	4563.2	30.6
$\tau = 0.30$	3374.8	4207.6	10.7	2733.6	5386.1	41.3
$\tau = 0.35$	2995.0	5355.6	25.8	2578.0	6337.9	50.8

to larger pension benefits. However, for $\tau = 0.35$, although the flexible system would still be supported by a majority of the population, it would be financially undesirable. This can be deduced from the existence of a percentage of people preferring the system with a legal retirement age. Workers with wages comprised between $w = 3232.1$ and $w = 5060.9$ (constituting 19.5% of the population) would be better off with a pension system with legal retirement age. As the theory states, this only happens when the pension benefits associated with the legal retirement age are larger than those of the flexible system, and this only occurs when the financing of the pension system is larger with a legal retirement age.

The importance of the wage distribution is highlighted when we compare the results obtained with those of the second distribution. In this case the main objective of the reform, to improve the financing of the system, is never achieved. Pension benefits are always larger with the legal retirement age, regardless of the contribution rate. This is because the higher median wage, $w_m = 2674.22$, implies a higher legal retirement age, resulting in larger pension benefits. In spite of that, in all cases the reform would be supported by the majority of the population. Notice that the percentage of people preferring the legal-retirement-age system is always lower than 50% for the three contribution rates. Also notice that for $\tau = 0.35$ we have $w_{lo} < w_m$, which implies that a coalition of the low- and high-wage individuals is needed to support a pension scheme with flexible retirement.

We shall now turn to table 2. The comparison of the results with the previous ones documents the importance of the elasticity of the labor force. The negative effect of the more inelastic labor force LF_2 on the financing of the flexible scheme and therefore on the pension benefits means that all the percentages of individuals preferring a pension system with a legal retirement age grow with respect to those related to the more elastic labor force LF_1. Now, only for $\tau = 0.25$ in the first wage distribution is the financing of the system still improved with the flexible scheme. Moreover, we can observe how for $\tau = 0.35$ in the second wage distribution a majority of the population,

those workers with wages comprised between $w = 2578$ and $w = 6337.9$, would be in favor of a legal retirement age.

6. Conclusions

This paper has studied the importance of retirement rules for the financing of the pension system and for the welfare levels of individuals by comparing two polar cases, total flexibility in the retirement decision versus a system with a legal (mandatory) retirement age.

We have shown that in the flexible system individuals retire earlier than their preferred legal retirement ages. This result suggests that eliminating the standard age at which pension benefits are available and imposing a flexible system might have a hidden risk. The legal retirement age divides the population into working people and retired people, and this may be a reference point for most individuals. It may be easier for them to comprehend the indirect macro effects related to this age (apart from the direct effects on their own working periods) – for instance, to perceive the positive effects on the financing of the pension system from a delay in the legal retirement age. They may easily see that the improvement is derived from a reduction in the number of retirees and an increase in the number of workers. However, if we shift to a flexible system, when individuals decide on their retirement ages they will not consider that their single decisions affect the financing of the pension system (which is in fact plausible). And this misperception may lead them to retire earlier than the existing legal retirement age.

Thus, for that flexible system to succeed, the legal retirement age should considerably limit the current retirement ages of a large percentage of the population, mainly those of high-wage workers. We have shown that this will crucially depend on the elasticity of the labor force and on the wage distribution.

Finally, in most of the pension systems observed in reality there are both a minimum and a maximum retirement age. Within these two limits, individuals decide freely when to retire. Our result provides a rationale for the existence of this minimum retirement age. The role of the minimum retirement age is to force some individuals to retire later than they would do otherwise. On the other hand, letting people choose when to retire above this age makes the system more flexible than a pure mandatory retirement scheme and thus may attract larger political support.

7. Appendix: Proof of Proposition 3

Since $S_F \geqslant S_L$, there exists a wage level $\hat{w}$ such that $R_F^*(\hat{w}) = R_L^e$. Given that $S_F \geqslant S_L$ implies that $\overline{p}_F \geqslant R_L^e \bar{w}$, then $\nu(R_F^*(\hat{w}), \hat{w}) \geqslant \nu(R_L^e, \hat{w})$.

Now, we obtain the first derivative of (25) and the first and second derivatives of (24) with respect to the wage, and we get

$$\left.\frac{\partial v\left(R_L^e, w\right)}{\partial w}\right|_{w=\hat{w}} = R_L^e(1-\tau) \qquad \text{(A1)}, \tag{26}$$

$$\left.\frac{\partial v\left(R_F^*(w), w\right)}{\partial w}\right|_{w=\hat{w}} = R_F^*(w)(1-\tau) \qquad \text{(A2)}, \tag{27}$$

$$\left.\frac{\partial^2 v\left(R_F^*(w), w\right)}{\partial w^2}\right|_{w=\hat{w}} = \frac{(1-\tau)^{\frac{\delta+1}{\delta}}}{\delta\gamma^{\frac{1}{\delta}}} w^{\frac{1-\delta}{\delta}} > 0 \qquad \text{(A3)}. \tag{28}$$

The strict convexity of (24) guarantees that $v\left(R_F^*(w), w\right) > v\left(R_L^e, w\right)$, and therefore, $V\left(R_F^*(w), w\right) > V\left(R_L^e, w\right)$ for any $w \in \left(w_-, w^+\right)$. ■

References

Blondal, S., and Scarpetta, S. (1998), The Retirement Decision in OECD Countries, OECD-EDWP 202.

Breyer, F. (1994), Voting on Social Security when the Labour Supply Is Endogenous, Economics and Politics 6, 119–130.

Burbidge, J. B., and Robb, A. L. (1980), Pensions and Retirement Behaviour, Canadian Journal of Economics 13, 421–437.

Casamatta, G., Cremer, H., and Pestieau, P. (2000a), Political Sustainability and the Design of Social Insurance, Journal of Public Economics 75, 341–364.

Casamatta, G., Cremer, H., and Pestieau, P. (2000b), The Political Economy of Social Security, Scandinavian Journal of Economics 102, 503–522.

Casamatta, G., Cremer, H., and Pestieau, P. (2005), Voting on Pensions with Endogenous Retirement Age, International Tax and Public Finance 12, 7–28.

Crawford, V. P., and Lilien, D. M. (1981), Social Security and the Retirement Decision, Quarterly Journal of Economics 96, 505–529.

Conde-Ruiz, J. I. and Galasso, V. (2003), Early Retirement, Review of Economics Dynamics 6, 12–36.

Conde-Ruiz, J. I., and Galasso, V. (2004), The Macroeconomics of Early Retirement, Journal of Public Economics 88, 1849–1869.

Epple, D., and Romano, R. (1996), Public Provision of Private Goods, Journal of Political Economy 104, 57–84.

Gruber, J., and Wise, D. (1999), Social Security Programs and Retirement around the World, The Chicago University Press.

Lacomba, J. A., and Lagos, F. M. (2006), Population Aging and Legal Retirement Age, Journal of Population Economics 19, 507–519.

Lacomba, J. A., and Lagos, F. M. (2007), Political election on Legal Retirement Age, Social Choice and Welfare 29, 1–17.

Michel, P., and Pestieau, P. (1999), Social Security and Early Retirement in an Overlapping-Generations Growth Model. CORE Discussion Paper 9951.

OECD, (2000), Maintaining Prosperity in an Ageing Society. Organisation for Economic Cooperation and Development.

Samwick, A. (1998), New Evidence on Pensions, Social Security, and the Timing of Retirement, Journal of Public Economics 70, 207–236.

Sheshinski, E. (1978), A Model of Social Security and Retirement Decisions, Journal of Public Economics 10, 337–360.

Simonovits, A. (2006), Optimal Design of Old-Age Pension Rule with Flexible Retirement: The Two-Type Case, Journal of Economics 89, 197–222.

Francisco Lagos
University of Granada
Campus La Cartuja Granada
E-18011 Granada
Spain
fmlagos@ugr.es

Decentralization and Local Governments' Performance: How Does Fiscal Autonomy Affect Spending Efficiency?

Lorenzo Boetti, Massimiliano Piacenza, and Gilberto Turati*

Received 22 November 2010; in revised form 14 March 2012; accepted 26 March 2012

All over the world, recent legislative reforms share the goal of increasing fiscal autonomy of lower tiers of governments. The aim of this policy is to reduce vertical fiscal imbalance (VFI) and enhance the efficiency in the provision of public services, via increased accountability of local politicians. The purpose of this paper is to assess whether inefficiency of local governments is really affected by the degree of VFI, relying on a sample of Italian municipalities to study the determinants of spending performance. Consistently with modern fiscal federalism theories, our results show that more fiscally autonomous municipalities exhibit less inefficient behavior, thus supporting the waves of reforms towards the devolution of taxing power to lower government tiers.

Keywords: municipalities, decentralization, fiscal autonomy, electoral accountability, spending inefficiency

JEL classification: D 78, H 71, H 72, R 51

1. Introduction

In Italy, as in other countries around the world, recent legislative reforms share the goal of increasing the fiscal autonomy of lower tiers of govern-

* We wish to thank the Editor, Alfons J. Weichenrieder; two anonymous referees; Marie Estelle Binet; Frédéric Blaeschke; Fabio Fiorillo; Marcel Gerard; Fabio Padovano; Stéphane Saussier; Judith Stallmann; and seminar participants at the 66th Congress of the International Institute of Public Finance (Uppsala University, Sweden, 2010), the Workshop on Fiscal Federalism (University of Roma Tre, 2010), the 2010 Meeting of the European Public Choice Socierty (Izmir, Turkey), the 56th North American Meetings of the Regional Science Association International (San Francisco, CA, 2009), the 50th Conference of the Italian Economic Association (LUISS Guido Carli, Rome, 2009), the 21st Conference of the Italian Society of Public Economics (University of Pavia, 2009), the Workshop HERMES–Chaire PPP (University of Paris 1 Panthéon-Sorbonne, 2009), and the conference on The Economics of Political Economics (Catholic University, Milan, 2009), for helpful comments. Financial support by HERMES is gratefully acknowledged. The usual disclaimers apply.

doi: 10.1628/001522112X653840
ISSN 0015-2218
© 2012 Mohr Siebeck

ments, from regions to municipalities.[1] Enhancing tax decentralization implies a better alignment between spending and funding responsibilities and, as remarked by several economists, a potential improvement of the efficiency (as well as of the effectiveness) of public services provided to citizens. The mechanism to explain these improvements in public-spending efficiency and voter welfare – suggested by the modern theoretical literature on fiscal federalism – highlights the importance of increasing electoral accountability of incumbent local politicians, by forcing them to collect autonomously a substantial part of tax revenues used to finance their expenditures (e.g., Oates, 2005; Weingast, 2009).

This normative prescription is at the heart of the so-called second-generation theory (SGT), as opposed to the first-generation theory (FGT), of fiscal federalism. More precisely, FGT looks at government agencies as entities managed by welfare-maximizing politicians and analyzes the desirability of decentralization in the light of a sort of trade-off between, on the one hand, the efficiency of a decentralized provision of local public goods in the presence of differentiated preferences and, on the other hand, the inefficiencies from not internalizing possible scale economies and spillovers across jurisdictions (e.g., Oates, 1972, 1999). However, the types of inefficiencies on which FGT concentrates are not those that typically make newspaper headlines, such as mismanagement of public resources and real cases of corruption. To understand the dissipation of public monies one needs to recognize: first, that politicians do not typically act to maximize social welfare, but their own interest; second, that their effort in pursuing public goals cannot be directly observed by voters; third, that political institutions affect the heterogeneity of politicians. These are the arguments at the core of SGT, which focuses on a different trade-off from that in FGT. In particular, the centralization-versus-decentralization argument is based on the comparison between the benefits from greater coordination of policies under centralization (which favors the internalization of scale economies and spillovers) and the higher degree of electoral accountability of local politicians obtainable through fiscal decentralization (e.g., Besley, 2006).

Hence, from a normative point of view, decentralization should be pursued not only when there are differences in tastes for local services, but also as an effective tool to achieve better control by voters of politicians' performance. To this end, Oates (2005) suggests implementing a reliable and effective

1 For Italy, see the framework law 42/2009 on fiscal federalism, now partially implemented through the legislative decrees 23/2011 (relative to municipalities) and 68/2011 (relative to regions). An overview of the evolution of taxing power of subcentral governments in 30 OECD countries over the years 1995–2005 is provided in Blöchliger and Rabesona (2009), and Stegarescu (2005) investigates the long-run trend in the degree of tax-revenue decentralization for 23 OECD countries between 1965 and 2001.

system of local taxation, such as a property tax, because of the incentives this type of tax provides to local governments toward the provision of public goods that maximize citizens' property values and – in turn – their revenues (e.g., Tiebout, 1956; Glaeser, 1996). As Weingast (2009) puts it, "subnational governments that raise a substantial portion of their own revenue – i.e., with a low degree of Vertical Fiscal Imbalance – tend to be more accountable toward the citizens, to provide market-enhancing public goods, and to be less corrupt."

Is tax decentralization really effective in ensuring better performance of local governments, in particular in terms of spending efficiency? Empirical studies on incentive effects stemming from local taxation – starting with the seminal paper by Oates (1985) – are mostly focused on how decentralization affects government size, implicitly assuming that large spending is inefficient (e.g., Jin and Zou, 2002; Rodden, 2003; Fiva, 2006; Borge and Rattsø, 2008; Eyraud and Lusinyan, 2011). A scant number of works have attempted to directly assess efficient spending by estimating production and cost frontiers (Coelli et al., 2005), which allows one to separate productive inefficiencies from structural expenditure and then to investigate the determinants of local governments' estimated inefficiency, exploring the role of different types of variables (socio-economic and political characteristics, spatial location, etc.). However, even in this literature, it is only in a very few cases that the determinants of spending inefficiency considered in the empirical analyses can be related to factors that SGT deems to be important in order to generate the right incentives for higher accountability. In particular, one should notice the ambiguous effects estimated for local taxes: though an inverse relationship between higher local tax rates and the inefficiency of municipalities emerges in De Borger et al. (1994), De Borger and Kerstens (1996), and Vanden Eeckaut et al. (1993), the recent study by Balaguer-Coll et al. (2007) points to a positive effect of greater per capita tax revenues on inefficient spending. SGT suggests the importance of vertical fiscal imbalance (VFI), not of local taxes per se; but none of the studies on the efficiency of local governments has ever analyzed the role of VFI.

The goal of this paper is to analyze the role of VFI as a determinant of spending efficiency. To do so, we rely on a cross section of Italian municipalities, filling another gap in the strand of literature on spending efficiency of local governments, which has never considered Italy so far. To assess local spending efficiency, we exploit both parametric and nonparametric frontier estimation techniques (SFA and DEA, respectively). Following the existing empirical literature (e.g., De Borger and Kerstens, 1996; Prieto and Zofio, 2001; Balaguer-Coll et al., 2007; Giménez and Prior, 2007), we selected output indicators that are proxies for the level of services provided by local governments with respect to their most fundamental functions, identified in

terms of both their incidence on municipal budgets and their relevance for the citizens: general administration, waste management, education, elderly care, road maintenance, and local mobility. Inputs of local governments' activities are represented by the corresponding costs as accounted in municipal budgets, by disaggregating current expenditure according to these specific items. This represents an additional improvement over previous literature, which has so far relied on a crude measure of current spending considered as a whole. After defining the efficient spending frontier, the impact of fiscal autonomy is assessed, considering the ratio of municipal own taxes to total current revenues, which represents a measure of VFI and – more importantly – the best proxy for the electoral accountability of local politicians. We also augment our empirical model by considering the potential incentives to higher efficiency stemming from fiscal restraints imposed by the central government on the largest municipalities (the so-called Domestic Stability Pact). Finally, we test the robustness of our findings, considering also the role played by other potential drivers of local governments' performance, which embrace a variety of spatial, political, and organizational variables. Our main result is that VFI does matter in reducing spending inefficiency.

The remainder of the paper is structured as follows. In section 2, after discussing some institutional characteristics of Italian municipalities, we present our data and we define the variables and the empirical strategy. The results showing the effects of fiscal autonomy and of other variables affecting spending inefficiency are presented in section 3. Section 4 provides concluding remarks.

2. Assessing the Spending Inefficiency of Italian Municipalities

2.1. Institutional Features of Italian Municipalities

Italian public administration is characterized by different layers of governments below the central level: regions, metropolitan areas, provinces, and municipalities. The Republican Constitution – implemented in 1948 and amended in 2001 – assigns different tasks to these different local governments. In particular, excluding metropolitan areas (which are basically a selection of the biggest cities in various regions), municipalities are in charge of a wide array of services: from administrative services provided directly to citizens (including, for instance, the registry office) to local police, and from local mobility to waste management and social services (like childcare or care for the elderly). Funding for municipalities also includes a number of different sources of revenue, from own taxes and fees and charges for specific services to grants received from regional and central governments. According to aggregate data at the national level, about 2/3 of municipal

expenditure is funded with autonomous revenues, while the remaining 1/3 is received as a transfer from upper-level governments.[2]

As for tax revenues, the most important source of fiscal autonomy is represented by the local property tax, the so-called Imposta Comunale sugli Immobili (ICI), which brings in almost 1/4 of total municipal revenues. It applies to both domestic and business properties, according to a set of rules defined at the national level. Local governments can however freely set both the tax rate, in a range between 0.4% and 0.7%, and – to a certain degree – total or partial exemptions for specific types of property. Two other important local taxes are a surcharge on the personal income tax (Addizionale Comunale IRPEF) and the specific tax for waste collection and management (TARSU). As for the first, which represents more than 10% of total revenues at the national level, municipalities can only modify – within a limited extent – the tax rate. As for the second – which is slowly being changed from a tax to a tariff for the service provided – it is computed relying on a vague proxy of waste production (viz., the size of the dwelling), and municipalities can freely decide both rates and exemptions; nationally, it represents almost 10% of total revenues for municipalities. The distribution of the taxable basis is of course very different across municipalities, especially for the local property tax and the surcharge on the personal income tax.

Differences among municipalities arise also in administrative and political rules, according to the size of the town as measured by total population. For instance, the size of the municipal council varies between 12 members (for municipalities below 3000 inhabitants) to 50–60 members (for municipalities above 500,000 inhabitants). The remuneration of the mayor and of council members increases with population size too. The monthly gross salary of the mayor ranges between 1291 and 7798 euros; for council members it is computed as a percentage of the mayor's: it is 15% for small municipalities, and increases up to 75% for the largest ones. Electoral rules are also different, with a threshold fixed at 15,000 inhabitants: below this limit there is a single round of voting, while above the threshold voting is according to the runoff plurality rule. Term limits for the mayor are however the same, and no more than two consecutive terms of five years each are allowed at present. A threshold operates also for the possibility to create neighborhood councils within the city: these are sublevels of local governments with independent budgets and are allowed for municipalities with more than 30,000 inhabitants. Finally, as local governments' budgets are consolidated in the Italian budget of the public administration and contribute to defining the national deficit – which is relevant for the fiscal rules defined in the European

2 It is worth highlighting that the situation is very much differentiated across the country. Indeed, northern municipalities are funded to the extent of about 3/4 from autonomous revenues, while for southern municipalities the corresponding figure is only around 1/4 .

Stability and Growth Pact[3] – Italy has implemented a so-called Domestic Stability Pact (DSP) since 1999. The fiscal rules for municipalities and other subnational governments have often been varied by the central government, which imposed restraints alternatively on expenditure growth or on deficit size. The scope of the law spans all levels of the decentralized administrative structure: regions, provinces, and municipalities. However, starting from 2001, the municipalities with less than 5000 inhabitants were excluded from the DSP.[4]

Besides tax structure, political rules, and fiscal restraints, a last important dimension along which the municipalities appear to be different concerns the managerial model adopted for providing a local service of particular relevance, namely waste collection and disposal.[5] The observed alternatives range from direct production within the municipality (i.e., the so-called in-house provision), to the assignment of the function to a specific firm (publicly or privately owned), up to the creation of a cooperative aggregating two or more municipalities in the management of the service.[6]

2.2. Data and Variables

The sample we use in our empirical analysis is composed of 262 municipalities, all belonging to the province of Turin. The province of Turin represents an interesting case study within the Italian landscape, because it is the province with the highest number of local governments (315), thus ensuring great variability in the data. This variability is confirmed not only by looking at population size (included are Moncenisio, with 48 inhabitants, as well as Moncalieri, with 55,000 inhabitants, besides Turin – the chief regional town of the Piedmont – with over 900,000 residents), but also in terms of topography (more than 10% of municipalities are located above 1000 meters in alti-

3 The Stability and Growth Pact, first introduced in 1997 and revised in 2005, is an agreement among EMU member states aimed at mantaining and enforcing fiscal discipline in the EMU. For more details, see Brunila et al. (2001).

4 For a discussion of strengths and weaknesses of Italian DSP fiscal rules, refer to Giuriato and Gastaldi (2009). A critical analysis of the main European experience is provided in Ambrosanio and Bordignon (2009).

5 In principle, the differences among municipalities in management form involve also education and social services, like those provided by nursing homes and child-care centers. However, differently from waste management, spending for education and social services included in the municipal budget only represents direct transfers to the citizens to subsidize the access to these services; operating costs are not included.

6 As for the environmental services, one should recognize the importance of the national law D.Lgs. 05/02/1997 (the so-called Decreto Ronchi), which assigns different competencies to central state, regions, provinces, and municipalities in this field. In particular, it establishes the power of municipalities to determine the management form for waste collection and disposal.

tude), the management of public services, and political and socioeconomic characteristics.

To some extent, however, this huge heterogeneity across units may introduce potential biases in our study, especially in view of the presence of some municipalities that produce the analyzed services within particular geographical contexts and are subject to a different voting mechanism for the election of the mayor and the municipal council. Therefore, we have decided to exclude from the sample – besides Turin, one of the Metropolitan Areas recognized by the Constitution – all the towns over 15,000 inhabitants, as they are hard to compare with smaller municipalities along two relevant dimensions: with regard to spending, for these largest municipalities the share absorbed by the four sectors considered in our analysis represents much less than 80% of total current expenditure (see the discussion below), given the higher incidence of the remaining functions (e.g., local police, cultural and sport services, economic development); moreover, they are subject to a runoff voting mechanism, which is likely to significantly influence political outcomes and subsequent policy choices (e.g., Osborne and Slivinsky, 1996; Bordignon and Tabellini, 2009). We also excluded municipalities located above 900 meters in altitude, as they show remarkably higher expenditure levels than other municipalities of the province (on average, 1800 euros against 560 euros per capita): this may be due to the fact that provision of services is strongly affected by both the topography of the territory and the heavy tourist inflows, which clearly influence the peak demand for services such as waste management and local mobility.[7]

The data have been provided by different public institutions and refer to the year 2005 (the last period for which all the relevant information is available). Expenditures and revenues come from the budgets of Italian municipalities published by the Ministry of Domestic Affairs (the so-called Certificati Consuntivi). Other important data – related to output indicators and explicative variables for spending inefficiency – have been obtained from statistical services of Regione Piemonte and Provincia di Torino.

2.2.1. Input and Output Indicators

The definition of input and output variables is strongly influenced by the Italian institutional framework discussed above. More precisely, we select the spending items and related output measures by looking at the most important competencies, in terms of both their effects on the municipal budgets and the relevance of the provided services for the citizens.

7 Dividing the municipalities according to their altitude, one can observe that just starting from 900 meters they show levels of average current spending beyond 1000 euros per capita.

Table 1
Macro Functions of Municipal Current Expenditure in the Province of Turin

CATEGORY OF PROVIDED SERVICES	AVERAGE SHARE
General administration	38.6%
Territorial and environmental management	21.9%
Educational services	12.5%
Social services	9.4%
Road maintenance and local mobility	7.5%
Local police	4.7%
Cultural services	2.4%
Sports and entertainment	0.8%
Economic development	1.4%
Tourism	0.4%
Support to productive activities	0.3%
Justice	0.1%

In Italy, municipal current expenditure is classified into 12 macro functions. More than 90% of current expenditure in our sample is represented by five of these functions (see table 1): "General administration" (39%); "Environmental management" (22%); "Educational services" (13%); "Social services" (including child care and elderly care, 9%); "Road maintenance and local mobility" (8%). Clearly, the share of each function in local current spending varies according to municipality size: for instance, moving from the smallest municipalities (0–500 inhabitants) to the biggest ones (between 10,000 and 15,000 inhabitants), the weight of "General administration" decreases from 54% to 31%, while the shares of "Educational services" and "Social services" increase from 6% and 5% to 13% and 12%, respectively. We use current expenditure of municipalities for each of these items as an aggregate input indicator, given by the sum of the corresponding budget values.

For the categories "General administration," "Educational services," and "Road maintenance and local mobility," we consider the total expenditure as registered in the municipal budget. In order to strengthen the connection between spending and the selected output indicators, for the categories "Social services" and "Environmental management," we just retain a fraction of the whole expenditure devoted to these functions: spending for "Environmental management" only includes the subitem "Waste collection and disposal," which represents an important share of the total expenditure related to this task (60–70%); similarly, related to total spending for "Social services," we consider only the component specifically devoted to public

welfare and elderly care. Our final input indicator (*EXP*) represents, on average, 86% of total current expenditure, with very little variation across demographical classes of municipalities. Notice that this selection procedure represents a significant improvement over previous literature on local governments' efficiency, which has so far relied on the crude measure of total current expenditure considered as a whole.

As remarked by Fox (2001), output measurement of government departments is rather difficult and is often a source of controversy. Often, mainly due to data limitations, one has to select proxies for the provision of services (like demand indicators) instead of direct output measures. Moreover, the quality of public services represents another source of concern, since it can vary across municipalities and lead to different expenditure levels for the same output quantities. Here we closely follow the available literature (in particular, De Borger and Kerstens, 1996, and Balaguer-Coll et al., 2007) and define the four output indicators that are more directly linked with our selected spending categories: (1) the total served population as a proxy for "General administration" services; (2) the total amount of garbage collected for "Waste collection and disposal"; (3) the total number of people in need of care (i.e., those under 14 years old – enrolled in nursery, primary, and secondary schools – and those over 75 years old) for "Educational services" and "Social services"; (4) the total length of municipal roads for "Road maintenance and local mobility."

Although the publicly available information does not go much beyond these data and our output measures strictly mirror the indicators used in previous analyses of local governments' efficiency in Europe, we are nevertheless aware that most of these variables are very loose proxies of real outputs produced by municipalities. Indeed, with the exception of the amount of garbage collected (*WASTE*), which can be viewed as a direct output of the expenditure in waste management, the indicators listed above reflect citizens' needs more than the quantities of services actually provided. This suggests caution in interpreting results from the estimation of spending inefficiency (not only for this paper), which properly has to be read as a relative measure of excess spending for given citizens' needs. In particular, as De Borger and Kerstens (1996) and Balaguer-Coll et al. (2007) point out, the size of the population (*POP*) is usually assumed to proxy for the needs for various administrative services supplied by municipalities (management of registers and release of certificates for births, marriages, and deaths, etc.). The number of people under 14 years old and over 75 years old represents a consistent fraction of the needy, and the indicator (*DEPEND*) is reasonably correlated with the demand for educational and elderly care services. Finally, the total length of municipal roads (*ROAD*) is aimed at proxying for the needs associated with the ordinary management of the existing road infrastructures

(surface maintenance, public street lighting, local public transport arrangements, etc.).

We also recognize that output quality is a crucial issue when trying to assess local governments' performance. For the same level of output, municipalities may differ in the quality of the services provided – e.g., certificates may be obtained online, be issued in one day, or take several weeks; waste collection may be weekly or biweekly, etc. – and ignoring this aspect could cause greater spending due to higher quality to be mistakenly identified as greater inefficiency. However, with regard to outputs, measuring the quality of public services is a long-standing problem in local public finance, not only because of the lack of relevant data, but also with respect to the definition of quality. So far, only Balaguer-Coll et al. (2007) have considered a direct indicator built from a survey on citizens' perceptions of the quality of services provided. However, despite representing a clear improvement, these subjective measures could bias results as well, because perceptions are affected for instance by different frameworks or by previous experiences. Although similar data are currently unavailable for Italian municipalities, one may recur also in this case to a proxy of the demand for quality. From this point of view, it is well known that richer communities demand higher service quality (e.g., Bergstrom and Goodman, 1973; Reiter and Weichenrieder, 1997). We then exploit information on average municipal income and include the variable (*INCOME*) as a control for the (demand of) quality of the public output in our spending frontier model. Notice that taking into account differences in local communities' income – besides controlling for different kinds of services and efficient levels of public expenditures, for given citizens' needs – also allows us to consider other potentially relevant issues for efficiency, such as the heterogeneity in tax bases and different incentives to monitor municipal expenditures (De Borger et al., 1994; De Borger and Kerstens, 1996), which in turn should ensure a more precise assessment of the effects of fiscal autonomy and other determinants of spending efficiency.[8]

Table 2 shows the summary statistics for input and output indicators used in the empirical analysis. In addition to the aggregate value of spending input (*EXP*), also spending for each category is reported separately; these disaggregated values (*EXPGA*, *EXPWM*, *EXPEE*, *EXPRM*) will be used in a preliminary step of the empirical analysis in which we explore the influence of different types of variables (output proxies and inefficiency determinants) on the expenditure for each function. Then, in a second stage, a global approach is adopted in order to evaluate jointly the spending performance of all sectors.

8 We thank the editor and an anonymous referee for having raised all these critical aspects and suggested the inclusion of muncipal income in the model specification.

Table 2
Summary Statistics for Input and Output Indicators of SFA and DEA Spending Models

VARIABLE DESCRIPTION	NAME	Mean	Std. Dev.	Min	Max
INPUTS					
Current expenditure (10^3 euros)	*EXP*	1297	1284	95	6743
a) general administration	EXP_{GA}	604	509	88	2672
b) waste management	EXP_{WM}	278	348	11	2189
c) education and elderly care	EXP_{EE}	296	356	5	1927
d) road maintenance and local mobility	EXP_{RM}	119	106	7	595
OUTPUTS					
Population (no. of served inhabitants)	*POP*	2657	2826	102	13,835
Total amount of waste collected (quintals)	*WASTE*	12,117	13,914	486	76,107
Total number of pupils and old people (pupils enrolled in nursery, primary, and secondary school + over-75 inhabitants)	*DEPEND*	466	488	16	2449
Total length of municipal roads (km)	*ROAD*	33	28	3	240
CONTROL VARIABLES					
Municipal income per capita (10^3 euros)	*INCOME*	18.39	1.48	13.90	26.40
Dummy for *less than 1000 inhabitants*	*POP-1000*	0.35	0.48	0	1
Dummy for *more than 10,000 inhabitants*	*POP-10,000*	0.04	0.19	0	1
Dummy for *altitude over 600 meters*	*OVER-600*	0.17	0.37	0	1

It is worth noticing that our sample does not show any variability in input prices at which the municipalities buy their inputs. Indeed, there is no wage flexibility, as salary scales and allowances of municipal personnel are completely fixed. Moreover, since we are considering only the province of Turin, all municipalities have access to the same capital market, and obtain most of their funds from the same specialized financial institutions at the same interest rate. Thus, the hypothesis of identical input prices across municipalities is quite plausible.[9] Consequently, throughout the analysis we focus on the measurement of overall cost or spending inefficiency (with the meaning explained above), as it is more closely related to the nature of our data than pure technical inefficiency (which would require information on input prices in order to disentangle the allocative component of excess cost).

9 About this issue, see also the discussion in De Borger and Kerstens (1996).

2.2.2. Fiscal Autonomy and Other Determinants of Spending Inefficiency

The study focuses on the effects of tax decentralization and other explicative factors for estimated spending inefficiency, by relying on both the SFA and DEA methodologies discussed in section 2.3. Besides a measure of fiscal autonomy – the key issue of our analysis – the other aspects considered among the potential determinants of local governments' performance embrace a variety of fiscal, spatial, political, and organizational variables. Summary statistics are presented in table 3.

(a) *Fiscal Indicators*:
Similarly to other countries, Italian municipalities rely on three main sources of revenues: local taxes, fees and charges for specific services, and grants from upper-level governments. As a measure of fiscal autonomy, we adopt the tax decentralization indicator proposed for the first time by Akai and Sakata (2002), defined as the share of own taxes[10] (ICI + Addizionale Comunale IRPEF + TARSU) in local governments' total revenues.[11] As remarked by the authors, this indicator (*FISCAUT*) reflects how much public spending of lower-tier governments is maintained on the basis of tax revenues collected at the local level. It is thus a measure of VFI particularly suitable for testing the theoretical prediction that effective electoral accountability of local politicians – here interpreted as lower excess spending for given citizens' needs – can be obtained by increasing their responsibilities in funding. It is worth highlighting that the inclusion of the average municipal income among the variables defining the efficient frontier allows a better identification of the effects associated to the use of own fiscal tools, since we are able to control (at least to some extent) for the different tax bases available to different local communities.[12]

10 See the discussion in section 2.1.

11 This measure of tax autonomy has been subsequently adopted in the studies on fiscal decentralization in OECD countries by Stegarescu (2005) and Baskaran (2010), among others.

12 Indeed, the ICI and TARSU tax bases (the property value and the dwelling size, respectively) are strongly correlated with citizens' income level, while in the case of Addizionale Comunale IRPEF the tax base just coincides with citizens' income. Nevertheless, we are also aware that the use of the *FISCAUT* indicator may lead to an endogeneity problem, as the degree of fiscal autonomy is affected by local tax rates, and the way these are set is likely to depend in part on the ability of municipalities to manage their expenditures efficiently. However, this problem is hard to solve with the available information, which does not allow defining other proxies of fiscal autonomy strictly exogenous with respect to spending performance. Thus, we decided to rely on this measure, while recommending caution in interpreting the estimated impact of tax decentralization on spending efficiency in terms of a pure causal relationship.

Table 3
Summary Statistics for the Determinants of Spending Inefficiency

VARIABLE DESCRIPTION	NAME	Mean	Median	Std. Dev.	Min	Max	%
FISCAL INDICATORS							
Fiscal autonomy (% of revenues from local taxes on total current revenues)	*FISCAUT*	32	31	10	13	69	–
Local tax revenues per capita		154	150	34	77	448	
High taxes (municipalities with local tax revenues per capita over the median)	*HTAX*	–	–	–	0	1	50
Fees and charges per capita		146	116	109	31	904	
High extra taxes (municipalities with fees and charges per capita over the median)	*HEXTRA*	–	–	–	0	1	50
Grants per capita		219	203	100	14	696	
High grants (municipalities with grants per capita over the median)	*HGRANT*	–	–	–	0	1	50
Domestic Stability Pact (municipalities subject to the DSP fiscal rule)	*PACT*	–	–	–	0	1	15
SPATIAL INDICATORS							
Distance of the municipality from Turin	*KMTO*	37.09	36.50	13.78	8	72	–
Population density (inhabitants per km^2)	*DENS*	2.04	1.35	2.11	0.06	15.80	–
POLITICAL INDICATORS							
Electoral mandate (number of years since election for the governing coalition in 2005)	*YGOV*	1.40	1	1.03	0	4	–
Governing coalition with a civic list	*CIVIC*	–	–	–	0	1	56
Center–left-wing governing coalition	*LEFT*	–	–	–	0	1	23
Mayor's gender (municipalities with a male mayor)	*MAYORSEX*	–	–	–	0	1	83
Mayor's age (age of the mayor in 2005)	*MAYORAGE*	53	54	11	25	79	–
WASTE MANAGEMENT INDICATORS							
Public management	*PUBLIC*	–	–	–	0	1	77
Public management by a firm	*PUBLIC* × *FIRM*	–	–	–	0	1	33
Public management by a cooperative firm	*PUBLIC* × *FIRM* × *COOP*	–	–	–	0	1	27

To fully understand the role played by the accountability of local politicians, we introduce other fiscal indicators in the analysis. First, we decompose per capita current revenues into their three main sources – own taxes, fees

and charges, and grants – and for each category we identify the municipalities with a per capita level exceeding the median, for which the dummies *HTAX*, *HEXTRA*, and *HGRANT* are equal to 1. Then, we interact the three dummies with the variable *FISCAUT*, hence controlling for the possible presence of opportunistic behavior due to excessive availability of resources in the richest communities (distinguishing by type of revenue, as in Balaguer-Coll et al., 2007), which could loosen the improvements of electoral accountability potentially obtainable by the reduction in VFI. The tightness of the budget constraint and its influence on the accountability of local politicians is further investigated through the dummy variable *PACT*, which distinguishes local governments subject to the DSP (see section 2.1) from the municipalities (with less than 5,000 inhabitants) that – starting from 2001 – have been excluded from the application of this fiscal discipline rule.[13]

(b) *Spatial Indicators*:
The importance of the spatial dimension in determining the spending performance of decentralized governments has been highlighted by the strand of fiscal federalism literature that relies on the spillover approach for explaining the presence of possible interactions among expenditure decisions of neighboring jurisdictions (e.g., Revelli, 2003; Baicker, 2005; Ermini and Santolini, 2010). According to these studies, the benefits or possible detrimental effects of public expenditure (concerning social services, local mobility and road maintenance, environmental management, etc.) spread across the administrative boundaries of a jurisdiction, and the spending decisions of each community will possibly depend, besides its own characteristics, also on policies chosen elsewhere. The specific nature of the spillover – which can results in a positive or a negative impact on expenditure levels – depends on the relationship of complementarity or substitutability among local public services provided by the neighboring jurisdictions. Considering our context (all communities belonging to the province of Turin), it is reasonable to assume that those closer to Turin may be affected by both positive and negative spillovers of public-good provision there, and can free-ride on certain services (e.g., educational and elderly services, waste disposal) and spend more for others (e.g., road maintenance). To take into account the potential role played by these effects, we follow some previous studies on cost efficiency of local governments (e.g., Lokkainen and Susiluoto, 2004; Afonso and Fernandes, 2005) and include a variable for the distance of each municipality from Turin (*KMTO*). Furthermore, we use population density in the municipal area (*DENS*) to control for the presence of both density economies and congestion effects in the provision of public goods within each community.

13 See Kornai et al. (2003) for an extensive discussion of the "soft budget constraint" concept.

(c) *Political Indicators*:
Some political features of municipal governments are considered as potential determinants of spending efficiency. In particular, we define the variable *YGOV*, which assumes values from 0 to 4 and represents the number of years since election for the mayor and the governing coalition, in order to test the presence of opportunistic behavior by local politicians attributable to the electoral budget cycle. The theoretical argument here is that incumbent politicians – in an effort to signal their competence to the voters, so as to increase their chances to be reelected – tend to enlarge spending (inefficiently) when they are close to new elections (e.g., Rogoff and Sibert, 1988), i.e., when more years have passed since the previous election. We also interact *YGOV* with the fiscal indicators *PACT* and *FISCAUT*, so as to check for the presence of possibly relevant interplays between the impact of an electoral mandate deadline on spending efficiency and the variables reflecting the accountability and fiscal constraints of local politicians.

A control is also included for the political orientation of the governing coalition, using two dummy variables that assume the value 1 if coalition parties belong to a center–left list (*LEFT*) or to a so-called civic list with no clear ideological orientation (*CIVIC*). We finally consider two variables controlling for the age (*MAYORAGE*) and the gender (*MAYORSEX*) of the mayor, looking at recent political-economy literature that stresses the roles of more experienced and of female representatives in determining policy preferences and spending outcomes (e.g., Edlund and Pande, 2002; Chattopadhyay and Duflo, 2004; Dal Bó and Rossi, 2008; Funk and Gathmann, 2008).

(d) *Waste Management Indicators*.
We also assess the effects of different management models of waste collection and disposal that are observed in our sample. Indeed, this particular service may be provided by the municipalities adopting several organizational forms. The weight that waste management has recently gained in Italy for judging the behavior of local politicians can be easily understood in the light of yardstick competition between municipalities (e.g., Salmon, 1987; Revelli, 2006) and highlights the importance of reaching efficiency in spending for local administrations.[14] Waste collection can be managed directly by the local government; directly by a consortium of local governments with the possibility for a municipality to be either consortium head or a simple participant; through a specialized external firm, which can be either publicly or privately

14 Examples of the importance of waste management for the comparative evaluation of local administrators include the recent garbage crisis and the subsequent scandals in Naples and Palermo. See, e.g., "Naples burns as residents protest at garbage crisis," The Guardian, May 27, 2007.

owned; or through a publicly owned cooperative firm involving two or more municipalities. We summarize these six different organizational choices in three variables. A first dummy (*PUBLIC*) distinguishes public ownership from private; a second dummy ($PUBLIC \times FIRM$) indicates that the service is provided by an external firm, conditionally on this firm having public ownership; finally, a third dummy ($PUBLIC \times FIRM \times COOP$) represents a cooperative organization, conditionally on being a publicly owned firm.

2.3. The Empirical Strategy

In order to make our empirical analysis more transparent and easier to interpret, we first run simple OLS regressions for each sector separately, by investigating whether the variations in spending for general administration (*EXPGA*), waste management (*EXPWM*), education and elderly care (*EXPEE*), and road maintenance and local mobility (*EXPRM*) are actually related to their output proxies (*POP*, *WASTE*, *DEPEND*, and *ROAD*, respectively) and are affected by fiscal autonomy and the other explicative factors of inefficiency discussed above. Notice that, in this explorative analysis, we include in the model for each spending function only the corresponding output proxy and the variable *INCOME* as an indirect control for output quality, a set of dummies controlling for potential scale effects (viz., two variables measuring the effect on spending of extreme size classes: $POP\text{-}1000 = 1$ for the municipalities with less than 1000 inhabitants, and $POP\text{-}10{,}000 = 1$ for those with more than 10,000 inhabitants[15]; and a variable capturing the effect of altitude on spending: $ALT\text{-}600 = 1$ for the municipalities located above 600 meters)[16], and the determinants of inefficiency. As for the last mentioned, the three dummies summarizing the different organizational choices for waste management are included only in the cost models referring to this function and to the general administration – as also this category of spending is likely to be affected by the adopted schemes, especially in the case of public solutions (due to the presence of some common overhead costs) – while they are not considered in the analysis of spending relative to the other two categories

15 Following the classification adopted by the Ministry of Domestic Affairs, the municipalities have been divided into seven size classes: under 500 inhabitants (13% of observations), between 500 and 1000 (22%), between 1000 and 2000 (25%), between 2000 and 3000 (9%), between 3000 and 5000 (15%), between 5000 and 10,000 (11%), and finally over 10,000 (4%).

16 These thresholds were selected by looking at the distribution of per capita current spending of municipalities according to their population size and altitude. The municipalities under 1000 and over 10,000 inhabitants represent the extreme sides of a ∪-shaped trend that shows per capita spending along different dimensional classes. Moreover, the municipalities located at an altitude above 600 meters typically exhibit per capita spending levels significantly higher than the average of the sample. Interestingly, the 600-meter limit is also considered by the Italian Law 991/1952 to define mountain municipalities.

(*EXPEE* and *EXPRM*). After the sector-by-sector regressions, we proceed with a more global approach and run an OLS regression on the aggregated current expenditures (*EXP*) on the whole set of regressors (output proxies and structural control variables, as well as inefficiency determinants), so as to allow for potential interactions among the expenditure decisions related to the different functions.

As a final important step, we will try to disentangle the inefficient part of municipal spending (with the meaning clarified in section 2.2.1) from its structural part (which is driven by citizens' needs), and test whether such inefficiency is affected by the degree of tax autonomy and the other fiscal, spatial, political, and organizational variables. To this end, we will exploit the standard techniques adopted in the empirical literature to assess the efficiency of production units (firms as well as governments), which are usually classified as parametric and nonparametric methods. In particular, we estimate here both parametric stochastic frontiers (SFA model) and nonparametric deterministic frontiers (DEA model), since each technique presents advantages and disadvantages, but the literature has not been able so far to establish when one of them is strictly superior to the other (e.g., Coelli et al., 2005). Generally, when considering parametric techniques, the functional form of the best-practice frontier has to be defined a priori, while in the case of nonparametric techniques no functional form needs to be predetermined, and only the basic microeconomic properties of a production set are imposed as constraints on a linear programming problem. On the other hand, the SFA technique allows for both inefficiencies and random variables outside the control of the decision-maker that might affect the production performance, while standard deterministic frontiers like DEA are able to allow only for inefficiency, ruling out the role of stochastic disturbances. Given these pros and cons, it is important to check the robustness of our results, by using both approaches to investigate municipal spending inefficiency and the specific role played by tax autonomy.

More precisely, within the SFA approach, we focus here on the *cost function* representation of a given production technology for municipal services. For the *i*th observation, the cost function $C(q_i, w_i; \beta)$ defines a lower bound for spending, C_i, necessary to provide output levels q_i at given input prices w_i. The vector β is the set of technological (or structural) parameters to be estimated. Stochastic parametric frontiers are based on the specification of a composed error term (ε_i) that allows one to disentangle spending inefficiency from stochastic disturbances: a *symmetric* component (v_i) captures the usual random noise, while a *one-sided* (positive) error term (u_i) is introduced to measure cost inefficiency. When a Cobb–Douglas technology with no variability in input prices w_i is assumed (see De Borger and Kerstens, 1996, and the discussion at the end of section 2.2.1), the resulting SFA spending (or

cost) model – expressed in a logarithmic form – is

$$\ln C_i = \beta_0 + \sum_m \beta_m \ln q_{mi} + \sum_k \delta_k \ln d_{ki} + \varepsilon_i \quad (1)$$
$$\text{with } \varepsilon_i = v_i + u_i ,$$

where C represents current municipal spending in the selected functions (*EXP*), q_m are the corresponding output indicators (*POP*, *WASTE*, *DEPEND*, *ROAD*), and d_k are other structural variables controlling for the quality of public output (*INCOME*) and the presence of potential scale effects due to population size (*POP-1000* and *POP-10,000*) and geographical altitude (*ALT-600*). To estimate the SFA model (1), we rely on the maximum-likelihood technique proposed by Battese and Coelli (1995, called SFA-BC95 from now on), and assume the one-sided inefficiency term to be distributed as a truncated normal: $u_i \sim N^+(\eta' Z, \sigma_u^2)$. This specification allows the mean of spending inefficiency to be affected – through the vector of coefficients η to be estimated – by a set of observable exogenous factors Z, which includes tax autonomy and the other determinants discussed above. Finally, the symmetric random noise component v_i is assumed to be distributed as a standard $N(0, \sigma_v^2)$.

For the investigation of spending inefficiency within the DEA framework, we rely instead on a standard two-stage procedure (Coelli et al., 2005). Following De Borger and Kerstens (1996), in a first stage we compute an inefficiency score for each municipality, by fitting a *variable returns to scale* DEA model – the so-called DEA-VRS frontier, which allows us to take into account the presence of both scale economies and scale diseconomies in the production technology[17] – with the aggregate value of spending (*EXP*) used as input indicator and the proxies for the quantity (*POP*, *WASTE*, *DEPEND*, *ROAD*) and the quality (*INCOME*) of the services provided used as output indicators. Then, in the second stage, we take DEA-VRS inefficiency scores and regress them on the same set Z of inefficiency determinants specified in the SFA-BC95 model. The second-stage analysis relies on a tobit regression model, a censored model that allows us to make proper inference on the factors driving the estimated inefficiency, considering that, in the DEA framework, fully efficient municipalities show a value of 0 and no values below 0 can be observed.[18]

17 The generalization of DEA technique to the case of variable returns to scale – which is the most adopted approach in the literature since the early '90s – is due to the contribution of Banker et al. (1984). The orginal approach by Charnes et al. (1978) implicitly assumes constant returns to scale (DEA-CRS model) in the production technology.

18 Recent developments in DEA (Simar and Wilson, 2007) permit one to estimate the efficiency levels conditionally on the influence of exogenous variables Z, without assuming different distributions for the scores in the two stages of the analysis (which represents the main shortcoming of the standard tobit procedure). However, the implementation of this methodology is not essential in our context, as we check the robustness of our results by relying on both SFA and DEA to assess productive efficiency.

Table 4a

Preliminary Analysis of Spending Determinants: OLS Estimates for Each Sector Separately

Regressor	Dep. var. = EXP_{GA}		Dep. var. = EXP_{WM}		Dep. var. = EXP_{EE}		Dep. var. = EXP_{RM}	
	BASIC MODEL	EXTENDED MODEL	BASIC MODEL	EXTENDED MODEL	BASIC MODEL	EXTENDED MODEL	BASIC MODEL	EXTENDED MODEL
POP	0.820 (0.025)***	0.837 (0.027)***	–	–	–	–	–	–
WASTE	–	–	0.621 (0.089)***	0.606 (0.088)***	–	–	–	–
DEPEND	–	–	–	–	1.024 (0.041)***	1.006 (0.043)***	–	–
ROAD	–	–	–	–	–	–	0.294 (0.052)***	0.336 (0.053)***
INCOME	0.232 (0.146)	0.373 (0.179)**	0.852 (0.423)**	0.657 (0.476)	0.512 (0.302)*	0.278 (0.306)	1.576 (0.451)***	1.073 (0.505)**
POP-1000	0.112 (0.045)**	0.119 (0.046)***	−0.084 (0.114)	−0.071 (0.124)	−0.104 (0.072)	−0.100 (0.072)	0.741 (0.105)***	0.674 (0.106)***
POP-10,000	0.032 (0.044)	0.091 (0.047)**	0.292 (0.110)***	0.308 (0.099)***	0.191 (0.089)**	0.141 (0.071)**	0.559 (0.108)***	0.402 (0.115)***
ALT-600	0.113 (0.036)***	0.067 (0.044)	0.223 (0.092)**	0.148 (0.108)	−0.066 (0.068)	0.031 (0.085)	0.146 (0.104)	0.398 (0.125)***
FISCAUT	−0.368 (0.073)***	−0.352 (0.082)***	−0.154 (0.163)	−0.086 (0.172)	−0.237 (0.116)**	−0.205 (0.097)**	0.020 (0.200)	0.017 (0.207)
FISCAUT × HTAX	0.044 (0.005)***	0.042 (0.005)***	0.006 (0.014)	0.004 (0.015)	0.018 (0.010)*	0.017 (0.010)*	0.021 (0.016)	0.021 (0.016)
FISCAUT × HEXTRA	0.019 (0.005)***	0.015 (0.005)***	0.001 (0.015)	0.009 (0.016)	0.005 (0.008)	0.009 (0.008)	−0.030 (0.016)*	−0.027 (0.015)*
FISCAUT × HGRANT	0.006 (0.007)	0.006 (0.007)	−0.043 (0.037)	−0.039 (0.028)	−0.030 (0.012)**	−0.026 (0.012)**	−0.042 (0.021)**	−0.035 (0.022)
PACT	0.039 (0.040)	0.063 (0.054)	−0.405 (0.101)***	−0.344 (0.132)***	−0.014 (0.072)	−0.068 (0.090)	−0.429 (0.094)***	−0.310 (0.119)***
KMTO	–	0.047 (0.033)	–	0.201 (0.094)**	–	0.094 (0.054)*	–	0.067 (0.095)
DENS	–	−0.035 (0.020)*	–	0.002 (0.041)	–	0.059 (0.036)*	–	0.200 (0.060)***
YGOV	–	0.010 (0.013)	–	−0.008 (0.026)	–	0.014 (0.024)	–	0.020 (0.040)
YGOV × PACT	–	0.920 (0.439)**	–	1.814 (0.909)**	–	1.441 (0.701)**	–	0.605 (1.086)
YGOV × PACT × FISCAUT	–	−0.205 (0.096)**	–	−0.396 (0.197)**	–	−0.312 (0.152)**	–	−0.124 (0.235)
CIVIC	–	−0.001 (0.030)	–	−0.009 (0.073)	–	0.032 (0.048)	–	−0.048 (0.074)
LEFT	–	0.020 (0.032)	–	−0.028 (0.086)	–	0.027 (0.052)	–	−0.225 (0.091)**
SEXMAYOR	–	−0.027 (0.026)	–	0.071 (0.060)	–	0.034 (0.058)	–	0.022 (0.083)
AGEMAYOR	–	0.031 (0.054)	–	−0.244 (0.163)	–	−0.134 (0.100)	–	−0.415 (0.161)***
PUBLIC	–	−0.002 (0.025)	–	−0.163 (0.103)	–	–	–	–
PUBLIC × FIRM	–	0.073 (0.051)	–	0.259 (0.256)	–	–	–	–
PUBLIC × FIRM × COOP	–	−0.122 (0.052)**	–	−0.221 (0.055)***	–	–	–	–
No. of observations	262		262		262		262	
R^2	0.95	0.96	0.77	0.78	0.94	0.94	0.70	0.72
F test [*p*-value]	711.76 [0.000]	349.87 [0.000]	159.84 [0.000]	128.73 [0.000]	411.45 [0.000]	269.06 [0.000]	96.42 [0.000]	52.19 [0.000]

Note: All variables have been transformed with natural logarithm; robust standard errors are reported in parentheses; significance level: *** 1%, ** 5%, * 10%.

3. Results

3.1. Preliminary OLS Estimates

In table 4a we report the findings of the explorative investigation on spending determinants carried out for each function separately, while table 4b shows the OLS estimates for the aggregate current expenditure. The results refer both to a basic model specification, where only tax autonomy and other fiscal indicators are considered, and an extended model specification, which includes also the spatial, political, and organizational variables among the possible drivers of spending performance.

Looking at the sector-by-sector estimates, one can first notice that each spending item is positively and significantly driven by the corresponding output proxy, with the highest correlation observed for education and elderly care (the effect of *DEPEND* on EXP_{EE} is slightly greater than 1) and the lowest correlation in the case of road maintenance and local mobility (β_{ROAD} ranges between 0.29 and 0.34). This clearly provides an empirical support to our choice of output variables in terms of proxies for citizens' needs. Second, the effect of output quality on expenditure turns out to be important, as *INCOME* plays a major role in almost all the estimated models, with a positive and significant effect that appears particularly large for EXP_{RM}. A possible explanation for this evidence is that the increase in income makes citizens more demanding toward road maintenance, asking their mayors, for instance, to promptly repair broken pavements or roads, and to improve urban living areas. Population size and altitude are also important in affecting spending. The positive and significant coefficients for the dummies *POP-1000*, *POP-10,000*, and *ALT-600* in most specifications (again, remarkably for EXP_{RM} in the extended model) points to the presence of some adverse scale effect for the smallest and the biggest municipalities (notice, in particular, the presence in all sectors of a statistically significant cost increase for local governments with more than 10,000 inhabitants), as well as for the mountain (and touristic) resorts.

Turning attention now to the influence of tax autonomy and the other fiscal constraints, *FISCAUT* appears to play an important role in the provision of general administration services (EXP_{GA}) and care services (EXP_{EE}), two categories for which the estimated coefficients show a negative and statistically significant sign: this finding supports the theoretical argument that higher accountability of local politicians can be reached by raising their responsibilities for funding; indeed, the higher is the share of current revenues derived from own taxes, the lower is the spending level of local government, even if we are not yet able in this preliminary stage to attribute the decrease observed in expenditure to a reduction of excess costs compared to citizens'

needs (for which a frontier analysis is required; see next section).[19] For the same spending categories, given the positive sign of the coefficients associated with the interaction of *FISCAUT* with *HTAX* and *HEXTRA*, our results also show that fiscal incentives due to revenue autonomy are partially offset by spending increases when local taxes per capita, as well as fees and charges in the case of EXP_{GA}, are higher than the median. This evidence supports the existence of opportunistic behavior by politicians (discussed also in Balaguer-Coll et al., 2007) when a local government can rely on a large amount of own revenues. The external imposition of a tighter budget constraint – such as the limit on spending growth established by the DSP for the municipalities with more than 5000 inhabitants – seems to be effective in motivating incumbent politicians to better control their expenditures, at least in some functions like waste management or road maintenance and local mobility, for which the estimated coefficient of *PACT* is negative and statistically significant.

When considering the extended model, results on the effects of fiscal variables are very similar to those discussed above, showing their robustness to alternative specifications of spending functions. But this model offers other interesting insights into the effects associated with the spatial, political, and organizational factors. There is evidence of positive spillovers from being closer to Turin, since we observe a decrease in spending (statistically significant only for waste management and for education and elderly care) when the distance from Turin (*KMTO*) reduces. As for the effect of a higher population density (*DENS*), potential congestion effects seem to prevail over density economies in the sectors of education and elderly care and of road maintenance and local mobility, while the reverse is observed for general administration services.

Looking at the political features of the municipalities, the variable measuring the number of years after election for the mayor and the governing coalition (*YGOV*) exerts a significant influence on spending decisions only when it is interacted with the fiscal indicators *PACT* and *FISCAUT*: this reveals that the opportunistic behavior of incumbent politicians highlighted by the empirical literature on the electoral budget cycle (e.g., Galli and Rossi, 2002; Veiga and Veiga, 2007) is conditioned by the presence of fiscal constraints imposed on the local government. More precisely, the positive coefficient of $YGOV \times PACT$ (statistically significant for EXP_{GA}, EXP_{WM}, and EXP_{EE}) confirms the recent findings by Mink and De Haan (2005) and Bartolini and Santolini (2009) of a strong electoral-budget-cycle effect for the municipalities subject to a fiscal discipline rule: the introduction of the

19 Indeed, our evidence is consistent with the results of the empirical studies that consider how tax decentralization affects local governments' size, without separating inefficient spending from its structural component (see section 1).

DSP provides incentive for opportunistic spending by incumbents who are closer to the end of their mandate; this could be due to forward-looking behavior, which leads incumbents to intensify compliance to DSP in the early years of their terms, so as to exploit higher margins for increasing the expenditure when close to new elections. We also consider the interaction $YGOV \times PACT \times FISCAUT$, to study the interplay between the fiscal constraints and the degree of tax autonomy in influencing the electoral budget cycle. Interestingly, one can notice that higher revenue autonomy has the effect of dampening the electoral budget cycle's effect on spending observed for the municipalities under the DSP, hence increasing the importance of the argument for stronger accountability of local governments obtainable through tax decentralization. The control for the political orientation of the government points to negative and significant effects on spending for center–left coalitions ($LEFT$) and for older mayors ($AGEMAYOR$), although these variables seem to influence only the expenditure on road maintenance and local mobility.

The last set of regressors included in the extended model aims at controlling for the influence of different organizational choices for waste management on the expenditures in that sector and, possibly, in general administration. The results indicate a significant effect only for the dummy $PUBLIC \times FIRM \times COOP$: neither the public versus private ownership of the firm nor the externalization of the service seems to matter in itself; it is instead relevant that, besides being publicly owned and run through a firm, garbage collection and disposal is managed cooperatively. The organizational scheme of the publicly owned cooperative firm would then represent the most efficient solution, in terms of reduced spending both for waste management and for general administration. These cost savings are likely to result from the advantage of sharing large fixed costs (typical of the consortium option) combined with the benefit of increasing expenditure control (typical of the external-firm option).

The estimates for aggregate current expenditure in the four selected functions (table 4b) generally confirm the main findings discussed above of separate models for each spending category. Considering the multi-output production function underlying this global specification of the cost model, one can notice that population size is the most important proxy of citizens' needs for explaining variability in current spending (β_{POP} is about 0.66), while $WASTE$, $DEPEND$, and $ROAD$ play secondary roles. Moreover, constant returns to scale seem to dominate the aggregate provision of municipal services, as the sum of estimated elasticities with respect to the four output indicators is very close to one (it ranges between 0.92 and 0.93). Notice, however, that this result crucially depends on the fact that 61% of the municipalities in our sample do not belong to extreme size

Table 4b

Preliminary Analysis of Spending Determinants: OLS Estimates for the Aggregated Current Expenditure

Regressor	Dep. var. = *EXP*[a]	
	BASIC MODEL	EXTENDED MODEL
POP	0.669 (0.056)***	0.655 (0.058)***
WASTE	0.167 (0.039)***	0.173 (0.039)***
DEPEND	0.063 (0.032)**	0.068 (0.033)**
ROAD	0.024 (0.010)**	0.026 (0.010)**
INCOME	0.356 (0.101)***	0.402 (0.115)***
POP-1000	0.066 (0.026)**	0.065 (0.026)**
POP-10,000	0.078 (0.038)**	0.115 (0.029)***
ALT-600	0.053 (0.022)**	0.046 (0.026)*
FISCAUT	−0.450 (0.042)***	−0.429 (0.044)***
FISCAUT × *HTAX*	0.041 (0.003)***	0.039 (0.003)***
FISCAUT × *HEXTRA*	0.012 (0.003)***	0.011 (0.003)***
FISCAUT × *HGRANT*	−0.006 (0.004)	−0.005 (0.004)
PACT	−0.025 (0.023)	−0.057 (0.031)*
KMTO	–	0.004 (0.019)
DENS	–	−0.010 (0.012)
YGOV	–	0.010 (0.008)
YGOV × *PACT*	–	1.039 (0.281)***
YGOV × *PACT* × *FISCAUT*	–	−0.232 (0.061)***
CIVIC	–	−0.013 (0.017)
LEFT	–	−0.017 (0.019)
SEXMAYOR	–	−0.003 (0.015)
AGEMAYOR	–	−0.026 (0.034)
PUBLIC	–	−0.010 (0.015)
PUBLIC × *FIRM*	–	0.023 (0.030)
PUBLIC × *FIRM* × *COOP*	–	−0.049 (0.020)**
No. of observations	262	
R^2	0.99	0.99
F test [*p*-value]	1845.57 [0.000]	1134.63 [0.000]

Notes: All variables have been transformed to natural logarithms; robust standard errors are reported in parentheses; significance level: *** 1%, ** 5%, * 10%. [a] $EXP = EXP_{GA} + EXP_{WM} + EXP_{EE} + EXP_{RM}$.

classes (under 1000 and over 10,000 inhabitants) and are located below 600 meters in altitude. Indeed, as already observed before, *EXP* significantly increases for the groups of the smallest (*POP-1000*) and the biggest (*POP-10,000*) municipalities, as well as for those above 600 meters (*ALT-600*), thus suggesting the likely presence of scale economies/diseconomies in the

production of municipal services, which will be discussed in more detail in the next section, when comparing results from SFA and DEA frontier estimations. Aggregate current spending also shows a positive elasticity with respect to the average level of municipal income (β_{INCOME} ranges between 0.36 and 0.40), thus confirming the importance of controlling for the demand for higher-quality services in richer communities. Looking at the effects of fiscal indicators, the coefficient of tax autonomy (*FISCAUT*) has the expected negative sign – again partially offset by a relatively small increase in spending for the municipalities with a per capita level of own revenues (both taxes and fees and charges) higher than the sample median – and supports the argument of higher accountability of local governments induced by the increase of their funding responsibility. When the model is extended so as to include the spatial, political, and organizational variables, the introduction of DSP (*PACT*) exerts a statistically significant effect in containing spending.[20]

Finally, also the link of the electoral budget cycle with fiscal constraints and the higher efficiency of managing waste collection through a publicly owned cooperative firm are confirmed. This analysis does not allow us, however, to understand whether the degree of fiscal autonomy affects structural expenditure – which is related to citizens' needs – or inefficient spending. To do so, we move a step further and consider frontier models.

3.2. Analysis of Spending Inefficiency

Table 5 reports summary statistics for the inefficiency scores obtained with the SFA, considering both the basic and the extended specification of the model $\eta' Z$ for the mean of u_i in equation (1) and the DEA-VRS models. The average inefficiency is between 0.24 (extended model) and 0.26 (basic model) for SFA and about 0.20 for DEA-VRS[21], which means that municipalities, on average, could satisfy citizens' needs with respect to the analyzed services with a 20–26% reduction in the current level of spending. The distributions of inefficiency levels appear concentrated around the mean in both SFA

20 We note that this result is robust to a different specification of the aggregate spending function, in which returns to scale are allowed to fully vary across the municipalities. In particular, we replaced the dummies for extreme size classes (*POP-1000* and *POP-10,000*) with a quadratic term for the population. According to this specification, the estimated coefficient for *PACT* is net of possible confounding effects due to population size.

21 It is worth remarking that the estimation of the efficient frontier within the DEA framework is based on the identification of a group of fully efficient municipalities (for which the inefficiency score is equal to zero: 33 cases in our analysis) that are used as benchmarks for assessing the performance of the other units of the sample. Therefore, by construction, the average efficiency (inefficiency) computed using DEA is typically higher (lower) than the values resulting from SFA estimation (see, e.g., Coelli et al., 2005).

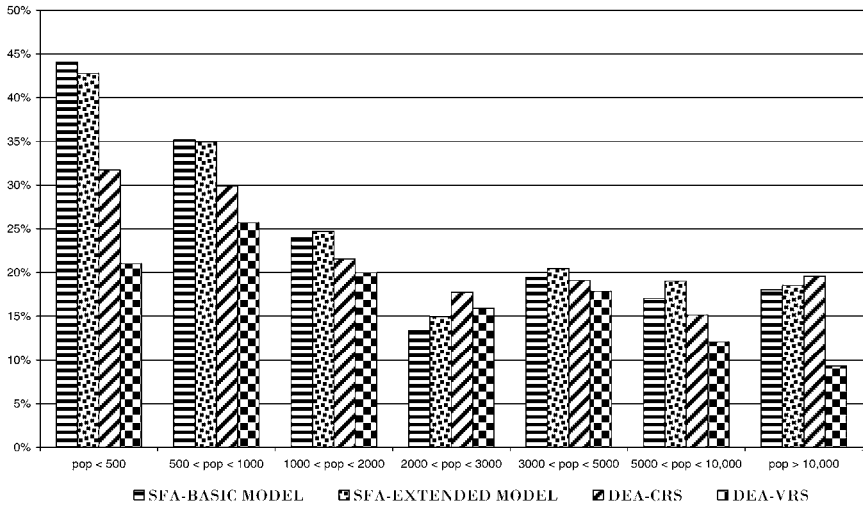
Table 5
Analysis of Spending Inefficiency: Summary Statistics for SFA and DEA Scores

	SFA BASIC MODEL	SFA EXTENDED MODEL	DEA-VRS
Mean	0.26	0.24	0.20
Standard deviation	0.17	0.17	0.12
Min	0.02	0.01	0.00
25th percentile	0.12	0.10	0.12
Median	0.25	0.22	0.20
75th percentile	0.35	0.33	0.28
Max	0.93	0.94	0.52
Fully efficient municipalities	–	–	33

and DEA models, as they exhibit a median value very close to their mean, and 75% of observations show a spending inefficiency lower than 0.33–0.35 using SFA and 0.28 using DEA. Not surprisingly, the standard deviation is generally small and is larger for SFA estimates, due to the presence of more extreme scores (maximum values are around 0.93–0.94, against 0.52 in the DEA model).

More importantly, the correlation between SFA and DEA inefficiencies is very high (ranging between 0.7 and 0.8) for both DEA-VRS and DEA-CRS models (see footnote 17). As discussed above, the inclusion of population size and altitude dummies in the cost frontier (1) helps control for the effects of variable returns to scale on efficiency estimates, as in a DEA-VRS framework, even if these effects do not vanish completely. Indeed, also for SFA parameter estimates the sum of elasticities with respect to the four output proxies is very close to one (0.94 in the basic model and 0.93 in the extended model; see table 6a), highlighting a multi-output production technology mainly characterized by constant returns to scale, which is the basic assumption of a DEA-CRS model. Such a result is probably driven by the prevalence in our sample of medium-sized municipalities, for which returns to scale appear to be actually constant on looking at the difference in inefficiency levels between DEA-CRS and DEA-VRS (figure 1). Variable returns to scale seem instead to characterize municipalities under 1000 and over 10,000 inhabitants (notice that the coefficients of *POP-1000* and *POP-10,000* in SFA estimates are both positive and statistically significant). More precisely, the former mainly exhibit increasing returns to scale, perhaps because of the stronger influence of fixed costs on current spending (especially with regard to waste management and general administration

Figure 1
Distribution of SFA and DEA Average Inefficiency by Municipal Size Classes



services), while the latter mostly show decreasing returns to scale, probably because they produce a wider range of more complex services (this is particularly true for social-welfare spending). As for the definition of the proper scale for providing the essential services analyzed in this study, municipalities with a number of served inhabitants from 3,000 to 5,000 apparently correspond to the optimal size. This size emerges on looking at both the differences between DEA-CRS and DEA-VRS scores and the SFA inefficiencies in figure 1. It is also worth noticing that, in the DEA-VRS model, spending inefficiency (net of scale inefficiency) tends to decrease with increasing municipal size. In the light of the positive correlation between municipal size and degree of tax autonomy observed in our sample (0.62), a possible interpretation of this evidence relies on the argument that local politicians are probably subject to more severe control from their citizens when the latter can ask for differentiated and more effective services that are financed to a significant extent through taxes collected at local level. To explore this issue more in depth, we turn now our attention to the investigation of the factors that could help explain the estimated inefficiency.

Tables 6a and 6b show the estimates of the SFA-BC95 frontier model (1) and DEA-VRS tobit regression, respectively, relying on the same vector Z

Table 6a

Analysis of Spending Inefficiency: SFA-BC95 Parameter Estimates

Regressor	**Dep. var. = *EXP*[a]**	
	BASIC MODEL	EXTENDED MODEL
Frontier variables		
POP	0.697 (0.044) ***	0.697 (0.046) ***
WASTE	0.156 (0.027) ***	0.155 (0.026) ***
DEPEND	0.060 (0.029) **	0.059 (0.029) **
ROAD	0.023 (0.010) **	0.022 (0.010) **
INCOME	0.324 (0.098) ***	0.387 (0.102) ***
POP-1000	0.070 (0.023) ***	0.068 (0.022) ***
POP-10,000	0.077 (0.036) **	0.110 (0.034) ***
ALT-600	0.045 (0.020) **	0.036 (0.021) *
Wald test$_{(8)}$ [*p*-value]	8156.81 [0.000]	9338.74 [0.000]
Inefficiency determinants		
FISCAUT	−0.528 (0.054) ***	−0.529 (0.052) ***
FISCAUT × *HTAX*	0.052 (0.007) ***	0.053 (0.006) ***
FISCAUT × *HEXTRA*	0.015 (0.004) ***	0.014 (0.004) ***
FISCAUT × *HGRANT*	−0.002 (0.005)	−0.002 (0.005)
PACT	0.007 (0.032)	−0.047 (0.047)
KMTO	–	0.003 (0.024)
DENS	–	−0.014 (0.013)
YGOV	–	0.012 (0.007) *
YGOV × *PACT*	–	1.325 (0.527) ***
YGOV × *PACT* × *FISCAUT*	–	−0.296 (0.116) ***
CIVIC	–	−0.044 (0.020) **
LEFT	–	−0.050 (0.023) **
SEXMAYOR	–	−0.014 (0.019)
AGEMAYOR	–	−0.014 (0.035)
PUBLIC	–	−0.017 (0.018)
PUBLIC × *FIRM*	–	0.031 (0.033)
PUBLIC × *FIRM* × *COOP*	–	−0.066 (0.034) **
$\sigma^2 = \sigma_u^2 + \sigma_v^2$	0.010 (0.001) ***	0.009 (0.001) ***
$\gamma = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$	0.587 (0.190) ***	0.597 (0.174) ***
LR test$_{(r)}$ [*p*-value]	181.42$_{(7)}$ [0.000]	205.93$_{(19)}$ [0.000]

Notes: Number of observations: 262; all variables have been transformed to natural logarithms; standard errors are reported in parentheses; significance level: *** 1%, ** 5%, * 10%.
[a] $EXP = EXP_{GA} + EXP_{WM} + EXP_{EE} + EXP_{RM}$.

of inefficiency determinants (basic and extended specifications) used in the preliminary OLS analysis.

Table 6b
Analysis of Spending Inefficiency: Tobit Parameter Estimates

Regressor	Dep. var. = *DEA-VRS scores*[a]	
	BASIC MODEL	EXTENDED MODEL
FISCAUT	−0.272 (0.035) ***	−0.259 (0.038) ***
FISCAUT × *HTAX*	0.025 (0.003) ***	0.025 (0.004) ***
FISCAUT × *HEXTRA*	0.009 (0.003) ***	0.007 (0.003) **
FISCAUT × *HGRANT*	−0.006 (0.004)	−0.004 (0.004)
PACT	−0.062 (0.020) ***	−0.043 (0.021) **
KMTO	–	0.029 (0.024)
DENS	–	0.012 (0.011)
YGOV	–	0.007 (0.004) **
YGOV × *PACT*	–	1.599 (0.294) ***
YGOV × *PACT* × *FISCAUT*	–	−0.352 (0.065) ***
CIVIC	–	−0.021 (0.017)
LEFT	–	−0.027 (0.021)
SEXMAYOR	–	0.006 (0.017)
AGEMAYOR	–	−0.057 (0.028) **
PUBLIC	–	0.013 (0.016)
PUBLIC × *FIRM*	–	−0.022 (0.036)
PUBLIC × *FIRM* × *COOP*	–	−0.018 (0.008) **
F test [*p*-value]	28.73 [0.000]	12.92 [0.000]

Notes: Number of observations: 262; all variables have been transformed to natural logarithms; robust standard errors are reported in parentheses; significance level: *** 1%, ** 5%, * 10%.
[a] Inefficiency scores are computed using *EXP* as input variable and *POP*, *WASTE*, *DEPEND*, *ROAD*, and *INCOME* as output variables in a DEA-VRS frontier model.

First, the SFA parameter estimates of frontier variables (output proxies and other control factors) are very similar to those discussed above for the average spending function (table 4b), both in magnitude and in statistical significance, thus confirming that these are important drivers of the structural component of aggregate current expenditure in the selected functions. Most importantly, the model highlights the importance of the spending inefficiency u_i in comparison with random noise v_i in determining the global error term ε_i: the parameter γ – the share of residual variance due to deviations from the best-practice frontier, $\sigma_u^2/(\sigma_u^2 + \sigma_v^2)$ – is very close to 0.60 in both specifications. This evidence supports the hypothesis that the municipalities of our sample are not cost-minimizing and a traditional *average* spending function with the term u_i equal to zero does not adequately represent the observed performance. Looking now at the parameter estimates of

inefficiency determinants, SFA-BC95 and tobit models perform both well in terms of general statistical fit (as indicated by LR and F tests) and provide similar results for the sign and the significance of most coefficients, showing that our findings are robust to alternative methodological approaches for analyzing spending inefficiency.

As for the key issue of our study, the sign of the coefficient of *FISCAUT* is negative and highly statistically significant in all models. Hence, the negative impact on expenditure stemming from a greater tax autonomy of municipalities targets *inefficient* spending, i.e., the waste of resources with respect to the amount required to satisfy citizens' needs. This finding supports the theoretical prediction of SGT that higher accountability of local politicians can be obtained by reducing VFI (e.g., Weingast, 2009), specifying some previous results in the literature that highlighted the positive effects of higher local tax rates on municipal efficiency (Vanden Eeckaut et al., 1993; De Borger et al., 1994; De Borger and Kerstens, 1996). At the same time, both SFA-BC95 and tobit estimates confirm that large availability of own resources – as reflected in per capita levels higher than the sample median – counteract the incentive effect of fiscal decentralization: the interactions of *FISCAUT* with *HTAX* and *HEXTRA* show positive and significant coefficients, even if the first-order effect is much larger (see Balaguer-Coll et al., 2007). Finally, also the DSP seems to work well as a mechanism of fiscal discipline, leading to cuts in excess spending, although the reducing effect on inefficiency is statistically significant only in the tobit model, probably because the variable *PACT* partly captures a size effect in SFA models.[22]

As for the role played by the other variables included in the extended model, the two spatial indicators are never significant (as in the OLS estimates of table 4b); indeed, at the aggregate spending level, it is likely that positive and negative spillovers associated with a higher proximity to the capital (*KMTO*) are compensated, and the same holds for congestion effects and cost savings resulting from increased population density (*DENS*). Looking at the political variables, there emerges now a positive and significant coefficient also for the first-order effect of the shorter time period before new elections (*YGOV*), in both SFA and tobit estimates. However, the effect of the electoral budget cycle conditional on the presence of the DSP and its interaction with tax autonomy continue to play the major role, as shown by the magnitude of the parameters associated with the interactions of *YGOV* with *PACT* and *FISCAUT*. The relevant aspect of our results compared to the existing evidence on the electoral budget cycle in local governments (e.g.,

22 Recall that the variability of returns to scale over the sample is not fully controlled in our SFA model specification, where only the effect of extreme size classes on spending is taken into account through the dummies *POP-1000* and *POP-10,000*, while the nonparametric DEA framework allows returns to scale to vary freely across all productive units.

Veiga and Veiga, 2007; Bartolini and Santolini, 2009) is that they clarify that the observed increase in municipal spending when close to new elections can be interpreted as a greater waste with respect to an efficient expenditure level.[23] As for the political orientation, the coefficient of the dummy for center–left coalitions (*LEFT*) is significant and negative, as well as the one associated with the presence of a civic list (*CIVIC*), even if only in SFA estimates. The effect of *LEFT* on spending inefficiency can be added to the existing political-economy literature, which often found a propensity of left-wing governments towards larger expenditure (e.g., Blais et al., 1993; De Haan and Sturm, 1994): observing larger spending carried out by left-wing coalitions does not imply higher inefficiency; indeed, the latter even seems to decrease.

Finally, the presence of older mayors significantly reduces inefficiency (*AGEMAYOR*), at least in tobit estimates. Again, this result adds to the recent literature on the effects of term length on politicians' behavior, which points out a positive role of older representatives in determining good legislative performance of governments (e.g., Dal Bò and Rossi, 2008). As for the effects of the organizational choice for waste management, both SFA and tobit results imply that higher efficiency in spending can be reached if garbage collection and disposal is provided by a publicly owned cooperative firm.

4. Conclusions

This paper studies the role played by tax decentralization (measured as the degree of VFI, i.e., the fiscal autonomy in covering the costs associated with the provision of essential public services) in influencing the spending efficiency of local governments. The study relies on a sample of 262 Italian

23 A control for the robustness of these findings is also provided by introducing the dummy *2GOV*, which distinguishes the municipalities with a mayor facing a second-term limit from those with a mayor who can be reelected. The theoretical literature suggests that impossibility of reelection influences the opportunistic behavior of the incumbents, especially in proximity to new elections (e.g., Besley and Case, 1995, 2003; Smart and Sturm, 2006). However, testing the effect of this variable – considered alone or interacted with the years of mandate – we did not observe any significant effect on the inefficiencies. A plausible explanation for this evidence can be found by analyzing more in depth the municipalities included in our sample: the dummy *2GOV* is equal to one for 26.5% of these municipalities; among these, in the subsequent elections (between 2006 and 2009), a person belonging to the previous governing coalition was elected mayor in 58% of the cases; in another 22%, a person belonging to the previous governing coalition was presented as one of the main candidates to become mayor. These statistics stress the role of a party affiliation and party discipline in identifying candidates and their behavior once elected. Therefore, incumbents' interests, merging in the party's ones, do not vanish simply with their impossibility of reelection.

municipalities belonging to the province of Turin and exploits both standard regression analysis and efficiency-frontier techniques (SFA and DEA) to study local governments' spending performance and its main determinants, considering four main spending categories: general administration, waste management, education and elderly care, and road maintenance and local mobility.

Consistently with modern fiscal federalism theories, our results show that more autonomous municipalities – i.e., local governments with a low VFI – exhibit lower spending for satisfying citizens' needs, thus supporting the argument that effective electoral accountability of local politicians can be obtained by increasing their responsibilities with respect to funding. We also find some evidence on the possibility for the central government to control spending efficiency through fiscal rules (here the Domestic Stability Pact). The analysis was then extended to allow for the role played by other potential factors, including spatial, political, and organizational variables. Among these, a major influence on spending performance seems to be exerted by the opportunistic behavior of incumbent politicians, as highlighted by theoretical literature on the electoral budget cycle. We provide new empirical support for the (inefficient) increase of spending observed for mayors closer to new elections, and we also find that this effect is strongly conditioned by the presence of fiscal restraints imposed on local governments (i.e., the DSP) and by the degree of accountability deriving from their fiscal autonomy.

From a policy perspective, the evidence emerging in this study supports the recent waves of reforms towards the devolution of taxing power to lower government tiers – from regions to municipalities – observed in Italy as well as in other countries around the world, with the purpose of reducing VFI and increasing the accountability of local politicians and, through this mechanism, improving both the efficiency and the effectiveness of public services provided to the citizens.

References

Afonso, A., and Fernandes, S. (2005), Local Government Spending Efficiency: DEA Evidence for the Lisbon Region, Regional Studies 40, 39–53.

Akai, N., and Sakata, M. (2002), Fiscal Decentralization Contributes to Economic Growth: Evidence from State-Level Cross-Section Data for the United States, Journal of Urban Economics 52, 93–108.

Ambrosanio, F., and Bordignon, M. (2009), Internal Stability Pacts: The European Experience, European Economic Governance Monitor, EEGM Paper 4, European Economic Governance Monitor.

Baicker, K. (2005), The Spill-Over Effects of State Spending, Journal of Public Economics 89, 529–544.

Balaguer-Coll, M. T., Tortosa-Ausina, E., and Prior, D. (2007), On the Determinants of Local Government Performance: A Two Stage Nonparametric Approach, European Economic Review 51, 425–451.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984), Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, Management Science 30, 1078–1092.

Bartolini, D., and Santolini, R. (2009), Fiscal Rules and the Opportunistic Behavior of the Incumbent Politician: Evidence from Italian Municipalities, CESifo Working Paper No. 2605.

Baskaran, T. (2010), On the Link between Fiscal Decentralization and Public Debt in OECD Countries, Public Choice 145, 351–378.

Battese, G. E., and Coelli, T. J. (1995), A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, Empirical Economics 20, 325–232.

Bergstrom, T. C., and Goodman, R. P. (1973), Private Demands for Public Goods, American Economic Review 63, 280–296.

Besley, T. (2006), Principled Agents?, Oxford University Press, Oxford.

Besley, T., and Case, A. (1995), Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits, Quarterly Journal of Economics 110, 769–798.

Besley, T., and Case, A. (2003), Political Competition and Policy Choices: Evidence from the United States, Journal of Economic Literature 41, 7–73.

Blais, A., Blake, D., and Dion, S. (1993), Do Parties Make a Difference? Parties and the Size of Governments in Liberal Democracies, American Journal of Political Science 37, 40–62.

Blöchliger, H., and Rabesona, J. (2009), The Fiscal Autonomy of Sub-Central Governments: An Update, OECD Working Paper 9, OECD Network on Fiscal Relations Across Levels of Government, Paris.

Bordignon, M., and Tabellini, G. (2009), Moderating Political Extremism: Single Round vs Runoff Elections under Plurality Rule, CESifo Working Paper No. 2600.

Borge, L. E., and Rattsø, J. (2008), Property Taxation as Incentive for Cost Control: Empirical Evidence for Utility Services in Norway, European Economic Review 52, 1035–1054.

Brunila, A., Buti, M., and Franco D. (2001), The Stability and Growth Pact, Palgrave Macmillan.

Charnes, A., Cooper, W., and Rhodes, E. (1978), Measuring the Efficiency of Decision Making Units, European Journal of Operational Research 2, 429–444.

Chattopadhyay, R., and Duflo, E. (2004), Women as Policy Makers: Evidence from a Randomized Policy Experiment in India, Econometrica 72(5), 1409–1443.

Coelli, T. J., Prasada Rao, D. S., O'Donnell, C. J., and Battese, G. E. (2005), An Introduction to Efficiency and Productivity Analysis, 2nd Edition, Springer, New York.

Dal Bó, E., and Rossi, M. (2008), Term Length and Political Performance, NBER Working Paper 14511, Cambridge, MA.

De Borger, B., and Kerstens, K. (1996), Cost Efficiency of Belgian Local Governments: A Comparative Analysis of FDH, DEA, and Econometric Approaches, Regional Science and Urban Economics 26, 145–170.

De Borger, B., Kerstens, K., Moesen, W., and Vanneste, J. (1994), Explaining Differences in Productive Efficiency: An Application to Belgian Municipalities, Public Choice 80, 339–358.

De Haan, J., and Sturm, J. E. (1994), Political and Institutional Determinants of Fiscal Policy in the European Community, Public Choice 80, 157–172.

Edlund, L., and Pande, R. (2002), Why Have Women Become Left-Wing? The Political Gender and the Decline in Marriage, Quarterly Journal of Economics 117, 917–961.

Ermini, B., and Santolini, R. (2010), Local Expenditure Interaction in Italian Municipalities: Do Local Council Partnerships Make a Difference?, Local Government Studies 36, 655–677.

Eyraud, L., and Lusinyan, L. (2011), Decentralizing Spending More than Revenue: Does It Hurt Fiscal Performance?, IMF Working Paper No. 11/226.

Fiva, J. (2006), New Evidence on the Effect of Fiscal Decentralization on the Size and Composition of Government Spending, FinanzArchiv/Public Finance Analysis 62, 250–280.

Fox, K. J. (2001), Efficiency in the Public Sector. Studies in Productivity and Efficiency, Vol. 1, Kluwer Academic Publishers, Boston.

Funk, P., and Gathmann, C. (2008), Gender Gaps in Policy Making: Evidence from Direct Democracy in Switzerland, Working Paper, Universitat Pompeu Fabra.

Galli, E., and Rossi, S. (2002), Political Budget Cycles: The Case of the Western German Länder, Public Choice 110, 283–303.

Giménez, V., and Prior, D. (2007), Long- and Short-Term Cost Efficiency Frontier Evaluation: Evidence from Spanish Local Governments, Fiscal Studies 28, 121–139.

Giuriato, L., and Gastaldi, F. (2009), The Domestic Stability Pact in Italy: A Rule for Discipline?, MPRA Paper 15183.

Glaeser, E. (1996), The Incentive Effects of Property Taxes on Local Governments, Public Choice 89, 93–111.

Inman, R. P. (2003), Transfers and Bailouts: Enforcing Local Fiscal Discipline with Lessons from U. S. Federalism, in: Rodden, J. A., and Eskeland, G. S. (Eds.), Fiscal Decentralization and the Challenge of Hard Budget Constraints, MIT Press, Cambridge, MA, 35–83.

Jin, J., and Zou, H. (2002), How Does Fiscal Decentralization Affect Aggregate, National, and Subnational Government Size?, Journal of Urban Economics 52, 270–293.

Kornai, J., Maskin, E., and Roland, G. (2003), Understanding the Soft Budget Constraint, Journal of Economic Literature 41, 1095–1136.

Lokkainen, H., and Susiluoto, I. (2004), Cost Efficiency of Finnish Municipalities 1994–2002. An Application of DEA and Tobit Methods, Paper Presented at the 44th Congress of the European Regional Science Association, Porto, Portugal.

Lott, J. R., and Kenny, L. W. (1999), Did Women's Suffrage Influence the Size and Scope of Government?, Journal of Political Economy 107, 1163–1198.

Mink, M., and De Haan, J. (2005), Has the Stability and Growth Pact Impeded Political Budget Cycles in the European Union?, CESifo Working Paper No. 1535.

Oates, W. E. (1972), Fiscal Federalism, Harcourt Brace Jovanovich, New York, NY.

Oates, W. E. (1985), Searching for Leviathan: An Empirical Study, American Economic Review 75, 748–757.

Oates, W. E. (1999), An Essay on Fiscal Federalism, Journal of Economic Literature 37, 1120–1149.

Oates, W. E. (2005), Toward a Second-Generation Theory of Fiscal Federalism, International Tax and Public Finance 12, 349–373.

Osborne, M. J., and Slivinsky, A. (1996), A Model of Political Competition with Citizen–Candidates, Quarterly Journal of Economics 111, 65–96.

Prieto, A. M., and Zofio, J. L. (2001), Evaluating Effectiveness in Public Provision of Infrastructure and Equipment: The Case of Spanish Municipalities, Journal of Productivity Analysis 15, 41–58.

Reiter, M., and Weichenrieder, A. (1997), Are Public Goods Public? A Critical Survey of the Demand Estimates for Local Public Services, FinanzArchiv/Public Finance Analysis 54, 374–408.

Revelli, F. (2003), Reaction or Interaction? Spatial Process Identification in Multi-Tiered Government Structures, Journal of Urban Economics 53, 29–53.

Revelli, F. (2006), Spatial Interaction among Governments, in: Ahmad, E., and Brosio, G. (Eds.), Handbook in Fiscal Federalism, Edward Elgar, Cheltenham, UK.

Rodden, J. (2003), Reviving Leviathan: Fiscal Federalism and the Growth of Government, International Organization 57, 695–729.

Salmon, P. (1987), Decentralization as an Incentive Scheme, Oxford Review of Economic Policy 3, 24–43.

Simar, L., and Wilson, P. W. (2007), Estimation and Inference in Two Stage Semi-Parametric Models of Productive Efficiency, Journal of Econometrics 136, 31–64.

Smart, M., and Sturm, D. (2006), Term Limits and Electoral Accountability, CEP Discussion Paper No. 770.

Stegarescu, D. (2005), Public Sector Decentralization: Measurement Concepts and Recent International Trends, Fiscal Studies 26, 301–333.

Tiebout, C. (1956), A Pure Theory of Local Expenditures, Journal of Political Economy 64, 416–424.

Vanden Eeckhaut, P., Tulkens, H., and Jamar, M. A. (1993), Cost Efficiency in Belgian Municipalities, in: Fried, H. O., Lovell, C. A. K., and Schmidt, S. S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press.

Veiga, L. G., and Veiga, F. J. (2007), Political Business Cycles at the Municipal Level, Public Choice 131, 45–64.

Weingast, B. R. (2009), Second Generation Fiscal Federalism: The Implication of Fiscal Incentives, Journal of Urban Economics 65, 279–293.

Massimiliano Piacenza
University of Torino
School of Economics
Department of Economics and Statistics
Corso Unione Sovietica 218 bis
10134 Torino
ITALY
piacenza@econ.unito.it

Revenue Equalization Systems in a Federation with Tax Evasion

Lisa Grazzini and Alessandro Petretto*

Received 30 June 2010; in revised form 12 January 2012; accepted 10 May 2012

We analyze how vertical or horizontal fiscal equalization affects the overprovision of local public goods due to vertical fiscal externality, when there is tax evasion. The overspending incentive of regional governments is examined in the cases of fiscal equalization based on pretax earned income and on reported taxable income. We show that a more efficient level of regional public expenditure is associated with a vertical equalization scheme based on a proxy of earned income taken from national-accounts data.

Keywords: fiscal federalism, equalization, marginal cost of public funds, tax evasion

JEL classification: H 2, H 41, H 71, H 77

1. Introduction

It is well known in the literature on fiscal federalism that a vertical fiscal externality arising from two-level taxation on the same tax base may imply overprovision of local public goods as local governments perceive a marginal cost of public funds lower than the (true) social one (Dahlby, 2008). Moreover, such an incentive to overprovide local public goods increases when an equalization system with redistributive purposes is applied (Smart, 1998; Kelders and Köthenbürger, 2010).[1] It is also shown in the literature on optimal taxation that the introduction of sheltering (illegal tax evasion and/or legal tax avoidance) may affect the criterion for measuring the excess burden, and thus also the marginal cost of public funds (Chetty, 2009; Saez et al., 2012). This may in turn affect the efficiency of local public-good provision. Accordingly, when the effects of equalization and sheltering are considered simultaneously, one interesting question is whether or not the incentive to overprovide local public goods will prevail.

* We are deeply grateful to Alberto Zanardi, Alfons J. Weichenrieder and three anonymous referees for their extremely helpful comments.

1 This is not the case when the transfers aim instead to correct fiscal externalities (Dahlby (2008), ch. 9).

FinanzArchiv 68 (2012), 303–330 doi: 10.1628/001522112X653859
ISSN 0015-2218
© 2012 Mohr Siebeck

To analyze this issue, we consider a federation where a central government is precommitted, i.e., it acts as a Stackelberg leader while regional governments act as Stackelberg followers (Köthenbürger, 2008, (forthcoming)). Local public expenditures are financed through two different instruments: local governments choose a surtax on the local fiscal base of a national tax, and may be entitled to receive an equalization grant. In particular, we suppose that a linear labor income tax is devoted to financing public expenditure, at both the central and the local level, so that a federal tax rate and a local one are applied to the same basis. Further, the equalization system, which in our setup only has redistributive purposes, is chosen by the federal government, and it is like that described by Smart (1998).[2] In particular, it is based on the so-called *fiscal capacity equalization criterion*, a scheme of transfers applied in many countries such as Australia, Canada, Germany, Switzerland, and now, after a recent reform, Italy too. Accordingly, regions are (partially or totally) compensated by federal revenues for the difference between a standard level of tax revenue and the revenue regions are deemed to be able to raise if standard tax rates are applied to their tax basis.[3]

Different specifications of the equalization system (ES) are analyzed in this paper: vertical versus horizontal, and based on actual versus taxable income. With a *vertical*, or "gross" (*horizontal*, or "net") ES, *poor* regions, whose fiscal capacity is lower than the average or standard one, receive a grant from the central government (directly from *rich* regions). Further, both vertical and horizontal ESs may be implemented with a fiscal capacity based on pretax earned income or on reported taxable income (the former being net of sheltering). For instance, art. 9(g) of the recent Italian bill n. 42/2009 on fiscal federalism reform applies a horizontal ES by referring

2 The present paper refers to the literature on the *fiscal gap*, i.e., the gap between a lower tier of government's tax revenues and its expenditure (Dahlby, 2008, ch. 9; Boadway and Tremblay, 2010). Efficiency would require that the marginal cost of public funds be the same at both the regional and the federal level; otherwise a vertical fiscal imbalance arises. In that case, intergovernmental transfers could be designed to equalize the marginal cost of public funds across levels of government. Further, intergovernmental transfers could be also designed to counteract fiscal distortions due to horizontal tax and expenditure externalities. Beyond such efficiency aims, intergovernmental grants are also used for redistributive purposes, i.e., to favor horizontal equity across regions of a federation, as in Italy, Germany, Canada, and Australia. Equalization grants with only redistributive purposes are the object of our model.

3 See, in particular, Bucovetsky and Smart (2002), Boadway et al. (2002), Baretti et al. (2002), Dahlby and Warren (2003), Köthenbürger (2002), Boadway (2004), Figueiras et al. (2004), Buettner (2006), Grazzini and Petretto (2006), Smart (2007), Hindriks et al. (2008), Egger et al. (2010). Most of these papers deal with horizontal tax competition versus equalization, while in the present model we concentrate our attention on vertical tax competition. Kotsogiannis (2010) deals with both tax competition typologies versus equalization. Finally, Kotsogiannis and Schwager (2008) study the effect of equalization on the accountability of local politicians.

to the comparison between the regional per capita tax yield and the average tax revenue over all regions. The tax considered as a proxy of fiscal capacity is a regional income surtax at a standardized rate. Therefore, it seems that the Italian ES will be horizontal and based on reported taxable income.

When there are sheltering phenomena – in particular, tax evasion, on which we focus here – it is interesting to consider the activity of an authority devoted to the auditing and monitoring of tax compliance, and the degree of cooperation between regional and federal governments in running it.[4] Usually, both government levels participate in this activity, providing resources and agreeing on the way to share the clawed-back tax revenue.[5] Such aspects are also analyzed in the present paper.

To the best of our knowledge, the main novelty of this paper is its comparison of different types of ES within the same framework, taking also into account evasion and the consequent auditing and monitoring activity of an authority carried on at a regional level. In particular, we are interested in comparing different types of ES along two dimensions: vertical versus horizontal, and based on pretax earned income versus reported taxable income. Such a comparison allows us to identify the most efficient type of ES in redistributing resources across regions, with respect to reduction in the incentive for local public-goods overprovision.

In this respect, our main results are the following. From the point of view of regional governments, for a poor region, a shift from a vertical to a horizontal ES or vice versa does not change the incentive for overprovision of a public service, either in the case of an ES based on pretax earned income or in the case of one based on reported taxable income (a poor region is always a receiving one). However, such an incentive for a poor region is strengthened (reduced) when an ES based on pretax earned (reported taxable) income is replaced by one based on reported taxable (pretax earned) income, in the cases of both a vertical and a horizontal ES. For a rich region, a shift from a vertical (horizontal) to a horizontal (vertical) ES increases (decreases) the incentive for overprovision of a public service, in the cases of an ES based on pretax earned income and on reported taxable income. Further, such an incentive for a rich region does not change when a vertical ES based on pretax earned income is replaced by one based on reported taxable income.

4 For a treatment of evasion and auditing issues in a federal context, see Cremer and Gahvari (2000), Bartolini and Fiorillo (2009), and Stöwhase and Traxler (2005). In particular, the last paper analyzes how the decentralized choice of audit rates may be affected both by fiscal competition among regions and by the type of the fiscal equalization scheme (gross or net revenue sharing). For an extension of such a model allowing for a multiregional firm structure, see Traxler and Reutter (2008).

5 The cited bill on fiscal federalism in Italy explicitly considers the possibility that a regional agency can cooperate with the *Agenzia delle Entrate*, the national tax authority. For some taxes, regional authorities may even fully replace the *Agenzia delle Entrate*.

Instead, in the case of a horizontal ES, a shift from one based on pretax earned (reported taxable) income to another based on reported taxable (pretax earned) income increases (decreases) the incentive for a rich region for overprovision of a public service.

From the point of view of the federal government, we show that the degree of equalization is higher with a horizontal ES than with a vertical one, for both indexes of fiscal capacity. Since we show that a lower degree of equalization implies a lower incentive to overprovide local public goods, if the latter represents one of the aims of an ES, then a vertical ES should be preferred. Further, we show that the lowest degree of equalization is obtained when a vertical ES is based on pretax earned income rather than reported taxable income.

To conclude, our analysis suggests that a more efficient level of regional public expenditures is associated with a vertical ES based on pretax earned income at regional level.[6] The drawback of such a scheme is that it is more demanding in the information to be gathered. Tax records containing data on reported taxable income are widely available, while gathering data on earned incomes at a regional level is quite difficult. However, to apply such a scheme, the tax base has to be known at an aggregate and not at an individual level. Consequently, some proxy variables, such as GDP or NDP, which can be found in national-accounts data, could be used for our purposes. We are aware that national-accounts data may likewise be biased because they are usually derived from micro reporting. However, since such micro reporting systems are usually based on the anonymity of responses, we may expect their bias to be lower than that of data on reported taxable income. Further, it is likely that the use of these proxy variables would be opposed by local governments because they cannot influence them, given their pure fixed-price nature. With an ES based on reported taxable income, local politicians might desire to favor their constituencies with generally permissive fiscal legislation, and simultaneously to compensate the loss of tax revenues due to evasion with equalization. On the contrary, with an ES based on pretax earned income, an increase in evasion would reduce the equalization grant received by poor regions, so that local politicians should be more concerned with reducing such evasion.

The plan of the paper is as follows. In section 2, we analyze consumer behavior with tax evasion, and we design the tax structure of the federation according to the type of ES (vertical versus horizontal) and the basis on which the grant is calculated (pretax earned income versus reported taxable income). In section 3, the incentive to overprovide local public goods when

6 Thus, such a solution is contrary to the one established by the mentioned Italian reform on fiscal federalism.

there is tax evasion at the local level is compared according to different typologies of ES. In section 4, different degrees of equalization chosen by the federal government are compared according to the type of ES. Finally, section 5 contains some concluding remarks.

2. The Model

We analyze a federation of regions i, $i = 1, ..., n$, each with a population size normalized to unity. The federation has two tiers of government: central and regional. The federal government has the power to levy a tax on national income, and to decide the degree of equalization among regions. The regional governments have the power to levy a tax on regional income, and regions that are classified as poor are entitled to receive an equalization grant. For reasons of equity, the constitution specifies that an equalization system based on fiscal capacity has to be implemented. It also specifies that the equalization rate α has to be not lower than a threshold $\overline{\alpha}$ in order to guarantee that all regions can afford to supply a minimum level of local public goods.[7] This means that the constitution does not specify either the type of ES or the level of the equalization rate, which instead are chosen by the federal government.

In order to describe the interaction between the federal government and the regional ones, we suppose a three-stage game. Events in the model unfold as follows. First, the federal government chooses its tax rate and the degree of equalization, acting as a Stackelberg leader with respect to regions. Second, regions choose their tax rate by playing a Nash game among them, and acting as Stackelberg followers with respect to the federal government.[8] Finally, agents in both countries make their consumption and labor decisions.

In each region i, consumers are immobile, and preferences of the representative consumer are described by the following additively separable utility function:

$$U^i(x_i, l_i, g_i, G) = u^i(x_i, l_i) + b(g_i) + B(G), \qquad i = 1, ..., n, \tag{1}$$

where $u^i(.)$ is a strictly quasiconcave sub-utility function of private consumption x_i (taken as the numeraire) and leisure l_i, and the functions $b(g_i)$ and $B(G)$ measure the benefits of a local public good g_i, $i = 1, ..., n$, and a federal public good G, respectively.[9] g_i and G are only locally consumed. The unit cost of both public goods is one; thus g_i, $i = 1, ..., n$, and G also represent local and federal public expenditure, respectively. Both public expenditures

7 For example, such a prescription can be found in art. 109 of the Italian Constitution.

8 The same assumption can be found, for example, in Goodspeed (2002) and Keen and Kotsogiannis (2002).

9 The separability assumption on the utility function implies that g_i, $i = 1, ..., n$, and G do not affect households' leisure–consumption decisions.

are financed through a labor income tax. Let t be the tax rate chosen by the federal government, and let ρ_i, $i = 1, ..., n$, be the surtax on the regional fiscal base, decided by the regional government, with the consolidated tax rate given by $\tau_i \equiv t + \rho_i$, $i = 1, ..., n$.

In each region i, $i = 1, ..., n$, the quantity $Y_i = w_i L_i$ is the pretax earned income from labor, with w_i denoting the gross wage paid by firms, and $L_i = 1 - l_i$ denoting labor.[10] Thus, the net wage rate received by a consumer in region i is obtained as $\widetilde{w}_i = (1 - \tau_i)w_i$, $i = 1, ..., n$. Further, let us suppose that a taxpayer living in region i may evade taxation on s_i euros of income. Thus, his reported taxable income is $TY_i = Y_i - s_i$.

Following the distinction proposed by Chetty (2009), tax evasion in region i implies a resource cost given by $h_i(s_i)$ with $h_i'(.) > 0$ and a transfer cost given by $z_i(\tau_i, s_i)$ with $\frac{\partial z_i}{\partial \tau_i} > 0$ and $\frac{\partial z_i}{\partial s_i} > 0$.[11] In our case, the resource cost function $h_i(s_i)$ describes the cost borne by a tax authority that is engaged in monitoring and auditing procedures,[12] and the transfer cost function $z_i(\tau_i, s_i)$ describes the expected private cost of tax evasion, mainly due to the risk of being caught and charged a fine. The latter cost implies a transfer of resources from the taxpayer to the tax authority.

In order to gather some more hints on this point, let us assume the following tax-authority resource cost function:

$$h_i(s_i) = c_i[p_i(\eta_i)]s_i\,, \tag{2}$$

where $c_i[.]$ is the unit cost of auditing, and it is an increasing function of $p_i(\eta_i)$, which equals the probability that a taxpayer living in region i is audited, with η_i denoting the effort variable measuring the x-efficiency of the tax authority. The probability of being audited is clearly an increasing function of the effort: $p'(\eta_i) > 0$. Further, if the taxpayer has evaded and is audited, he has to pay his tax bill on the evaded amount, $\tau_i s_i$, and a fine $F(s_i, \tau_i)$, which can be assumed to be linear, i.e., $F(s_i, \tau_i) = f_i \tau_i s_i$. Thus, the transfer cost function equals the sum of the expected repaid tax plus the fine, i.e.,

$$z_i(\tau_i, s_i) = p_i(\eta_i)(\tau_i s_i + f_i \tau_i s_i)\,. \tag{3}$$

The budget constraint of a consumer in region i is

$$x_i = (1 - \tau_i)TY_i + s_i - z_i(\tau_i, s_i), \qquad i = 1, ..., n\,. \tag{4}$$

Each consumer living in region i, $i = 1, ..., n$, chooses leisure and how much income to conceal by maximizing the utility function in (1) subject to

10 We suppose that w_i, $i = 1, ..., n$, is constant, and thus it is not affected by taxation.

11 We also assume $\frac{\partial z_i(\tau_i, 0)}{\partial s_i} = 0$ and $\frac{\partial z_i(\tau_i, Y_i)}{\partial s_i} = \infty$ to guarantee an interior optimum in s_i.

12 Actually, in Chetty (2009) there is also a direct cost borne by the taxpayer, for instance, for shifting reported money rewards to nontaxable fringe benefits in order to reduce his taxable income.

his budget constraint (4). The solution of this maximization problem implies the following indirect utility function:

$$V^i(\widetilde{w}_i, g_i, G) = v^i(\widetilde{w}_i) + b(g_i) + B(G), \qquad i = 1, ..., n\,, \tag{5}$$

and, through Roy's identity, the following condition:

$$\frac{\partial v^i}{\partial \tau_i} = -v_I^i\left(TY_i + \frac{\partial z_i}{\partial \tau_i}\right), \qquad i = 1, ..., n\,, \tag{6}$$

where $v_I^i = u_x^i$ is the marginal utility of income. This means that the marginal cost for the taxpayer due to the cost of tax evasion, $-v_I^i \frac{\partial z_i}{\partial \tau_i} < 0$, adds to the standard cost of the increase of the tax payment, $-v_I^i TY_i$. By using (3), and applying the envelope theorem, according to which

$$\frac{\partial z_i}{\partial \tau_i} = p_i(\eta_i)s_i(1 + f_i)\,, \tag{7}$$

the condition (6) can be rewritten as follows:

$$\frac{\partial v^i}{\partial \tau_i} = -\,v_I^i[TY_i + p_i(\eta_i)s_i(1 + f_i)] = -v_I^i\{Y_i - [1 - p_i(\eta_i)(1 + f_i)]s_i\},$$
$$i = 1, ..., n\,. \tag{8}$$

This condition shows that the traditional individual welfare cost of a tax increase, $v_I^i Y_i$, is reduced by the expected net benefit from evasion, $v_I^i[1 - p_i(\eta_i)(1 + f_i)]s_i$, when $p_i(\eta_i) < \frac{1}{1+f_i}$, i.e., evasion is worthwhile ($s_i > 0$).

The structure of the federal and regional public budget constraints depends on the agreement between federal and regional governments on running the tax authority, i.e., how the expected benefits and costs of monitoring and auditing activity are shared between central and local governments. Moreover, the budget constraints of the two levels of governments depend on whether the adopted ES is vertical or horizontal, and on whether it is based on pretax earned income or reported taxable income.

Let R_i denote the revenue available for a region i, $i = 1, ..., n$. The regional public budget constraint is

$$R_i = \rho_i TY_i + e_i + \beta_i D_i = g_i, \qquad i = 1, ..., n\,. \tag{9}$$

Each region finances a local public good through three types of revenue: the yield from regional taxation, $\rho_i TY_i$; the yield from a federal government grant, e_i; and the yield from monitoring activity, $\beta_i D_i$. The latter represents the fraction of the tax authority's budget devoted to region i, where $D_i = z_i(\tau_i, s_i) - h_i(s_i) \geqslant 0$. For instance, by substituting (2) and (3) into the tax authority's budget devoted to region i , we obtain

$$D_i(\eta_i) = \{p_i(\eta_i)(1 + f_i)\tau_i - c_i[p_i(\eta_i)]\}s_i\,. \tag{10}$$

With $0 < \beta_i < 1$, we assume the existence of an agreement between central and regional governments according to which each region yields a share

β_i of the difference between the transfer from taxpayers and the cost of monitoring and auditing. This division might occur according to a political deal related to the ratio of the regional tax rate to the combined one, i.e., $\beta_i\left(\frac{\rho_i}{\tau_i}\right)$, with $\frac{\partial \beta_i}{\partial \rho_i} = \frac{\beta_i' t}{\tau_i^2} > 0$.

Denoting the federal revenue with a vertical and a horizontal ES as R_F^V and R_F^H, respectively, we obtain the corresponding federal public budget constraints as

$$R_F^V = tTY + \sum_{k=1}^{n}(1-\beta_k)D_k - \sum_{k\in N^p} e_k = G \tag{11}$$

and

$$R_F^H = tTY + \sum_{k=1}^{n}(1-\beta_k)D_k = G\,. \tag{12}$$

A federal public good is financed through the revenue from taxing labor income, tTY, with $TY = \sum_{k=1}^{n} TY_k$, plus the central government's fraction of the tax authority's budget, $\sum_{k=1}^{n}(1-\beta_k)D_k$, and in the case of a vertical ES, minus the sum of the grants to poor regions, $\sum_{k\in N^p} e_k$. The last term does not appear in (12) because a horizontal ES does not involve any transfer from the federal government to regions.

We are now in a position to describe the structure of regional and federal budget constraints, depending on the type of ES and on the type of income with respect to which the grant is calculated.

2.1. Vertical Equalization on Pretax Earned Income (*VY*)

When pretax earned income is the basis for the ES, let us define a poor (rich) region i as one for which $Y_i < (\geq)\overline{Y}$, with $i \in (\notin) N^p$, where N^p denotes the set of poor regions (n_p their number), and $\overline{Y}$ denotes the standard tax base. Specifically, $\overline{Y}$ is usually chosen as the average per capita tax base, $\overline{Y} \equiv \sum_{k=1}^{n} Y_k/n = Y/n$.

With a vertical ES, the federal government only pays grants to poor regions $i \in N^p$. When such a scheme is based on pretax earned income, the equalization grants paid by the federal government are of the following type:

$$e_{i\in N^p}^{VY} = \alpha\overline{\rho}(\overline{Y} - Y_i) > 0\,, \tag{13}$$

$$e_{i\notin N^p}^{VY} = 0\,, \tag{14}$$

where the parameter α, $\overline{\alpha} \leqslant \alpha \leqslant 1$, describes the chosen degree of equalization ($\alpha = 1$ is the case of *full* equalization), and $\overline{\rho}$ denotes the standard regional tax rate. In particular, $\overline{\rho}$ could be a given reference surtax rate, established at a federal level, or a weighted-average surtax rate, $\overline{\rho} \equiv \sum_{k=1}^{n} \rho_k Y_k / \overline{Y}$. Notice that in (13), region i's per capita GDP can be used as a proxy for earned income at regional level, Y_i, and the average national per capita GDP can be used as a proxy for $\overline{Y}$. Under the assumption of a federation with a large number of regions, each of which is small enough with respect to the rest of the federation for us to neglect the effect of its own regional tax rate on the standard tax base and the standard tax rate, $\overline{Y}$ and $\overline{\rho}$ are fixed, i.e., independent of regional tax rates (Esteller-Moré and Solé-Ollè, 2002; Grazzini and Petretto, 2006).[13] This also implies that the sets of poor and rich regions do not change during the time of our analysis, so we are able to compare different ESs.

Let us define $A^Y \equiv \alpha\overline{\rho}\overline{Y}$, and insert it into (13). The equalization grant for a poor region can then be rewritten as

$$e^{VY}_{i \in N^p} = A^Y - \alpha\overline{\rho} Y_i \,. \tag{15}$$

For poor regions, this ES works as a linear (affine) matching grant based on (inversely correlated with) the actual local tax base.

Let R_i^{VY} denote the total revenue available for a region i, $i = 1, ..., n$, with vertical equalization on pretax earned income. By substituting (15) into (9), the regional public budget constraint is obtained as

$$R^{VY}_{i \in N^p} = (\rho_i - \alpha\overline{\rho}) T Y_i - \alpha\overline{\rho} s_i + A^Y + \beta_i D_i = g_i \tag{16}$$

for a poor region, and by substituting (13) into (9), we obtain

$$R^{VY}_{i \notin N^p} = \rho_i T Y_i + \beta_i D_i = g_i \tag{17}$$

for a rich region.

Let R_F^{VY} denote the federal revenue with a vertical equalization on pretax earned income. By substituting (15) into (11), the federal public budget constraint is obtained as

$$R_F^{VY} = tTY + \sum_{k=1}^{n} (1 - \beta_k) D_k - \sum_{k \in N^p} \left(A^Y - \alpha\overline{\rho} Y_k \right) = G \,. \tag{18}$$

13 For instance, in Italy there are 20 regions, and some of them are quite small. The 2009 per capita GDP varies from €12,776 in Campania to €26,755 in Val d'Aosta. The average per capita GDP is €24,400 . The average central government tax rate of the Imposta sul Reddito delle Persone Fisiche (IRPEF) is 37%, while the standard uniform surtax rate is 0.9%, and it can be changed by regions within the range $\mp 0.5\%$. It is reasonable to expect that an increase of this dimension in the surtax rate, say, in the median region (Liguria), can reduce its regional income, but it can only produce negligible effects on the average national per capita income.

2.2. Horizontal Equalization on Pretax Earned Income (*HY*)

With a horizontal ES, poor regions receive a subsidy while rich regions pay for them. When such a scheme is based on pretax earned income, the structure of the yields is as follows:

$$e^{HY}_{i\in N^p} = e^{VY}_{i\in N^p}\,, \tag{19}$$

$$e^{HY}_{i\notin N^p} = \alpha\overline{\rho}(\overline{Y} - Y_i) \leqslant 0\,, \tag{20}$$

with $\sum_{i\in N^p} e^{HY}_i + \sum_{i\notin N^p} e^{HY}_i = 0$.

Let R^{HY}_i denote the total revenue available for a region i, $i = 1, ..., n$, with a horizontal equalization on pretax earned income. Since the grant received by a poor region is the same as in the case with vertical equalization based on pretax earned income, the regional public budget constraint for a poor region is also the same, i.e.,

$$R^{HY}_{i\in N^p} = R^{VY}_{i\in N^p} = g_i\,. \tag{21}$$

By substituting (20) into (9), the regional public budget constraint for a rich region is obtained as

$$R^{HY}_{i\notin N^p} = (\rho_i - \alpha\overline{\rho})TY_i - \alpha\overline{\rho}s_i + A^Y + \beta_i D_i = g_i\,. \tag{22}$$

Finally, the federal public budget constraint is (12).

2.3. Vertical Equalization on Reported Taxable Income (*VTY*)

When reported taxable income is the basis for the ES, we continue to use the same criterion to define a poor region, viz., a poor (rich) region i is one such that $Y_i < (\geq)\overline{Y}$.[14] With a vertical scheme the grants paid by the federal government are of the following type:

$$e^{VTY}_{i\in N^p} = \alpha\overline{\rho}(\overline{TY} - TY_i) > 0\,, \tag{23}$$

$$e^{VTY}_{i\notin N^p} = 0\,. \tag{24}$$

Let us define $A^{TY} \equiv \alpha\overline{\rho}\,\overline{TY}$, and use it in (23). The grant for a poor region can then be rewritten as

$$e^{VTY}_{i\in N^p} = A^{TY} - \alpha\overline{\rho}(Y_i - s_i)\,. \tag{25}$$

Given the actual earned income, the grant increases with the level of tax evasion.

14 Notice that when the type of ES changes, regional policy incentives may change, and thus each regional income position relative to the exogenous threshold may also change. To be sure the sets of poor and rich regions are unchanged, we also assume that poor and rich regions are sufficiently far away from the exogenous threshold, i.e., the implied change in regional (earned or reported) income is small. We thank a referee for suggesting this point.

Let R_i^{VTY} denote the total revenue available for a region i, $i = 1, ..., n$, with a vertical equalization on reported taxable income. By substituting (25) into (9), the regional public budget constraint is obtained as

$$R_{i\in N^p}^{VTY} = (\rho_i - \alpha\overline{\rho})TY_i + A^{TY} + \beta_i D_i = g_i \tag{26}$$

for a poor region, and by substituting (24) into (9), it is obtained as

$$R_{i\notin N^p}^{VTY} = \rho_i TY_i + \beta_i D_i = g_i \tag{27}$$

for a rich region. Thus $R_{i\notin N^p}^{VTY} = R_{i\notin N^p}^{VY}$.

Now let R_F^{VTY} denote the federal revenue with a vertical equalization on reported taxable income. By substituting (25) into (11), the federal public budget constraint is obtained as

$$R_F^{VTY} \equiv tTY + \sum_{k=1}^{n}(1-\beta_k)D_k - \sum_{k\in N^p}(A^{TY} - \alpha\overline{\rho}TY_k) = G\,. \tag{28}$$

2.4. Horizontal Equalization on Reported Taxable Income (*HTY*)

With a horizontal equalization based on reported taxable income, the structure of the yields is as follows:

$$e_{i\in N^p}^{HTY} = e_{i\in N^p}^{VTY}\,, \tag{29}$$

$$e_{i\notin N^p}^{HTY} = \alpha\overline{\rho}(\overline{TY} - TY_i) \leqslant 0\,, \tag{30}$$

with $\sum_{i\in N^p} e_i^{HTY} + \sum_{i\notin N^p} e_i^{HTY} = 0$.

Let R_i^{HTY} denote the total revenue available for a region i, $i = 1, ..., n$, with a horizontal equalization on reported taxable income. Since the grant received by a poor region is the same as in the case with vertical equalization based on pretax earned income, the regional public-budget constraint for a poor region is also the same, i.e.,

$$R_{i\in N^p}^{HTY} = R_{i\in N^p}^{VTY} = g_i\,. \tag{31}$$

By substituting (30) into (9), the regional public budget constraint for a rich region is obtained as

$$R_{i\notin N^p}^{HTY} = (\rho_i - \alpha\overline{\rho})TY_i + A^{TY} + \beta_i D_i = g_i\,. \tag{32}$$

Finally, the federal public-budget constraint is (12).

3. The Incentives for Local Public-Goods Overprovision with Different Equalization Schemes and Tax Evasion

In this section, we solve the game described above by backward induction. In the previous section, we solved the third stage of the game, when consumption and labor decisions are made. We now turn to the solution of the

second stage of the game. Since regional governments play a noncooperative game among themselves, and act as Stackelberg followers with respect to the federal government, they do not take into account the effects of their fiscal decisions on the federal government's budget constraint. In particular, ρ_i and g_i are chosen in order to maximize the welfare of a representative consumer in region i (see (5)), given the regional public-sector budget constraint, and taking as given the fiscal decisions of both the federal government and other regions.[15]

In accord with wide agreement in the literature, we adopt the following definition:

Definition 1 In a noncooperative fiscal (subgame) equilibrium, the incentive to overprovide local public services is measured by the gap between the regional perceived marginal cost of public funds (MCF_i) and the social one ($SMCF_i$), i.e., the one perceived at the second-best optimum.

According to this definition, the wider the wedge between MCF_i and $SMCF_i$, the higher the incentive to overprovide local public goods.

3.1. Benchmark Case: No Equalization and No Tax Evasion

Without an ES and tax evasion, by solving the maximization problem of a region i, $i = 1, ..., n$, the following noncooperative subgame equilibrium condition is obtained:[16]

$$\frac{b'}{v_I^i} = \frac{Y_i}{\frac{\partial R_i}{\partial \rho_i}} = \frac{1}{1 - \rho_i \Delta_i \varepsilon_i} = MCF_i, \qquad i = 1, ..., n\,, \tag{33}$$

where $v_I^i \equiv -\frac{\partial v^i / \partial \tau_i}{Y_i}$, $\Delta_i \equiv \frac{1}{1-\tau_i}$, $\varepsilon_i \equiv \frac{\partial L_i}{\partial \widetilde{w}_i} \frac{\widetilde{w}_i}{L_i}$, and we recall the assumption that in the federal country there are so many regions and each of them is sufficiently small relative to the rest of the federation that the effect of a regional surtax rate on the federal budget and, accordingly, on its own residents' utility function (1) through the B-term is negligible.[17] This means that the marginal benefit of one euro invested in the local public-good provision is equal to its regionally perceived marginal cost. In contrast, the second-best cooperative equilibrium condition that takes into account the negative tax externality,

15 As is well known, this corresponds to the case when regions are fiscally independent (Köthenbürger, 2008a; 2008b). A possible interaction is investigated in Kotsogiannis (2010).

16 See, for example, Dahlby (2008) and Keen (1998).

17 As has been pointed out by Kelders and Köthenbürger (2010), if in the federation one has $n \to \infty$, and each region receives an equal share of the federal budget, this "see-through" effect vanishes. For more insights into this type of effect, see Köthenbürger (forthcoming), and for a discussion of such assumptions, see Keen (1998).

$\frac{\partial R_F}{\partial \rho_i} < 0$, is as follows:

$$\frac{b'}{v_I^i} = \frac{Y_i}{\frac{\partial R_i}{\partial \rho_i} + \frac{\partial R_F}{\partial \rho_i}} = \frac{1}{1 - \tau_i \Delta_i \varepsilon_i} = SMCF_i, \qquad i = 1, ..., n\,. \tag{34}$$

This means that the socially perceived marginal cost of public funds, $SMCF_i$, takes into account the fact that the "true" tax distortion depends on τ_i and not only on ρ_i. Thus, it is easy to check that

$$MCF_i < SMCF_i, \qquad i = 1, ..., n\,,$$

i.e., without an ES scheme, the noncooperative equilibrium results in an incentive for overprovision.

3.2. Case with Different Equalization Schemes and No Tax Evasion

When an ES is introduced, the grants are as in (13) and (14) in the case of a vertical equalization, and as in (19) and (20) in the case of a horizontal equalization. Without tax evasion, let $R^V_{i\in N^p}$ and $R^H_{i\in N^p}$ denote the total revenue available for a poor region with a vertical and a horizontal equalization, respectively. By using (13) and (19), the regional public budget constraint for a poor region is obtained as

$$R^V_{i\in N^p} = R^H_{i\in N^p} = (\rho_i - \alpha\overline{\rho})Y_i + A^Y = g_i\,. \tag{35}$$

Since a rich region does not receive any grant with a vertical equalization (see (14)), its public budget constraint is the same as in the benchmark case:

$$R^V_{i\notin N^p} = \rho_i Y_i = g_i\,. \tag{36}$$

Instead, a rich region pays a transfer (20) with a horizontal equalization, and thus its public budget constraint is obtained as

$$R^H_{i\notin N^p} = (\rho_i - \alpha\overline{\rho})Y_i + A^Y = g_i\,. \tag{37}$$

Let MCF_i^V and MCF_i^H denote the marginal cost of public funds without tax evasion, and with a vertical and a horizontal equalization, respectively. By using (35), it is easy to check that the following condition for a poor region holds:

$$MCF^V_{i\in N^p} = MCF^H_{i\in N^p} = \frac{1}{1 - (\rho_i - \alpha\overline{\rho})\Delta_i\varepsilon_i} < MCF_i < SMCF_i\,. \tag{38}$$

This condition shows that the perceived marginal cost of public funds is lower with both vertical and horizontal equalization than without an ES. Accordingly, the incentive for overprovision of public goods is strengthened when an equalizing grant is directed to a poor region, for which the effective tax rate becomes $\rho_i - \alpha\overline{\rho}$. Indeed, the increase in the regional tax rate ρ_i of a poor region results in a decrease in its regional income due to the elasticity of labor supply. However, such a decrease in regional income allows

a greater equalization grant. Accordingly, the perception of the marginal cost of taxation decreases.

By using (36) and (37), it is easy to check that the following condition for a rich region holds:

$$MCF^{H}_{i\notin N^p} = \frac{1}{1-(\rho_i - \alpha\overline{\rho})\Delta_i\varepsilon_i} < MCF^{V}_{i\notin N^p} = MCF_i < SMCF_i\,. \tag{39}$$

The result concerning overprovision of the public service also applies to a rich region in the case of horizontal equalization. Actually, if region $i \notin N^p$ increases its tax rate ρ_i, then by decreasing its income it is enabled to pay less for the transfer to the set of poor regions, so its perceived marginal cost of taxation is also lower.

3.3. Case with Different Equalization Schemes and Tax Evasion

Now let us introduce the possibility of tax evasion. We will compare the social marginal cost of public funds with tax evasion, denoted by $SMCF^S_i$, with the perceived marginal cost of public funds, depending on the type of ES adopted, and on the basis on which the grant is calculated.

Since $SMCF^S_i$ is calculated taking into account the sum of all regional and federal public budget constraints, it is easy to check that it is independent not only of the type of ES (vertical or horizontal), but also of the basis on which the yield is calculated (pretax earned income or reported taxable income). Indeed, when all constraints on regional public budgets are summed along with the federal one, positive and negative equalization transfers cancel out. Thus, the social marginal cost of public funds is obtained as[18]

$$SMCF^S_i = \frac{TY_i + \frac{\partial z_i}{\partial \rho_i}}{TY_i\left(1-\tau_i\Delta_i\varepsilon^{TY}_i\right) + \frac{\partial D_i}{\partial \rho_i}}, \qquad i = 1, ..., n\,, \tag{40}$$

where $\varepsilon^{TY}_i \equiv \frac{\partial TY_i}{\partial \widetilde{w}_i}\frac{\widetilde{w}_i}{TY_i}$ denotes the elasticity of reported taxable income. Let us briefly consider the differences from the previous scenario without tax evasion.

Now in the numerator of $SMCF^S_i$ there is the sum of two types of marginal cost of taxation for the taxpayer in region i: TY_i and $\frac{\partial z_i}{\partial \rho_i}$. On using $TY_i = Y_i - s_i$, (7), and (8), the numerator in (40) becomes $Y_i - [(1 - p_i(\eta_i)(1 + f_i)]s_i$, which is lower than Y_i, the numerator in (34), the formula without tax evasion. Two observations can be made about the denominator of $SMCF^S_i$. First, the tax distortion is measured not in terms of the elasticity of labor supply, but in terms of the elasticity of reported taxable income, which is now the appropriate method for calculating the excess burden of income

18 Note the similarity to the condition (27) in Chetty (2009), although drawn in a different context.

taxation.[19] Second, there is the term $\frac{\partial D_i}{\partial \rho_i}$, which is positive if the marginal revenue coming from the transfer cost is higher than the marginal cost of organizing the tax authority. As is shown in lemma 1 in the appendix, this is indeed the case under the plausible assumptions that the tax authority is efficient in choosing the effort level and that evasion is increasing with respect to the tax rate.[20] The term $\frac{\partial D_i}{\partial \rho_i} > 0$ in the denominator of (40) means that the authority can translate the increase in evasion, coming from a tax-rate increase, into an opportunity for greater yield, partially compensating for the reduction of revenue due to the taxable-income decrease. The latter effect is, of course, captured by the term with the elasticity of taxable income in the condition (40).

3.3.1. The Incentive for Overprovision for Poor Regions

With both a vertical and a horizontal ES based on pretax earned income, for a poor region $i \in N^p$, the local marginal cost of public funds is obtained as

$$MCF^{VY}_{i\in N^p} = MCF^{HY}_{i\in N^p} = \frac{TY_i + \frac{\partial z_i}{\partial \rho_i}}{TY_i\left(1 - \rho_i \Delta_i \varepsilon_i^{TY}\right) + \alpha\overline{\rho}\Delta_i \varepsilon_i Y_i + \frac{\partial(\beta_i D_i)}{\partial \rho_i}}. \qquad (41)$$

Let us now compare the local marginal cost of public funds in (41) with the social one in (40). The numerator of the two expressions is the same, $TY_i + \frac{\partial z_i}{\partial \rho_i}$, as it captures the full effect of taxation on a taxpayer in region i. The denominator is different. In particular, the denominator in (41) contains three terms. The first term, $TY_i\left(1 - \rho_i \Delta_i \varepsilon_i^{TY}\right)$, represents the net gain in yield from an increase in regional tax rate. The second term, $\alpha\overline{\rho}\Delta_i \varepsilon_i Y_i$, represents the compensation coming from the ES, measured in terms of the labor-supply elasticity and pretax earned income. Finally, the third term, $\frac{\partial(\beta_i D_i)}{\partial \rho_i} = \frac{\partial \beta_i}{\partial \rho_i} D_i + \beta_i \frac{\partial D_i}{\partial \rho_i} > 0$, describes the effect of the regional tax rate on the share of the authority budget going to region i. More precisely, the three terms represent three types of fiscal externalities due to an increase of the regional tax rate, which determine the difference between the local marginal cost of public funds and the social one. The first term corresponds to a tax distortion depending on τ_i and not only on ρ_i; the second term corresponds to the compensation received by region i according to the vertical or horizontal ES, which cancels out at a social level; and finally, the third term captures the fact that region i perceives only a fraction of the increase in the tax authority's yield from monitoring activity. In particular, $MCF \gtrless SMCF$ if $\frac{\partial D_i}{\partial \rho_i} \gtrless \frac{\partial(\beta_i D_i)}{\partial \rho_i}$. In this connection, it is easy to check that $\frac{\partial D_i}{\partial \rho_i} \gtrless \frac{\partial(\beta_i D_i)}{\partial \rho_i}$ if $\frac{\partial D_i}{\partial \tau_i}\frac{\tau_i}{D_i} \gtrless 1$ and $\beta_i = \frac{\rho_i}{\tau_i}$.

19 See Chetty (2009) and the cited literature.

20 For a deeper analysis of the decentralized choice of auditing rates, see Stöwhase and Traxler (2005).

In other words, the regionally perceived MCF_i tends to be higher (lower) than the social one, $SMCF_i$ (the benchmark being the share of auditing costs in the total tax authority's revenue) if β_i is exactly equal to the ratio of local tax rate on federal tax rate, $\frac{\rho_i}{\tau_i}$, and the net revenues of monitoring agency is relatively elastic (inelastic) with respect to the total tax rate τ_i.

Similarly, with both a vertical and a horizontal ES based on reported taxable income, for a poor region $i \in N^p$, the local marginal cost of public funds is obtained as

$$MCF^{VTY}_{i\in N^p} = MCF^{HTY}_{i\in N^p} = \frac{TY_i + \frac{\partial z_i}{\partial \rho_i}}{TY_i\left(1 - \rho_i \Delta_i \varepsilon_i^{TY}\right) + \alpha\overline{\rho}\Delta_i \varepsilon_i^{TY} TY_i + \frac{\partial(\beta_i D_i)}{\partial \rho_i}}. \tag{42}$$

On comparing the regional marginal costs of public funds in (42) with the ones in (41), notice that the only different term is the second one in the denominator: the elasticity of labor supply times pretax earned income, $\varepsilon_i Y_i$, in (41) is replaced by the elasticity of reported taxable income times reported taxable income, $\varepsilon_i^{TY} TY_i$ in (42).

To conclude, for a poor region, in the case of an ES based on both pretax earned income and reported taxable income, a vertical and a horizontal ES provide the same incentive to overprovide the public service, i.e., $MCF^{VY}_{i\in N^p} = MCF^{HY}_{i\in N^p}$ and $MCF^{VTY}_{i\in N^p} = MCF^{HTY}_{i\in N^p}$. Instead, in the cases of both a vertical and a horizontal scheme, the incentive to overprovide the public service is higher when the scheme is based on reported taxable income instead of pretax earned income, i.e., $MCF^{VTY}_{i\in N^p} < MCF^{VY}_{i\in N^p}$ and $MCF^{HTY}_{i\in N^p} < MCF^{HY}_{i\in N^p}$, under the plausible assumption that $\frac{\partial s_i}{\partial \rho_i} = (\varepsilon_i^{TY} TY_i - \varepsilon_i Y_i) > 0$.[21]

We summarize the previous reasoning with the following proposition:

Proposition 1 At the equilibrium subgame of the regional governments,

(i) for a poor region, in the case of an ES based on pretax earned income or reported taxable income, a shift from a vertical to a horizontal one or vice versa does not change the incentive for overprovision of the local public service;

(ii) for a poor region, in the cases of both a vertical and a horizontal equalization, a shift from an ES based on pretax earned (reported taxable) income to one based on reported taxable (pretax earned) income increases (decreases) the incentive for overprovision.

Part (i) of proposition 1 is straightforward, because the grant received by a poor region does not depend on the type of ES, i.e., vertical or horizontal.

21 Notice that $\frac{\partial s_i}{\partial \rho_i} = (\varepsilon_i^{TY} - \varepsilon_i) TY_i - \varepsilon_i s_i$. The assumption that $\frac{\partial s_i}{\partial \rho_i} > 0$ is sustained by several empirical investigations in optimal income taxation, where it is found that the elasticity of taxable income is much higher than that of the labor supply. See, for example, Saez et al. (2010).

Part (ii) deserves further comment. Recall that the incentive for overprovision depends on the regional perceived marginal cost of taxation. If that cost decreases, the gap between the social marginal cost of public funds and the regionally perceived marginal cost of public funds increases. Let us now consider an increase in the tax rate of a poor region. This leads to two kinds of distortionary cost from the point of view of the regional revenue, in that it leads to a decrease in pretax earned income, but also to an increase in evasion. When the ES is based on pretax earned income, the cost due to an increase in evasion is not compensated by an equalization grant, and thus the loss of revenue is entirely borne by the poor region. On the contrary, when the ES is based on reported taxable income, the cost due to an increase in evasion leads to a decrease in reported taxable income, and that decrease is partially compensated by an increase in the equalization grant. Thus, the cost of taxation is lower and the incentive for local public-good overprovision is higher with an ES based on reported taxable income than with an ES based on pretax earned income.

3.3.2. The Incentive for Overprovision for Rich Regions

With a vertical ES based on both pretax earned income and reported taxable income, the local marginal cost of public funds, for a rich region $i \notin N^p$, is obtained as

$$MCF^{VY}_{i\notin N^p} = MCF^{VTY}_{i\notin N^p} = \frac{TY_i + \frac{\partial z_i}{\partial \rho_i}}{TY_i\left(1 - \rho_i \Delta_i \varepsilon_i^{TY}\right) + \frac{\partial(\beta_i D_i)}{\partial \rho_i}}.$$

Since a rich region pays nothing in the case of a vertical ES, its local marginal cost of public funds does not depend on the fact that the ES is based on pretax earned income or reported taxable income.

The local marginal cost of public funds is obtained as

$$MCF^{HY}_{i\notin N^p} = \frac{TY_i + \frac{\partial z_i}{\partial \rho_i}}{TY_i\left(1 - \rho_i \Delta_i \varepsilon_i^{TY}\right) + \alpha\overline{\rho}\Delta_i \varepsilon_i Y_i + \frac{\partial(\beta_i D_i)}{\partial \rho_i}} \tag{43}$$

with a horizontal ES based on pretax earned income, and

$$MCF^{HTY}_{i\notin N^p} = \frac{TY_i + \frac{\partial z_i}{\partial \rho_i}}{TY_i\left(1 - \rho_i \Delta_i \varepsilon_i^{TY}\right) + \alpha\overline{\rho}\Delta_i \varepsilon_i^{TY} TY_i + \frac{\partial(\beta_i D_i)}{\partial \rho_i}} \tag{44}$$

with a horizontal ES based on reported taxable income.

Two remarks are in order. First, $MCF^{HY}_{i\notin N^p} > MCF^{HTY}_{i\notin N^p}$ because $\varepsilon_i Y_i < \varepsilon_i^{TY} TY_i$. Second, both in the case of an ES based on pretax earned income and in the case of one based on reported taxable income, the comparison between the local marginal costs of public funds shows that $MCF^{HY}_{i\notin N^p} < MCF^{VY}_{i\notin N^p}$ because of the term $\alpha\overline{\rho}\Delta_i \varepsilon_i Y_i > 0$ in the denominator of (43), and $MCF^{HTY}_{i\notin N^p} <$

$MCF^{VTY}_{i \notin N^p}$ because of the term $\alpha \overline{\rho} \Delta_i \varepsilon_i^{TY} TY_i > 0$ in the denominator of (44). In particular, the tax distortion perceived by a rich region is lower with a vertical ES than with a horizontal one, i.e., if a rich region increases its local tax rate ρ_i, its income decreases, but it is enabled to pay less for the transfer to poor regions. Accordingly, a shift from a horizontal to a vertical ES augments the marginal cost of public funds, and thus reduces the gap with respect to the social marginal cost of public funds.

We summarize the previous reasoning as follows:

Proposition 2 At the equilibrium subgame of the regional governments,

(i) for a rich region, in the case of an ES based on pretax earned income or reported taxable income, a shift from a vertical (horizontal) to a horizontal (vertical) ES increases (decreases) the incentive for overprovision of the local public service;

(ii) for a rich region, in the case of a vertical equalization, a shift from an ES based on pretax earned income to one based on reported taxable income does not change the incentive for overprovision. Instead, in the case of a horizontal equalization, a shift from an ES based on pretax earned (reported taxable) income to one based on reported taxable (pretax earned) income increases (decreases) that incentive.

The result in part (i) of proposition 2 is due to the fact that, with a vertical ES, a rich region does not pay any grant, while with a horizontal ES it pays a grant to poor regions. Let us consider an increase in the tax rate of a rich region. This leads again to two kinds of distortionary cost from the point of view of the regional revenue, because it leads to a decrease in pretax earned income, but also to an increase in evasion. Only with a horizontal ES is one or both of these costs compensated because the grant paid by a rich region is reduced.

The intuition behind part (ii) of proposition 2 is linked to the fact that with a reported taxable income, both previous costs are partially compensated by a reduction of the grant paid by a rich region, while with a pretax earned income, only the first kind of cost is partially compensated, as the cost due to evasion is entirely borne by the rich region. Consequently, the cost of taxation is lower with reported taxable income than with pretax earned income, and the incentive for local public-good overprovision is higher.

4. The Federal Government's Decision on the Equalization Parameter

At the first stage of the game, the federal government chooses the federal tax rate on income, t, and the equalization parameter, α, depending on the type

of ES. To simplify the exposition, let us suppose a utilitarian federal government that maximizes the sum of regional indirect utility functions subject to the federal budget constraint, each regional budget constraint, and the constitutional constraint, i.e., $\overline{\alpha} \leqslant \alpha \leqslant 1$. Further, as the federal government acts as a Stackelberg leader with respect to regional governments, it takes into account all the regions' reaction functions, i.e., $\rho_i = \rho_i(t, \alpha)$, $i = 1, ..., n$.

The federal government's maximization problem is obtained as

$$\begin{aligned} &\max_{t,\alpha} \sum_{k=1}^{n} v^k(\widetilde{w}_k) + b(g_k) + B(G)\,, \\ &\text{s.t.} \quad R_F = G\,, \\ &\qquad R_i = g_i\,, \qquad i = 1, ..., n\,, \\ &\qquad \overline{\alpha} \leqslant \alpha \leqslant 1\,, \end{aligned} \tag{45}$$

where the federal and the regional budget constraints have to adapt to each type of ES.

Being only interested in comparing different ESs, let us concentrate our attention on the optimal choice of the equalization parameter α. In particular, what we are interested in is to identify which ES (vertical or horizontal, based on pretax earned income or reported taxable income) provides the lowest incentive for overspending by regions. Since the incentive for local public-good overprovision is increasing with respect to α,[22] from a positive point of view, the less distortive ES in terms of regional overspending is the one with the lowest level of α compatible with the constitutional requirement $\alpha \geqslant \overline{\alpha}$.[23]

In particular, we proceed as follows. First, we compare vertical and horizontal ESs, i.e., we compare the optimal levels of α chosen by the federal government with a vertical and a horizontal ES, both when the basis of comparison is pretax earned income and when it is reported taxable income; i.e., VY versus HY, and VTY versus HTY. Second, since we show that a vertical ES has the lowest level of α, we check whether the basis of comparison should be pretax earned income or reported taxable income, i.e., VY versus VTY.

22 This is due to the fact that the wedge between the consolidated income tax rate, $\tau_i = t + \rho_i$, and the effective one, $\rho_i - \alpha\overline{\rho}$, is increasing with respect to α (see (38) and (39)). For an analogous statement, see proposition 1 in Kelders and Köthenbürger (2010).

23 Notice that we are simply interested in identifying which ES has the lowest equalization rate, thus minimizing the incentive for overspending for regions. This does not mean that such an ES would be the one preferred from the federal government's point of view; for in choosing the equalization rate that maximizes a utilitarian social welfare function, the federal government could face a trade-off between equity concerns, in terms of redistribution of resources among regions via the equalization system fixed by the Constitution, and efficiency concerns, in terms of reduction of the incentive for overspending for regions. Accordingly, from a normative point of view, we cannot mantain that the minimum level of the equalization rate is optimal. We are indebted to an anonymous referee for this point.

The first-order condition with respect to α for the problem in (45) is obtained as[24]

$$\sum_{k=1}^{n} \frac{\partial v^k}{\partial \alpha} + b' \sum_{k \in N^p} \frac{\partial R_k^V}{\partial \alpha} + nB' \frac{\partial R_F^V}{\partial \alpha} = 0 \tag{46}$$

in the case of vertical ES, and

$$\sum_{k=1}^{n} \frac{\partial v^k}{\partial \alpha} + b' \left(\sum_{i \in N^p} \frac{\partial R_k^H}{\partial \alpha} + \sum_{i \notin N^p} \frac{\partial R_k^H}{\partial \alpha} \right) = 0 \tag{47}$$

in the case of horizontal ES, where $\frac{\partial v^i}{\partial \alpha} = -v_I^i \left(TY_i + \frac{\partial z_i}{\partial \rho_i} \right) \frac{\partial \rho_i}{\partial \alpha} < 0$, given that $\frac{\partial \rho_i}{\partial \alpha} > 0$, as is well known from the literature on revenue equalization (see Section 3.1). Notice that in (46), $\sum_{k \notin N^p} \frac{\partial R_k^V}{\partial \alpha} = 0$, because rich regions do not transfer grants to poor regions with a vertical ES, and in (47), $\frac{\partial R_F^H}{\partial \alpha} = 0$, because the federal government does not transfer grants to poor regions with a horizontal ES. Thus, by taking into account the signs of the partial derivatives, (46) and (47) can be written as follows:

$$\left| \sum_{k=1}^{n} \frac{\partial v^k}{\partial \alpha} \right| = b' \sum_{k \in N^p} \frac{\partial R_k^V}{\partial \alpha} - nB' \left| \frac{\partial R_F^V}{\partial \alpha} \right| \tag{48}$$

and

$$\left| \sum_{k=1}^{n} \frac{\partial v^k}{\partial \alpha} \right| = b' \left(\sum_{k \in N^p} \frac{\partial R_k^H}{\partial \alpha} - \sum_{k \notin N^p} \left| \frac{\partial R_k^H}{\partial \alpha} \right| \right), \tag{49}$$

respectively. Notice that, in both (48) and (49), we may interpret the left-hand side as the marginal cost, and the right-hand side as the net marginal benefit, of α from the point of view of the federal government. The marginal cost is related to the regional tax distortion in both cases, and given the properties of the indirect utility function (5), it is easy to check that it is increasing with respect to α.[25] In the case of a vertical ES, the net marginal benefit is due to the benefit in local public-good consumption in the set of poor regions, net of the loss in federal public-good consumption. In the case of a horizontal ES, the net marginal benefit is still due to the benefit in local public-good consumption in the set of poor regions, net however of the loss in local public-good consumption for rich regions. We assume that the net marginal benefit is decreasing with respect to α in both a vertical and a horizontal ES, by a standard assumption on decreasing marginal benefit. To ensure that an interior solution for the equalization parameter always exists, we also assume

24 Since we are only interested in interior solutions, the first-order condition with respect to α holds as an equality.

25 Recall that the indirect utility function is convex with respect to net tax wage and the latter is decreasing with respect to the tax rate, which is increasing with respect to α.

that the net marginal benefit is larger than the marginal cost at $\alpha \to \overline{\alpha}$, and that the reverse holds at $\alpha = 1$.

We are now in a position to compare the optimal value of α, $\widehat{\alpha}$, chosen by the federal government according to the different ESs. Since the ES does not affect the marginal cost in either (48) or (49), we check how the net marginal benefit varies according to each ES.

By comparing vertical versus horizontal ESs, we can state the following

Proposition 3 If $nB' > b'$, and the positive effect of regional taxation on the yield from monitoring activity at a federal and at a regional level is negligible with respect to all the other negative terms, then $\widehat{\alpha}^{VY} < \widehat{\alpha}^{HY}$; if also $\sum_{k \in N^p} s_k / n_p > \overline{s}$, then $\widehat{\alpha}^{VTY} < \widehat{\alpha}^{HTY}$.

Proof. See the appendix. ■

This proposition shows that the equalization parameter chosen by the federal government is higher with a horizontal ES than with a vertical one, with both pretax and taxable income as measures of fiscal capacity. That result is obtained under the following plausible assumptions. First, the sum over all the regions of the marginal benefit of the federal public good has to be greater than the single marginal benefit of a local public good for the representative consumer of each region. Second, the positive effect of regional taxation on the yield from monitoring activity at a federal and regional level, $nB' \sum_{k=1}^{n} \frac{\partial(1-\beta_k)D_k}{\partial \rho_k} - b' \sum_{k \notin N^p} \frac{\partial(\beta_k D_k)}{\partial \rho_k} > 0$, is negligible with respect to all other negative effects. Third, the average evasion of poor regions has to be greater than the national average.[26]

In order to grasp the intuition of this result, let us compare (48) and (49). On the right-hand sides of both equations, the first term represents the marginal benefit of α for the public revenues of the poor regions, and those terms are equal under vertical and horizontal ESs, both in the case of pretax earned income and in that of reported taxable income. Thus, the only difference in (48) and (49) is in the second term on the right-hand side. As is shown in the proof of the proposition, $nB' \frac{\partial R_F^V}{\partial \alpha} < b' \sum_{k \notin N^p} \frac{\partial R_k^H}{\partial \alpha}$, and this is due mostly to the fact that the distortive effect of regional taxation on the tax base is greater under a vertical than under a horizontal ES, because in the former case it affects all the regions, while in the latter it only affects rich regions. Thus, the net marginal benefit of α from the point of view of the federal government (the difference between the first and the second term on the right-hand sides of (48) and (49)) is lower under a vertical than under

26 For example, this is true for Italy. Agenzia delle Entrate provides estimates that evasion for every €100 of collected revenue is €65.87 in some poor regions in the south of Italy, while it is €10.93 in some rich regions in the north of Italy (the national average is €38.19). For more details, see http://www.corriere.it/economia/11_aprile_03/index.shtml.

a horizontal ES, implying that the optimal level of α is lower under a vertical than under a horizontal ES.

As a lower degree of equalization implies a lower incentive for overprovision of local public goods, we concentrate our attention only on the vertical ES. In this respect, we can state the following:

Proposition 4 If $\sum_{k \in N^p} s_k/n_p > \bar{s}$ and $\frac{\partial s_i}{\partial \rho_i} > 0$, then $\widehat{\alpha}^{VY} < \widehat{\alpha}^{VTY}$.

Proof. See the appendix. ■

This proposition shows under which conditions the equalization parameter chosen by the federal government is lower with a vertical ES based on pretax earned income than with a vertical ES based on reported taxable income.[27]

Finally, propositions 1 and 2 allow us to state the following corollary:

Corollary 1 Assume the following timing of decisions: first, the type of equalization is chosen (vertical versus horizontal), and then the index of fiscal capacity is chosen (pretax earned income versus reported taxable income). Then, under the assumptions of propositions 3 and 4, a vertical ES based on pretax earned income guarantees the minimal level of the equalization parameter.

This corollary confirms the result obtained in the previous section, i.e., a vertical ES based on pretax earned income lowers the incentives for overprovision of local public goods. On the one hand, from the point of view of regions, propositions 1 and 2 show that such a result holds for any choice of the equalization rate made by the federal government. On the other hand, corollary 1 compares the optimal rates of equalization chosen by the federal government under the different types of ES, and shows that the lowest one is associated with a vertical ES based on pretax earned income.

5. Concluding Remarks

When the central government is precommitted, it is well known that a vertical fiscal externality originating from an increase of a local surtax on a federal income tax base implies the overprovision of local public goods. This is due to the fact that local governments perceive a marginal cost of public funds lower than the (true) social one. This incentive further increases if there is in place an equalization system based on "fiscal capacity" like the one studied by Smart (1998) and now applied in many federal countries.

In this paper, we have analyzed how such an incentive to overprovide local public goods is changed when the model also allows for income evasion. Indeed, with tax sheltering the criterion to measure tax distortion and

27 For a discussion on the assumption $\frac{\partial s_i}{\partial \rho_i} > 0$, we refer to footnote 15.

excess burden must be changed in order to take into account the elasticity of taxable income instead of the elasticity of labor supply (Chetty, 2009; Saez et al., 2012). Consequently, the formula for the marginal cost of public funds must be changed too, and this has an impact on the overprovision concern. In this respect, we have shown how the local public-good overprovision phenomenon is affected by considering simultaneously tax evasion and different specifications of an equalization system, viz., vertical or horizontal, and based on pretax earned income or reported taxable income.

As far as the federal government's strategies are concerned, first, we have shown that the chosen degree of equalization is higher with a horizontal ES than with a vertical one, with both pretax and taxable income as measures of fiscal capacity, and second, that the ES based on pretax earned income has the lower degree of equalization of the vertical ESs.

Our main result shows that a more efficient level of regional public expenditures is more likely to be achieved with a vertical equalization system based on pretax earned income, because such an ES requires the lowest equalization rate. In this respect, two remarks are in order. First, our aim has been to examine the efficiency implications of different types of equalization schemes once they are in place, not to identify, from the point of view of the federal government, the optimal type of equalization scheme, which will depend on the federation's distributional objectives. Indeed, a vertical equalization system based on pretax earned income turns out to be the most efficient in terms of reduction of the incentive for overspending for regions, but it could imply a low level of redistribution of resources among regions via the equalization system fixed at a constitutional level. Thus, from a normative point of view, there could be a trade-off between equity and efficiency concerns. Of course, such a result relies on the assumption of a utilitarian federal government, and considering alternative tastes for redistribution may affect the results. Second, it should be stressed that such an equalization system may be more difficult to apply because it is more demanding of information to be gathered. However, as a proxy of pretax earned income at the regional level, some national-accounts indexes, like regional per capita GDP or NDP, can be used. Further, although the level of evasion is a household choice, it can be influenced by regional fiscal legislation under which the household makes its choices. Indeed, the level of evasion depends on local tax rates but can also be somewhat influenced by the struggle of local governments against underreporting of the tax base. Thus, the opportunity of limiting tax payments by local voters can be strategically played by local politicians to gain electoral consensus.

Finally, notice that the aim of our model has been very specific: to compare different types of equalization schemes on efficiency grounds alone. Thus, more general problems – as, for example, those related to corrective devices

that could be used to restore an optimal outcome – are not taken into account. The analysis of such issues within a more general approach than the present one is an open field for further research.

6. Appendix

6.1. Proof of Lemma 1

Lemma 1 If $\tau_i(1+f_i) \geq c_i'[p_i(\eta_i^*)]$ and $\frac{\partial s_i}{\partial \rho_i} > 0$, then $D_i(\eta_i^*) \geq 0$ and $\frac{\partial D_i(\eta_i^*)}{\partial \rho_i} > 0$, $\eta_i^* \equiv \operatorname{argmax}_{\eta_i} D_i(\eta_i)$.

Proof. The first-order condition for the maximization of (10) with respect to η_i is obtained as

$$\left\{\tau_i(1+f_i) - c_i'[p_i(\eta_i)]\right\}p'(\eta_i)s_i + \{p_i(\eta_i)(1+f_i)\tau_i - c_i[p_i(\eta_i)]\}\frac{\partial s_i}{\partial p_i}p'(\eta_i) = 0\,,$$

or

$$\tau_i(1+f_i) - c_i'[p_i(\eta_i)] - \left\{\tau_i(1+f_i) - \frac{c_i[p_i(\eta_i)]}{p_i(\eta_i)}\right\}\xi_i^p = 0\,,$$

where $\xi_i^p \equiv -\frac{\partial s_i}{\partial p_i}\frac{p_i}{s_i} > 0$, because s_i is decreasing in p_i for a risk-averse taxpayer. If $\tau_i(1+f_i) \geq c_i'[p_i(\eta_i^*)]$, it must be that $\tau_i(1+f_i) \geq \frac{c_i[p_i(\eta_i^*)]}{p_i(\eta_i^*)}$, and then $D_i(\eta_i^*) \geq 0$.

Let us now compute the following derivative:

$$\frac{\partial D_i(\eta_i^*)}{\partial \rho_i} = p_i(\eta_i^*)(1+f_i)s_i + \left\{p_i(\eta_i^*)(1+f_i)\tau_i - c_i\left[p_i(\eta_i^*)\right]\right\}\frac{\partial s_i}{\partial \rho_i}\,. \tag{50}$$

Thus, from (50), $D_i(\eta_i^*) \geq 0$, with $\frac{\partial s_i}{\partial \rho_i} > 0$, implies $\frac{\partial D_i(\eta_i^*)}{\partial \rho_i} > 0$. ■

6.2. Proof of Proposition 3

Proposition 3 If $nB' > b'$, and the positive effect of regional taxation on the yield from monitoring activity at the federal and regional levels is negligible with respect to all the other negative terms, then $\widehat{\alpha}^{VY} < \widehat{\alpha}^{HY}$; if also $\sum_{k\in N^p} s_k/n_p > \bar{s}$, then $\widehat{\alpha}^{VTY} < \widehat{\alpha}^{HTY}$.

Proof. In the case of pretax earned income, (21) implies that

$$\sum_{k\in N^p}\frac{\partial R_k^{VY}}{\partial \alpha} = \sum_{k\in N^p}\frac{\partial R_k^{HY}}{\partial \alpha}\,. \tag{51}$$

Thus, to compare the net marginal benefit with VY and HY, from (48) and (49), we need to check the sign of

$$nB' \frac{\partial R_F^V}{\partial \alpha} - b' \sum_{k \notin N^p} \frac{\partial R_k^H}{\partial \alpha}. \tag{52}$$

From (18), it is easy to check that

$$\begin{aligned} \frac{\partial R_F^{VY}}{\partial \alpha} = \sum_{k=1}^{n} \left(t \frac{\partial TY_k}{\partial \rho_k} + \frac{\partial (1-\beta_k) D_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha} + \alpha \overline{\rho} \sum_{k \in N^p} \frac{\partial Y_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} \\ + \overline{\rho} n_p \left(\frac{\sum_{k \in N^p} Y_k}{n_p} - \overline{Y} \right), \end{aligned} \tag{53}$$

and from (32), it is easy to check that

$$\begin{aligned} \sum_{k \notin N^p} \frac{\partial R_k^{HY}}{\partial \alpha} = \sum_{k \notin N^p} \left(\rho_k \frac{\partial TY_k}{\partial \rho_k} + \frac{\partial (\beta_k D_k)}{\partial \rho_k} + TY_k - \alpha \overline{\rho} \frac{\partial Y_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha} \\ - \overline{\rho} (n - n_p) \left(\frac{\sum_{k \notin N^p} Y_k}{n - n_p} - \overline{Y} \right). \end{aligned} \tag{54}$$

By using (53) and (54), (52) can be rewritten as

$$\Gamma + \Psi + \Omega - \sum_{k \notin N^p} TY_k \frac{\partial \rho_k}{\partial \alpha} + \overline{\rho} n_p (nB' - b') \left(\frac{\sum_{k \in N^p} Y_k}{n_p} - \overline{Y} \right), \tag{55}$$

where

$$\begin{aligned} \Gamma &\equiv \left(nB' t \sum_{k=1}^{n} \frac{\partial TY_k}{\partial \rho_k} - b' \sum_{k \notin N^p} \rho_k \frac{\partial TY_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha}, \\ \Psi &\equiv \alpha \overline{\rho} \left(nB' \sum_{k \in N^p} \frac{\partial Y_k}{\partial \rho_k} + b' \sum_{k \notin N^p} \frac{\partial Y_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha}, \\ \Omega &\equiv \left(nB' \sum_{k=1}^{n} \frac{\partial (1-\beta_k) D_k}{\partial \rho_k} - b' \sum_{k \notin N^p} \frac{\partial (\beta_k D_k)}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha}. \end{aligned}$$

The sign of (55) is negative because $\Gamma < 0$, $\Psi < 0$, $\sum_{k \notin N^p} TY_k \frac{\partial \rho_k}{\partial \alpha} > 0$, and $\sum_{k \in N^p} Y_k / n_p - \overline{Y} < 0$, and under the assumptions that $nB' > b'$ and $\Omega > 0$ is negligible with respect to the negative first, second, and fourth terms. Thus, $\widehat{\alpha}^{VY} < \widehat{\alpha}^{HY}$.

Mutatis mutandis, the same reasoning can be performed to show that $\widehat{\alpha}^{VTY} < \widehat{\alpha}^{HTY}$. This result arises for the same reasons as above, plus the assumption that $\sum_{k \in N^p} s_k / n_p > \overline{s}$. ■

6.3. Proof of Proposition 4

Proposition 4 $\sum_{k \in N^p} s_k / n_p > \overline{s}$ and $\frac{\partial s_i}{\partial \rho_i} > 0$, then $\widehat{\alpha}^{VY} < \widehat{\alpha}^{VTY}$.

Proof. To compare the net marginal benefit with VY and VTY, from (48), we need to check the sign of

$$-b' \sum_{k \in N^p} \frac{\partial R_k^{VY}}{\partial \alpha} - nB' \frac{\partial R_F^{VY}}{\partial \alpha} + b' \sum_{k \in N^p} \frac{\partial R_k^{VTY}}{\partial \alpha} + nB' \frac{\partial R_F^{VTY}}{\partial \alpha}. \tag{56}$$

From (28), it is easy to check that

$$\begin{aligned} \frac{\partial R_F^{VTY}}{\partial \alpha} = t \sum_{k=1}^{n} \frac{\partial TY_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} + \sum_{k=1}^{n} \frac{\partial (1-\beta_k) D_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} - n_p \overline{\rho}\,\overline{TY} \\ + \overline{\rho} \sum_{k \in N^p} TY_k + \alpha \overline{\rho} \sum_{k \in N^p} \frac{\partial TY_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha}. \end{aligned} \tag{57}$$

Further, for poor regions, from (16) and (26), it is easy to check that

$$\begin{aligned} \sum_{k \in N^p} \frac{\partial R_k^{VY}}{\partial \alpha} = \sum_{k \in N^p} \Bigg[\left(TY_k + \rho_k \frac{\partial TY_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha} + \frac{\partial (\beta_k D_k)}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} \\ - \overline{\rho} Y_k - \alpha \overline{\rho} \frac{\partial Y_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} + \overline{\rho}\,\overline{Y} \Bigg], \end{aligned} \tag{58}$$

$$\begin{aligned} \sum_{k \in N^p} \frac{\partial R_k^{VTY}}{\partial \alpha} = \sum_{k \in N^p} \Bigg[\left(TY_k + \rho_k \frac{\partial TY_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha} + \frac{\partial (\beta_k D_k)}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} \\ - \overline{\rho} TY_k - \alpha \overline{\rho} \frac{\partial TY_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial \alpha} + \overline{\rho}\,\overline{TY} \Bigg]. \end{aligned} \tag{59}$$

By using (53), (57), (58), and (59), (56) can be rewritten as

$$n_p \left(\overline{s} - \frac{\sum_{k \in N^p} s_k}{n_p} \right) - \alpha \sum_{k \in N^p} \left(\frac{\partial Y_k}{\partial \rho_k} - \frac{\partial TY_k}{\partial \rho_k} \right) \frac{\partial \rho_k}{\partial \alpha} < 0,$$

because we have assumed that $\sum_{k \in N^p} s_k / n_p > \overline{s}$ and $\frac{\partial Y_i}{\partial \rho_i} - \frac{\partial TY_i}{\partial \rho_i} = \frac{\partial s_i}{\partial \rho_i} > 0$, $i = 1, ..., n$. Thus, $\widehat{\alpha}^{VY} < \widehat{\alpha}^{VTY}$. ■

References

Baretti, C., Huber, B., and Lichtblau, K. (2002), A Tax on Tax Revenue: The Incentive Effects of Equalizing Transfers: Evidence from Germany, International Tax and Public Finance 9, 631–651.

Bartolini, D., and Fiorillo, F. (2009), Central and Local Tax Enforcement, in Longobardi, E., Petretto, A. (Eds.), Saggi di economia della tassazione, F. Angeli, Milan, 163–180.

Boadway, R. (2004), The Theory and Practice of Equalization, CESifo Economic Studies 50, 211–254.

Boadway, R., and Tremblay, J.-F. (2010), Mobility and Fiscal Imbalance, National Tax Journal 63, 1023–1054.

Boadway, R., Cuff, K., and Marchand, M. (2002), Equalization and the Decentralization of Revenue-Rasing in a Federation, Journal of Public Economic Theory 5, 201–228.

Bucovetsky, C., and Smart, M. (2002), The Efficiency Consequences of Local Revenue Equalization: Tax Competition and Tax Distortion, CESifo Working Paper No. 767.

Buettner, T. (2006), The Incentive Effect of Fiscal Equalization Transfers on Tax Policy, Journal of Public Economics 90, 477–497.

Chetty, R. (2009), Is Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance, American Economic Journal: Economic Policy 2, 31–52.

Cremer, H., and Gahvari, F. (2000), Tax Evasion, Fiscal Competition and Economic Integration, European Economic Review 44, 1633–1657.

Dahlby, B. (2008), The Marginal Cost of Public Funds: Theory and Applications, MIT Press, Cambridge, US.

Dahlby, B., and Warren, N. (2003), Fiscal Incentive Effects of the Australian Equalization System, Economic Record 79, 434–445.

Egger, P., Köthenbürger, M., and Smart, M. (2010), Does Fiscal Transfers Alleviate Business Tax Competition? Evidence from Germany, Journal of Public Economics 94, 235–246.

Esteller-Moré, A., and Solé-Ollè, A. (2002), Tax Setting in a Federal System: The Case of Personal Income Taxation in Canada, International Tax and Public Finance 9, 235–257.

Figuieras, C., Hindricks, J., and Myles, G. (2004), Revenue Sharing versus Expenditure Sharing in a Federal System, International Tax and Public Finance 11, 155–175.

Goodspeed, T. (2002), Tax Competition and Tax Structure in Open Federal Economies: Evidence from OECD Countries with Implication for the European Union, European Economic Review 46, 357–374.

Grazzini, L., and Petretto A. (2006), Vertical Tax Competition with Tax Sharing and Equalization Grants, Giornale degli Economisti e Annali di Economia 65, 75–93.

Hindriks, J., Peralta, S., and Weber, S. (2008), Competing in Taxes and Investment under Fiscal Equalization, Journal of Public Economics 92, 2392–2402.

Keen, M. J. (1998), Vertical Tax Externalities in the Theory of Fiscal Federalism, International Monetary Fund Staff Papers 45, 454–485.

Keen, M. J., and Kotsogiannis, C. (2002), Does Federalism lead to Excessively High Taxes?, American Economic Review 92, 363–370.

Kelders, C., and Köthenbürger, M. (2010), Tax Incentives in Fiscal Federalism: An Integrated Perspective, Canadian Journal of Economics 43, 683–703.

Köthenbürger, M. (2002), Tax Competition and Fiscal Equalization, International Tax and Public Finance 9, 391–408.

Köthenbürger, M. (2008), Federal Tax-Transfer Policy and Intergovernmental Pre-Commitment, Regional Science and Urban Economics 38, 16–31.

Köthenbürger, M. (forthcoming), How Do Local Governments Decide on Public Policy in Fiscal Federalism? Tax vs Expenditure Optimization, Journal of Public Economics.

Kotsogiannis, C. (2010), Federal Tax Competition and the Efficiency Consequences for Local Taxation of Revenue Equalization, International Tax and Public Finance 17, 1–14.

Kotsogiannis, C., and Schwager, R. (2008), Accountability and Fiscal Equalization, Journal of Public Economics 92, 2336–2349.

Saez, E., Slemrod, J., and Gierz, S. (2012), The Elasticity of Taxable Income with Respect to Marginal Tax Rates: a Critical Review, Journal of Economic Literature 50, 3–50.

Smart, M. (1998), Taxation and Deadweight Loss in a System of Intergovernmental Transfers, Canadian Journal of Economics 1, 188–205.

Smart, M. (2007), Raising Taxes through Equalization, Canadian Journal of Economics 40, 1188–1212.

Stöwhase, S., and Traxler, C. (2005), Tax Evasion and Auditing in a Federal Economy, International Tax and Public Finance 12, 515–533.

Traxler, C., and Reutter, A. (2008), Apportionment, Fiscal Equalization and Decentralized Tax Enforcement, Max Planck Institute for Research and Collective Goods Working Paper No. 2008/21.

Lisa Grazzini
University of Florence
Department of Economics
Via delle Pandette 9
50127 Florence
Italy
lisa.grazzini@unifi.it