TAX COMPETITION AND THE EFFICIENCY OF “BENEFIT-RELATED” BUSINESS TAXES

Elisabeth Gugl
Department of Economics, University of Victoria, Victoria, British Columbia, Canada

George R. Zodrow
Cline Chair of Economics and
Faculty Scholar, Center for Public Finance, Baker Institute for Public Policy,
Rice University, Houston TX, USA, and
International Research Fellow, Centre for Business Taxation,
Oxford University, Oxford, UK

A revised version of this paper is forthcoming in *International Tax and Public Finance*

Abstract: We construct a tax competition model in which local governments finance business public services with either a source-based tax on mobile capital, such as a property tax, or a tax on production, such as an origin-based Value Added Tax, and then assess which of the two tax instruments is more efficient. Many taxes on business apply to mobile inputs or outputs, such as property taxes, retail sales taxes, and destination-based VATs, and their inefficiency has been examined in the literature; however, proposals from several prominent tax experts to utilize a local origin-based VAT have not been analyzed theoretically. Our primary finding is that the production tax is less inefficient than the capital tax under many — but not all — conditions. The intuition underlying this result is that the efficiency of a user fee on the public business input is roughly approximated by a production tax, which applies to both the public input and immobile labor (in addition to mobile capital). In marked contrast, the capital tax applies only to mobile capital and is thus likely to be relatively inefficient.

Acknowledgements: We thank an anonymous referee, Steve Billon, Jed Brewer, Alexander Ebertz, Bill Fox, Timothy Goodspeed, Brad Hackinen, Andreas Haufler, James Hines, Margaret McKeehan, Matt Krzepkowski, David Wildasin, and John Wilson for helpful comments on earlier versions of this paper.

1. Introduction

At the state and local levels in the United States and at the provincial level in Canada, retail sales taxes often apply tax to business inputs, even though in principle they should be limited to final consumer goods. However, eliminating such taxes on business can provoke strong opposition from the electorate. For example, in 2010 British Columbia moved from a retail sales tax that was known to tax business inputs due to problems in its administration to a “pure” consumption tax — the Harmonized Sales Tax (HST), which combined the provincial retail sales tax and the federal Goods and Services Tax into a single value-added tax. This reform was projected to decrease tax payments from business by Canadian $730 million (The Independent Panel 2011), prompting critics of the HST to decry its shift away from taxing business (CBC News 2011). Premier Gordon Campbell introduced the HST shortly after winning office, despite denying any intent to do so during his campaign; as a result, the way the provincial government introduced the change also sparked strong opposition to the replacement of the provincial sales tax with the HST. The opposition was so fierce that continuation of the HST was the subject of a mail-in referendum, in which 55% of voters opposed the HST. As a result, British Columbia abandoned the HST and had to make the costly transition to re-establishing its provincial retail sales tax.

Motivated partly by the political argument that business should pay its “fair share” of taxes and partly by the economic argument that efficiency requires that businesses pay for the public services they consume (e.g., by paying user charges or being assessed “benefit taxes”), we analyze the impacts of imposing a production tax or a capital tax on local business in order to finance a public service that serves as an input in firms’ production functions. Because our model is characterized by a single consumption good and fixed labor supply, a consumption tax such as a destination-based VAT would result in an efficient equilibrium. Nevertheless, we analyze production and capital taxes on business within the context of this model because we are interested in the impact of such narrower (capital tax) and broader (production tax) levies on business and government, including their effects on the level of public services. These issues are especially relevant in light of common political perceptions that firms should pay their "fair share" for public services they utilize as well as the economic efficiency argument for requiring firms to pay benefit taxes or user charges, and in light of the widespread use of subnational taxes on business in practice.

Indeed, in addition to taxing retail sales of business inputs, subnational taxes in the United States and Canada and in many other countries often include property taxes on business capital as well. Both forms of business input taxation may distort a wide variety of decisions, including those regarding production, capital accumulation and allocation, and, as stressed in the tax competition literature, the level of public services.

In light of this situation, several prominent public finance experts have argued that, in the absence of the assessment of explicit benefit taxes or user charges on business, broad-based taxes on local production, such as an origin-based value added tax (VAT), are an attractive option (Bird 2000; Hines 2003). Their primary rationale is that such taxes serve as a proxy for the user charges that are desirable on economic efficiency grounds — that is, they are relatively efficient "benefit-related" taxes.

In this paper, we provide what we believe is the first attempt to analyze systematically the assertion that a production-based tax can be viewed as a proxy for a benefit tax, and is thus preferable to the often-used alternative of direct or indirect taxes on business inputs, especially capital. Our paper provides the theoretical basis underlying the simulation results presented in the context of CES production functions in Gugl and Zodrow (2015).

Many papers analyze the efficiency of capital taxation when firms receive public services and jurisdictions compete with each other for mobile capital; a partial list includes Zodrow and Mieszkowski (henceforth ZM) (1986), Oates and Schwab (1988, 1991), Noiset (1995), Sinn (1997), Keen and Marchand (1997), Bayindir-Upmann (1998), Richter (1994), Matsumoto (1998, 2000), and Dhillon, Wooders and Zissimos (2007). However, these papers focus on the inefficiency of capital taxation alone, and do not consider a tax on local production or compare such a production tax to a capital tax.

Our results provide some support for the idea that a production-based tax may be viewed as a “benefit-related” tax, although only under special circumstances does a production tax substitute perfectly for an explicit user charge. The characteristics of the firm’s production function — specifically, its log modularity properties (discussed at length below) — play a major role in determining the relative efficiency properties of production-based taxes and capital taxes, an issue that has thus far not been examined in models of interjurisdictional tax competition.

In our analysis, we assume that output is produced with a constant returns to scale (CRS) production function using labor, capital, and business public services as inputs. The log modularity properties of the production function with respect to capital and business public services indicate how the elasticity of output with respect to increases in business public services varies at different levels of the amount of capital employed in production. Specifically, if the production function is log modular (log submodular, log supermodular) in capital and business public services, then the elasticity of output with respect to business public services is constant (decreases, increases) as the capital employed in production increases. Note that all three possibilities are consistent with a CRS production function; in particular, a linearly homogenous Cobb-Douglas production function is log modular, while a constant elasticity of substitution

(CES) production function is log submodular (log supermodular) if the elasticity of substitution is greater than (less than) one.

In the special case of a production function that is log modular in capital and the public service, a production tax is effectively a benefit tax, and is thus analogous to a user charge for public services that ensures an efficient level of public service provision. In the same vein, if the production function is log submodular in capital and public services (e.g., a CES production function with an elasticity of substitution greater than one), a production tax is inefficient and leads to underprovision of the public service, but it results in less underprovision and thus is less inefficient than a capital tax. However, although we can show that the production tax leads to overprovision of public services in the case of a log supermodular production function (e.g., a CES production function with an elasticity of substitution smaller than 1), the ambiguity of the effect of a capital tax on public services in this case implies that a ranking of the relative efficiency properties of the two taxes is impossible without further restrictions on the production technology. In our companion paper (Gugl and Zodrow 2015), we provide a wide variety of simulations using the CES production function and find only modest efficiency gains in the instances in which capital taxes are more efficient than the production tax.

There is also a vast literature analyzing the relative efficiency of destination-based and origin-based value-added taxes (VATs). In a closed economy, there is no difference between a uniform tax on all consumption and a similar tax on production. In contrast, as economies become increasingly more open, the distinction between a tax on local production and a tax on local consumption becomes important in terms of efficiency (Mintz and Tulkens 1986; Kanbur and Keen 1993; Lockwood 2001; Haufler and Pflueger 2007). In tax competition models in which firms are perfectly competitive, a tax on local consumption such as a destination-based

VAT is efficient when countries are too small to affect world prices (Lockwood 2001). Haufler and Pflueger (2007) investigate the difference between a destination-based and an origin-based VAT in several settings of international duopoly and find that only the former is efficient when competition between countries is imperfect. In contrast, McLure (2003) notes that the growing importance of electronic commerce implies that a destination-based VAT is increasingly difficult to administer, providing an argument in favor of an origin-based VAT.[1]

In practice, most existing value-added taxes and retail sales taxes are destination-based taxes. However, there are numerous exceptions to this general rule, that is, cases in which these taxes are assessed on an origin basis and are thus similar to the origin-based production tax analyzed in this paper (Bird 2003). For example, 11 U.S. states tax intrastate sales on an origin basis (a business selling a good or service to a consumer in a different taxing jurisdiction in the same state is taxed in the home jurisdiction of the business), while California has a mixed system in which state, county, and municipal taxes are assessed on an origin basis but special districts use a destination-based approach.[2] Similarly, the subnational VATs in Brazil and Argentina and the regional business tax in Italy (imposta regionala sulle attività produttive or IRAP) are origin-based VAT, as is the VAT in Japan (Bird and Gendron 2007; Crawford, Keen and Smith 2010). Finally, note that the Hall and Rabushka (2007) flat tax proposal in the United States, as well as most of its many progeny which are still attracting the interest of some proponents of tax reform, is essentially a "two-part" origin-based VAT.[3]

---

[1] This problem has been mitigated recently in the United States by the decision of the U.S. Supreme Court in South Dakota v. Wayfair Inc., which specifies that remote sellers can under certain circumstances be required to collect sales taxes even in states in which they have no physical presence.

[2] The 11 states with origin-based treatment of intrastate sales are Arizona, Illinois, Mississippi, Missouri, New Mexico, Ohio, Pennsylvania, Tennessee, Texas, Utah, and Virginia (Faggiano 2017).

[3] The Hall-Rabushka flat tax uses the "subtraction" method rather than the almost universally used "invoice-credit" VAT method. Under the subtraction method wages are deductible from the tax base at

Despite the vast literature on different forms of VATs, an origin-based VAT has not been analyzed in a model of tax competition in which public services are provided to firms. To the best of our knowledge, this paper, along with Gugl and Zodrow (2015), is the first to examine this issue.

In the next two sections, we present the model and discuss our results on the production tax. We then contrast these results with the well-known properties of capital taxation in section 4. Section 5 compares the efficiency properties of a capital tax and a production tax, and Section 6 concludes. We provide a detailed discussion of log modularity in production in the Appendix.

2. The Model

Our model follows the ZM (1986) framework.[4,5] A federation or union consists of N jurisdictions, each with the same number of residents who are immobile across jurisdictions. All residents have identical preferences and endowments. Individuals work where they live, provide a fixed amount of labor, $L,$ and obtain utility from consumption of an aggregate composite good, X. The labor supply of each jurisdiction is therefore fixed.[6] People own an equal share of the union's capital stock $\bar{K}$, which is fixed in total supply at $N\bar{K}$.

Labor and capital are the private inputs in the production of the consumption good. In addition, the local government provides a fully congestible business public service, denoted by B_i, that is used directly in the production of the consumption good. Each jurisdiction produces with a technology characterized by constant returns to scale (CRS) in the two private inputs and

the business level but taxed, subject to a standard deduction and personal exemptions, at the individual level.

4 Gugl and Zodrow (2015) provide further discussion of the assumptions of the model.

5 See also Wilson (1986). For reviews of the tax competition literature, see Wilson (1999), Wilson and Wildasin (2004), and Zodrow (2003, 2010).

6 The fixed factor can also be thought of as a combination of labor and land, as assumed by ZM.

the public input. We assume throughout the paper that the production function is strictly concave in capital and the business public service. The consumption good is assumed to be tradable and is taken as the numeraire. We assume that the number of firms is fixed in each jurisdiction (or equivalently, given CRS, that there is a single representative firm) and focus therefore on the aggregate production function in each jurisdiction, given by $F(B_i, K_i, L)$. The marginal product of each input is positive and strictly decreasing and cross derivatives are positive.[7] That is, $F_B, F_K, F_L > 0, F_{BB}, F_{KK}, F_{LL} < 0$, and $F_{BK}, F_{BL}, F_{KL} > 0$. The consumption good can be transformed into the public service at a constant marginal rate of transformation of 1, so the unit cost of B is also equal to one,[8] i.e., for the economy as a whole

$$\sum_{i=1}^{N} F(B_i, K_i, L) = \sum_{i=1}^{N} X_i + B_i. \tag{1}$$

Before we analyze the decisions of the governments of the local jurisdictions, we derive the efficient amount of business public services in this economy, denoted by $\mathcal{B}$ and state the conditions that lead to efficient local production of business public services, $\mathcal{B}$, in each jurisdiction.

To create a benchmark, we begin by calculating the efficient amount of business public services if all inputs can move freely across jurisdictions. Given CRS,

$$F(\mathcal{B}, N\bar{K}, NL) = NF(\mathcal{B}/N, \bar{K}, L) \tag{2}$$

This function implies that production is independent of where labor is located. Although we assume labor is fixed and identically distributed across jurisdictions, for now we neglect this

[7] See, e.g., ZM, Bayindir-Upmann (1998), Keen and Marchand (1997), Dhillon et al. (2007) for the same assumptions.

[8] We follow most of the literature in assuming constant marginal costs for business public services (Oates and Schwab 1988 and 1991; Sinn 1997, Bayindir-Upmann 1998; Keen and Marchand 1997; Richter 1994; and Matsumoto 2000). Two alternative approaches, which Matsomoto (2000) points out are equivalent, would be to assume either an imperfectly congestible public input and a constant marginal cost of producing that public input, or a perfectly congestible public input (i.e., our publicly provided private service) and decreasing marginal costs of producing the public service.

constraint to determine the allocation of labor, capital, and business public services that is efficient without jurisdiction-specific constraints. Since the total amounts of capital and labor are fixed in the whole economy, and consumers care only about the consumption good, the sole choice variable in determining efficiency is how much business public services should be produced to maximize consumption, that is,

$$\max_{\mathcal{B}} F(\mathcal{B}, N\overline{K}, NL) - \mathcal{B}. \qquad (3)$$

The first order condition is

$$F_{\mathcal{B}} - 1 = 0, \qquad (4)$$

which simply indicates that the efficient aggregate amount of business public services occurs where the marginal product of such services equals the marginal rate of transformation of 1. Given our assumption of identical jurisdictions and fixed labor, the efficient amount of capital in each jurisdiction is equal to $\overline{K}$ and the efficiency condition for public services in each jurisdiction is

$$F_B(B, \overline{K}, L) - 1 = 0 \qquad (5)$$

In analyzing the impact of financing business public services either with a capital tax or a tax on production, we compare the marginal productivity of business public services in a jurisdiction under either of the two taxes with the efficient marginal productivity of one. If the marginal productivity of business public services is smaller (larger) than one, there is over (under) provision as, in equilibrium, the amount of public services that each jurisdiction provides is larger (smaller) than the efficient amount. Next, we turn to the financing of business public services with a production tax.

3. Production Tax

Suppose that local jurisdictions tax local production to finance the provision of public services to firms. The amount of capital a local jurisdiction might attract depends on the level of business public services and the production tax rate t_i; hence, K is a function of the policy mix B_i, t_i. We derive below the perceived change in capital in a jurisdiction as it increases its production tax to finance an increase in its level of business public services. For now, we simply note that capital is perceived to change with changes in the policy mix, i.e., $K_i(B_i, t_i)$.

The local government's budget constraint is given by

$$B_i = t_i F(B_i, K_i(B_i, t_i), L) + H, \tag{6}$$

where H is a fixed amount of public services that is financed externally. We will focus on the jurisdiction's problem when $H = 0$, but it is useful to consider the case of $H > 0$ when interpreting the Lagrange multiplier λ below.

Firms maximize profits, given factor prices, the tax on production, and the level of local business public services provided. They demand both labor and capital, but since labor is fixed in each jurisdiction, the local labor market will clear at L. Since capital is perfectly mobile across jurisdictions, the after-tax rate of return to capital, r, is the same in every jurisdiction. The jurisdiction takes as given the firm's profit maximization condition

$$(1 - t)F_K(B_i, K_i(B_i, t_i), L) = r. \tag{7}$$

We assume the local government acts to maximize the welfare of its representative resident, who owns all of the fixed labor input, and that any profits earned from production due to the provision of business public services will accrue to that resident as well. Thus, the profits in any given jurisdiction enter the local government's social welfare function together with locally earned wages and the capital income earned by its resident. To simplify the notation below, we suppress the subscript i in analyzing a single local government's optimization problem, which is to choose the production tax rate and the level of business public services to

$$\max_{B,t}[(1-t)F(B,K(B,t),L)-rK(B,t)]+r\bar{K}, \tag{8}$$

subject to equations (6) and (7).

The term in square brackets of (8) is the sum of wage income and profits generated in the jurisdiction. The Lagrange function for the government's optimization problem is

$$\max_{B,t}[(1-t)F(B,K(B,t),L)-rK(B,t)]+r\bar{K}+\lambda[H+tF(B,K(B,t),L)-B]. \tag{9}$$

The first order conditions with respect to B, t taking (7) into account are

$$(1-t)F_B+\lambda\left(tF_B+tF_K\frac{\partial K}{\partial B}-1\right)=0 \tag{10}$$

$$-F+\lambda\left(F+tF_K\frac{\partial K}{\partial t}\right)=0. \tag{11}$$

Note that at the optimal production tax and level of business public services, an increase in H and hence less reliance on self-financing of business public services leads to an increase in residents' income/consumption equal to λ.

Consider first the hypothetical case in which capital is fixed in a jurisdiction, i.e., $\partial K/\partial B$ and $\partial K/\partial t$ are zero. In this case, the first order conditions imply that the local jurisdiction provides business public services such that $F_B=1$ and that $\lambda=1$. By the Envelope Theorem, the Lagrange multiplier measures how much residents' income would increase if local jurisdictions faced a "soft" budget constraint and were given one unit of public services for free. Given that the Lagrange multiplier is 1 when local jurisdictions do not perceive a change in the capital in their jurisdiction, residents' income would go up by 1, which is equal to the marginal rate of transformation between the consumption good and business public services; as local governments receive a unit of the business public service for free, they save the marginal cost of business public services.

By comparison, if capital reacts to changes in the level of business public services and the production tax, then $\partial K/\partial B$ and $\partial K/\partial t$ are derived as follows. Differentiating (7) with respect to the production tax and capital and solving for $\partial K/\partial B$ yields

$$\frac{\partial K}{\partial B} = \frac{-F_{KB}}{F_{KK}} > 0. \qquad (12)$$

Differentiating (7) with respect to business public services and capital and solving for dK/dt yields

$$\frac{\partial K}{\partial t} = \frac{F_K}{(1-t)F_{KK}} < 0. \qquad (13)$$

Thus, from the local government's perspective, an increase in the production tax without an increase in public services would drive out capital, and an increase in public services without an increase in the production tax would attract capital to the jurisdiction. These two perceived distortions created by mobile capital need to be considered in setting the optimal production tax rate and the associated level of business public services. Note that the two distortions are characterized as perceived in the sense that, in equilibrium, each jurisdiction follows the same policies and thus adopts the same production tax rate, so that each jurisdiction receives an equal share of the total (mobile) national capital stock.

If we were to impose a consumption tax, *c*, instead of the production tax, this would be equivalent to imposing an income tax on residents on all sources of income. Thus would yield a budget balance condition of

$$B = c[F(B, K, L) - rK + r\overline{K}] + H,$$

and a profit maximization constraint

$$F_K = r.$$

With these changes to our model, a consumption tax, ie., a destination-based VAT, would lead to efficiency as all sources of distortion would be removed. However, as discussed in our introduction, we focus on taxes that are levied on business to examine the extent to which they can be viewed as benefit-related taxes.

Returning to our analysis of a production tax, substituting from (12) and (13) into the first order conditions (10-11) yields the optimal tax rate

$$t = \frac{(1-F_B)F_{KK}F}{F_K(F_B F_K - F F_{KB})}. \tag{14}$$

Note that by Young's Theorem, $F_{KB} = F_{BK}$. The elasticity of the marginal productivity of capital with respect to an increase in capital is $\varepsilon_{F_K,K} = -F_{KK}K/F_K$, the elasticity of the marginal productivity of public services with respect to an increase in capital is $\varepsilon_{F_B,K} = F_{BK}K/F_B$, and the output elasticity of capital is $\varepsilon_K = F_K K/F$. Using these elasticities, (14) becomes

$$t = \frac{(1-F_B)\varepsilon_{F_K,K}}{F_B(\varepsilon_{F_B,K} - \varepsilon_K)}. \tag{15}$$

Whether the production tax rate leads to overprovision or underprovision of business public services therefore depends on the difference between $\varepsilon_{F_B,K}$ and ε_K. The numerator of (15) is positive if there is overprovision of public services, i.e., if $F_B < 1$, and an interior solution for the production tax rate must thus be characterized by a positive difference between $\varepsilon_{F_B,K}$ and ε_K. We summarize our finding below.

Proposition 1: *With a positive production tax rate, overprovision of business public services occurs if at* $F_B < 1$, $\varepsilon_{F_B,K} > \varepsilon_K$. *Underprovision occurs if at* $F_B > 1$, $\varepsilon_{F_B,K} < \varepsilon_K$.

In the case in which production functions are log submodular or log supermodular, $\varepsilon_{F_B,K}$ and ε_K exhibit the same relationship to each other, regardless of the amounts of *B*, *K*, and *L*.

Lemma 1: *A production function is log sub (super) modular if and only if* $\varepsilon_{F_B,K} < \varepsilon_K$ $(\varepsilon_{F_B,K} > \varepsilon_K)$.

The proof of Lemma 1 is provided in the appendix, which includes a detailed discussion of log sub (super) modularity.

In the simulations performed in Gugl and Zodrow (2015), we use the CES production function with an elasticity of substitution less than 1 as an example of a log supermodular production function and the CES production function with an elasticity of substitution greater than 1 as an example of a log submodular production function. We provide more examples of log sub (super) modular production functions in the Appendix.

Our next result follows immediately from Lemma 1 and Proposition 1.

Corollary 1: *A production tax is inefficient whenever the production function is log submodular or log supermodular in capital and business public services. Log submodularity (supermodularity) leads to underprovision (overprovision) of business public services.*

Note that (15) cannot be used to assess the interior solution if the difference between $\varepsilon_{F_B,K}$ and ε_K is zero. Instead, we take a different approach to assess this case and ask what would happen when a budget-balanced increase in the tax rate and public services is considered, in which case both budget balance and the firm's profit maximization condition, that is, equations (6) and (7), must hold. Totally differentiating both equations with respect to the production tax, capital, and the public service, yields

$$\frac{dK}{dt} = \frac{(1-tF_B)F_K-(1-t)F_{KB}F}{(1-t)(F_{KB}tF_K+(1-tF_B)F_{KK})}. \quad (16)$$

Suppose jurisdictions choose the efficient amount of business public services, i.e. $F_B = 1$. In this case (16) becomes

$$\frac{dK}{dt} = \frac{F_K-F_{KB}F}{(F_{KB}tF_K+(1-t)F_{KK})}. \quad (17)$$

Governments will set the level of business public services efficiently only if they are assured that no capital flows in or out of their jurisdiction at the tax rate that leads to $F_B = 1$. Put differently, $dK/dt = 0$, if at $F_B = 1$ the numerator of (17) is zero and hence $F_B F_K = F_{KB} F$ at $F_B = 1$. In terms of elasticities, we need the difference between $\varepsilon_{F_B,K}$ and ε_K to be zero at $F_B = 1$.

We summarize our result in the next proposition.

Proposition 2: *With a positive production tax rate, efficient provision of business public services occurs if at* $F_B = 1$, $\varepsilon_{F_B,K} = \varepsilon_K$.

Lemma 2: *The equality* $\varepsilon_{F_B,K} = \varepsilon_K$ *is always satisfied when the production function is log modular in B and K.*

The proof is in the Appendix, where we provide examples of log modular production functions. As noted previously, a prominent example of the log modular case is the Cobb-Douglas function.

Corollary 2: *If a production function is log modular, a production tax is efficient.*

The intuition underlying these results is as follows. The government of the jurisdiction considering an increase in the production tax balances two factors. First, an increase in the tax will reduce profits, and thus tends to drive capital out of the jurisdiction. Second, the increase in business public services financed with the tax will increase the productivity of capital (generating more output at any given level of capital), and thus attract capital to the jurisdiction, which will in turn increase the productivity of local business public services. The case of log modularity corresponds to a production function for which these two effects are exactly offsetting. In this case, the increase in the capital stock attracted by the increase in public services generates an "average" increase in the productivity of public services, that is, one equal to the increase in output due to the same increase in capital.

By comparison, in the case of a log supermodular production function, the production tax increase has an additional benefit: the production function is such that, as the capital stock increases in response to the provision of additional public services, the productivity of public services increases more than proportionately. This additional benefit implies that the total marginal benefit to capital of the additional production-tax-financed public services outweighs the cost of paying the additional tax, and the production tax attracts capital to the jurisdiction. This ability to attract capital is sufficient to cause the government of the jurisdiction to overprovide business public services in order to attract additional capital (it does not take into account the cost to other jurisdictions that are losing capital as a result of their policy).

These results are reversed in the case of a log submodular production function, as the productivity of public services increases less than proportionately (relative to the increase in output) as additional capital is attracted to the jurisdiction with production-tax-financed public services, in which case the tax effect outweighs the productivity effect, and capital is driven out of the jurisdiction. As noted above, these two distortions attributable to the mobility of capital are only perceived distortions in the sense that, in equilibrium, each jurisdiction follows the same policies and thus adopts the same production tax rate, so that each jurisdiction receives an equal share of the total (mobile) national capital stock.

4. Capital Tax

Many authors have analyzed capital taxes using the original ZM framework.[9] Our analysis of capital taxation follows the approach used above to analyze the production tax in order to facilitate comparisons of the effects of the capital tax to those of the production tax.

[9] For a detailed literature review, see Gugl and Zodrow (2015).

Suppose that local governments impose a tax on capital at rate τ to finance business public services. The firm's condition for profit maximization under the capital tax is

$$F_K(B, K(B,\tau), L) = r + \tau. \qquad (18)$$

Analogous to the production tax case, the jurisdiction solves the problem

$$\max_{B,\tau}[F(B, K(B,\tau), L) - (r+\tau)K] + r\overline{K} + \mu[H + \tau K - B], \qquad (19)$$

subject to (18),where the first term in square brackets is the sum of wage income and profits generated in the jurisdiction and the second term in square brackets reflects the jurisdiction's budget constraint. The first order conditions for the government's optimization problem are

$$F_B + \mu\left(\tau\frac{\partial K}{\partial B} - 1\right) = 0 \qquad (20)$$

$$-K + \mu\left(K + \tau\frac{\partial K}{\partial \tau}\right) = 0. \qquad (21)$$

As in the previous section, if the capital stock is unchanged in the jurisdiction, $F_B = \mu = 1$. The jurisdiction takes the firm's profit maximization condition as given when deriving its perceived change in capital due to changes in the capital tax rate and the level of business public services. Differentiating (19) with respect to the capital tax and capital and solving for $\partial K/\partial \tau$ yields

$$\frac{\partial K}{\partial \tau} = \frac{1}{F_{KK}} < 0. \qquad (22)$$

The reaction of capital to an increase in the level of public services is the same under both taxes and is given by (12).

From (12), (20), (21) and (22) we find the optimal capital tax

$$\tau = \frac{-F_{KK}K(1-F_B)}{F_{KB}K - F_B}. \qquad (23)$$

Expressing the optimal capital tax (23) in terms of the elasticity of the marginal productivity of business public services with respect to an increase in capital,

$$\tau = \frac{F_K(1-F_B)\varepsilon_{F_K,K}}{F_B(\varepsilon_{F_B,K}-1)}, \tag{24}$$

which implies our next result.

<u>Proposition 3</u>: *With a positive capital tax, underprovision of business public services occurs if at* $F_B > 1$, $\varepsilon_{F_B,K} < 1$. *Overprovision of business public services occurs if at* $F_B < 1$, $\varepsilon_{F_B,K} > 1$.

Proposition 3 states that if we deal with production functions that exhibit $\varepsilon_{F_B,K} < 1$ for any *B* and *K*, we will always have underprovision of public services under a capital tax.[10]

Note that (24) cannot be used to assess the interior solution if $\varepsilon_{F_B,K} = 1$. In order to examine this case, consider again the perceived outflow of capital after taking into account budget balance, following the approach we utilized for the production tax in section 3, or

$$\frac{dK}{d\tau} = \frac{1-F_{KB}K}{F_{KK}+\tau F_{KB}} = \frac{F_B(1-\varepsilon_{F_B,K})}{F_{KK}+\tau F_{KB}}. \tag{25}$$

Local jurisdictions do not perceive an outflow of capital if $\varepsilon_{F_B,K} = 1$. This implies

<u>Proposition 4</u>: *With a positive capital tax, efficient provision of business public services occurs if and only if at* $F_B = 1$, $\varepsilon_{F_B,K} = 1$.

Propositions 3 and 4 relate to the existing literature on the efficient provision of business public services with a capital tax. Matsumoto (1998, p. 471) notes: "If the number of firms is constant in each jurisdiction (normalized to one), […] $[F_B]$ may be below one because the sign of $[F_{KB}K - F_B]$ is indeterminate under linear homogeneity with respect to all inputs. This argument corresponds to the Noiset (1995) result of potential overprovision in the [ZM] model where the variability of the number of firms is not explicitly considered." Dhillon et al. (2007) construct a model with two production inputs (capital and public services only) and then show that

[10] Such is the case for a Cobb-Douglas production function and any CES production function with substitution elasticity greater than 1; see, e.g., Gugl and Zodrow (2015).

overprovision, efficient provision, or underprovision of business public services can occur depending on whether the production function is only locally or globally strictly concave in capital and the public service. In Gugl and Zodrow (2015), we use a CES production function with a substitution elasticity less than one to provide numerical examples in which the capital tax is efficient.

5. Efficiency Properties of the Production and Capital Taxes

The various equilibrium outcomes under the production tax and capital tax regimes can be summarized as follows:

Table 1: Comparison of Production Tax and Capital Tax Equilibria

	efficient provision $\boldsymbol{F_B = 1}$	**underprovision $\boldsymbol{F_B > 1}$**	**overprovision $\boldsymbol{F_B < 1}$**
production tax	$\varepsilon_{F_B,K} = \varepsilon_K$	$\varepsilon_{F_B,K} < \varepsilon_K$	$\varepsilon_{F_B,K} > \varepsilon_K$
capital tax	$\varepsilon_{F_B,K} = 1$	$\varepsilon_{F_B,K} < 1$	$\varepsilon_{F_B,K} > 1$

We turn next to an evaluation of the relative efficiency properties of the two taxes.

Corollaries 1 and 2 identify the properties of the production function that guarantee efficient provision, overprovision, or underprovision under the production tax. We first consider production functions for which log modularity (with $\varepsilon_{F_B,K} = \varepsilon_K$) or log submodularity (with $\varepsilon_{F_B,K} < \varepsilon_K$) holds and investigate the relative efficiency properties of the production tax and the capital tax under the same technologies. In order to do this, it is important to note how these properties relate to the question of whether $\varepsilon_{F_B,K} < 1$.

<u>Proposition 5</u>: *If* $F(B,K,L)$ *is log submodular or log modular and strictly concave in* (B,K)*, then* $\varepsilon_{F_B,K} < 1$.

Proof: For a strictly concave function in K, $F_K K < F$ and hence $\varepsilon_K < 1$. By Lemmas 1 and 2, a log (sub) modular production function in (B,K) exhibits $\varepsilon_{F_B,K} \leq \varepsilon_K$. Together these two properties imply the stated result.

Note, however, that even if $\varepsilon_K < 1$ holds, there is no restriction on the log modularity property of $F(B,K,L)$ in (B,K).

Proposition 5 in turn implies

<u>Proposition 6</u>: *If* $F(B,K,L)$ *is log modular, then the production tax leads to efficient provision while the capital tax leads to underprovision of business public services; hence the production tax is more efficient than the capital tax.*

By comparison, in the case of log submodular production functions, underprovision occurs under both tax regimes. Nevertheless, it is possible to rank the efficiency properties of the two taxes, as shown in

<u>Proposition 7</u>: *Suppose that the production function is log submodular in* (B,K) *and that interior solutions to both the optimal capital tax and the optimal production tax problems exist. Under these conditions, the production tax is unambiguously more efficient than the capital tax.*

Proof: We first provide an outline of the proof. Suppose that under the capital tax each jurisdiction chooses the tax rate optimally and the economy is in equilibrium so that $B^{\tau} = \tau\overline{K}$. Total output is the same in all jurisdictions (given symmetry and a fixed national aggregate capital stock) and thus all derivatives of the production function with respect to *B* and *K* are also identical across jurisdictions. Using the optimality condition for the capital tax (24) we can calculate the production tax equivalent that would result at the same level of business public services as under the optimal capital tax equilibrium if all jurisdictions assumed that capital reacts to their provision of public services and to a tax increase exactly as under the equilibrium capital tax. We then show that the optimality condition for the production tax (15) does not hold simultaneously at this production tax rate. In particular, the level of the production tax given the calculated elasticities at B^{τ} and $\overline{K}$ is higher, which implies a larger *B*; this in turn implies that, for a given increase in business public services *B*, the jurisdiction perceives a smaller capital outflow under a production tax than under the capital tax, which implies a higher level of *B* under the optimal production tax than under the optimal capital tax. Since both tax regimes lead to underprovision, the production tax must thus result in less underprovision than the capital tax.

More formally, the production tax equivalent of the capital tax rate under the optimal capital tax equilibrium is given by

$$t^{\tau} = \frac{B^{\tau}}{F(B^{\tau},\overline{K},L)} = \frac{F_K^{\tau}(1-F_B^{\tau})\varepsilon_{F_K,K}^{\tau}\overline{K}}{F_B^{\tau}\left(\varepsilon_{F_B,K}^{\tau}-1\right)F^{\tau}} \,. \qquad (26)$$

The optimal production tax, however, at the elasticities evaluated at B^{τ} and $\overline{K}$ is

$$t\left(B^{\tau},\overline{K}\right) = \frac{(1-F_B^{\tau})\varepsilon_{F_K,K}^{\tau}}{F_B^{\tau}\left(\varepsilon_{F_B,K}^{\tau}-\varepsilon_K^{\tau}\right)}. \qquad (27)$$

Thus, if $t^{\tau} < t\left(B^{\tau},\overline{K}\right)$, then the capital tax results in more underprovision of business public services than the production tax and hence is less efficient. Note that

$$\frac{F_K^{\tau}(1-F_B^{\tau})\varepsilon_{F_K,K}^{\tau}\overline{K}}{F_B^{\tau}\left(\varepsilon_{F_B,K}^{\tau}-1\right)F^{\tau}} < \frac{(1-F_B^{\tau})\varepsilon_{F_K,K}^{\tau}}{F_B^{\tau}\left(\varepsilon_{F_B,K}^{\tau}-\varepsilon_K^{\tau}\right)} \qquad (28)$$

if and only if

$$-\frac{1}{\varepsilon_K^{\tau}}\left(\varepsilon_{F_B,K}^{\tau}-1\right) > -\left(\varepsilon_{F_B,K}^{\tau}-\varepsilon_K^{\tau}\right), \qquad (29)$$

which is unambiguously true, as both sides are positive and strict concavity of the production function in *K* implies that the left hand side is greater than the right hand side.

Our results in the cases of log modular and log submodular production functions validate the conjecture that a broad-based tax on local production is more efficient than a tax on capital. However, there is no analog to Proposition 7 in the case of log supermodular production functions, as either the production tax or the capital tax can be the more efficient business tax option. Moreover, there are prominent examples of such functions. As noted previously, the family of CES production functions includes examples of all three log modularity properties — log submodularity, log modularity, and log supermodularity — depending on whether the elasticity of substitution between capital and public services/labor is greater than, equal to, or less than one. In Gugl and Zodrow (2015), we examine more closely CES production functions characterized by log supermodularity in the context of tax competition in the provision of business public services. We show that both the production tax and the capital tax can be the more efficient tax option, but in the latter case, the efficiency gain from utilizing the capital tax instead of the production tax is relatively small. By comparison, relatively large efficiency gains can sometimes be obtained by using the production tax rather than the capital tax in those cases in which the production tax is the more efficient tax option.

6. Conclusion

Can a production tax, such as an origin-based value-added tax, approximate a benefit tax for public services provided to businesses, as suggested by Bird (2000) and Hines (2003)? How does a source-based capital tax such as the property tax compare to a production tax as a proxy for a benefit tax? Using the ZM model of interjurisdictional tax competition, we find that a production tax more closely approximates a benefit tax than does a capital tax in many instances. In particular, although a production tax is efficient only when the production function is log modular in business public services and capital, it is always *less* inefficient than a capital tax in the case of log submodular production technologies.

The basic intuition underlying the result that the production tax is often more efficient than the capital tax is that the capital tax applies to a narrower and relatively mobile tax base, while the production tax applies to a broader base that includes both relatively immobile labor and local business public services. However, if there is very strong complementarity between capital and business public services (as assumed by Sinn 1997 and Oates and Schwab 1988, 1991), then the capital tax is more likely to approximate a benefit tax than when capital and business public services are more substitutable. By comparison, in the presence of significant substitution opportunities between capital and business public services, a broader tax base that includes both immobile labor and the contribution of public services to output mimics more closely efficient user charges on public services.[11]

Our results have some interesting implications for potential reforms of state/provincial tax systems, as they suggest that a production tax may be a viable business tax alternative to the state/provincial retail sales tax, which is typically characterized by significant taxation of business inputs similar to that which occurs under a capital tax. By comparison, at least under certain circumstances in our admittedly highly stylized model, a production tax is less distortionary than the capital tax portion of a retail sales tax.

---

[11] In an earlier version of this paper (Gugl and Zodrow, CESifo Working Paper No. 5555, available at https://www.cesifo-group.de/DocDL/cesifo1_wp5555.pdf), we also compared the production tax to a uniform tax on only private inputs, under the assumption of Cobb-Douglas production functions. The latter tax is efficient in Matsumoto and Sugahara (2017) where the production function is CRS in private inputs only, but we show that it is inefficient in the case of many Cobb-Douglas production functions if the function is CRS in all inputs including public services even though such technology implies that the production tax is efficient.

References

Bird, R. M. (2000). Subnational VATs: Experience and prospects. *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association,* 93, 223-228.

Bird, R. M. (2003). A new look at business taxes. *Tax Notes International,* 30(7), 695-711.

Bird, R. M., & Gendron, P-P. (2007) *The VAT in developing and transitional countries*. Cambridge, UK: Cambridge University Press.

Bayindir-Upmann, T. (1998). Two games of interjurisdictional competition when local governments provide industrial public goods. *International Tax and Public Finance,* 5, 471–487.

CBC News (2011). B.C. promises HST cut to 10% by 2014: Tax hike for corporations and cheques for families also promised. http://www.cbc.ca/news/canada/britishcolumbia/story/2011/05/25/bc-hst-changes.html. Accessed 15 August 2018.

Crawford, I., Keen, M., & Smith, S. (2010). Value-added tax and excises. In Institute for Fiscal Studies (Ed.), *Dimensions of Tax Design: The Mirrlees Review* (pp 275-362). London, UK: Oxford University Press.

Dhillon, A., Wooders, M., & Zissimos, B. (2007). Tax competition reconsidered. *Journal of Public Economic Theory,* 9(3), 391–423.

Faggiano, M. (2017). Origin-based and destination-based sales tax collection 101. https://blog.taxjar.com/charging-sales-tax-rates/. Accessed 31 August 2018.

Gugl, E., & Zodrow, G. R. (2015). Competition in business taxes and public services: Are production-based taxes superior to capital taxes? *National Tax Journal,* 68(3S), 767–802.

Hall, R., & Rabushka, A. (2007). *The flat tax*. Stanford, CA: Hoover Institution Press.

Haufler, A., & Pflueger, M. (2007). International oligopoly and the taxation of commerce with revenue-constrained governments. *Economica,* 74(295), 451–473.

Hines, J. R. Jr. (2003). Michigan's Flirtation with the Single Business Tax. In C. L. Ballard, P.N. Courant, D. C. Drake, R. C. Fischer, & E. R. Gerber (Eds.), *Michigan at the Millennium: A Benchmark and analysis of its fiscal and economic structure* (pp 603-28). East Lansing: Michigan State University Press.

Kanbur, R., & Keen, M. (1993). Jeux sans frontiers: Tax competition and tax coordination when countries differ in size. *American Economic Review,* 83(4), 877–892.

Keen, M., & Marchand, M. (1997). Fiscal competition and the pattern of public spending. *Journal of Public Economics,* 66(1), 33–53.

Lockwood, B. (2001). Tax competition and tax co-ordination under destination and origin principles: A synthesis. *Journal of Public Economics,* 81(2), 279–319.

Matsumoto, M. (1998). A note on tax competition and public input provision. *Regional Science and Urban Economics,* 28(4), 465–473.

Matsumoto, M. (2000). A tax competition analysis of congestible public inputs. *Journal of Urban Economics,* 48(2), 242–259.

Matsumoto, M. & Sugahara, K. (2017) A note on production taxation and public-input provision. *The Annals of Regional Sciences*, 59(2), 419-426.

McLure, C. E., Jr. (2003). The value added tax on electronic commerce in the European Union. *International Tax and Public Finance,* 10(6), 753–762.

Mintz, J. M., & Tulkens, H. (1986). Commodity tax competition between member states of a federation, equilibrium and efficiency. *Journal of Public Economics,* 29(2), 133–372.

Noiset, L. (1995). Pigou, Tiebout, property taxation, and the underprovision of local public goods: A comment. *Journal of Urban Economics,* 38(3), 312–316.

Oates, W. E., & Schwab, R. M. (1988). Economic competition among jurisdictions: Efficiency enhancing or distortion inducing?" *Journal of Public Economics*, 35(3), 333–354.

Oates, W. E., & Schwab, R. M. (1991). The allocative and distributive implications of local fiscal competition. In D. A. Kenyon & J. Kincaid (Eds.), *Competition among states and local governments: Efficiency and equity in American federalism* (127-146). Washington, DC: The Urban Institute Press.

Richter, W. F. (1994). The efficient allocation of local public factors in Tiebout's tradition. *Regional Science and Urban Economics*, 24(3), 323–340.

Sinn, H-W. (1997). The selection principle and market failure in systems of competition. *Journal of Public Economics,* 66(2), 247–274.

The Independent Panel (2011). HST or PST/GST: It's your decision. http://www.slideshare.net/hstinbc/its-your-decisionhstgstpst. Accessed 15 August 15 2018.

Wilson, J. D. (1986). A theory of interregional tax competition. *Journal of Urban Economics,* 19(3), 296–315.

Wilson, J. D. (1999). Theories of tax competition. *National Tax Journal,* 52(2), 269–304.

Wilson, J. D., & Wildasin, D. E. (2004). Tax competition: Bane or boon?" *Journal of Public Economics,* 88(6), 1065–1091.

Zodrow, G. R. (2003). Tax competition and tax coordination in the European Union. *International Tax and Public Finance,* 10, 651–671.

Zodrow, G. R. (2010). Capital mobility and capital tax competition. *National Tax Journal,* 63(4), 865–901.

Zodrow, G. R., & Mieszkowski, P. (1986). Pigou, Tiebout, property taxation and the underprovision of local public goods. *Journal of Urban Economics,* 19(3), 356–370.

Appendix: Log (Sub/Super) Modularity

Before discussing log (sub/super) modularity, we first review (sub/super) modularity and then point out the differences between these two related concepts.

A.1 Definition and Examples

To define (sub/super) modularity of *B* and *K*, consider any two input vectors with the same amount of labor but different amounts of *B* and *K*, with one input vector containing strictly less of both $(\underline{B}, \underline{K}, L)$ than the other $(\overline{B}, \overline{K}, L)$. Since we keep labor the same in all our analysis, L is suppressed in our analysis. By definition, modularity holds if and only if

$$F(\underline{B}, \underline{K}) + F(\overline{B}, \overline{K}) = F(\underline{B}, \overline{K}) + F(\overline{B}, \underline{K}) \ . \tag{A1}$$

Under modularity, a given increase in one input results in the same increase of output regardless of how much output is being produced. Another way of thinking about the implications of modularity is by choosing the input combinations on the RHS of equation (A1) such that

$$F(\underline{B}, \overline{K}) = F(\overline{B}, \underline{K}) = Y^0 \ . \tag{A2}$$

With modularity, the arithmetic mean of $F(\underline{B}, \underline{K})$ and $F(\overline{B}, \overline{K})$ must equal Y^0. For super (sub) modularity, the arithmetic mean of $F(\underline{B}, \underline{K})$ and $F(\overline{B}, \overline{K})$ must be greater (less) than Y^0. For example, in the case of submodularity, total output of Y^0 can be produced with the arithmetic mean of $F(\underline{B}, \underline{K})$ and some other output level generated with larger amounts of both inputs, but the output level $F(\overline{B}, \overline{K})$ is not large enough to achieve this; instead, inputs larger than $(\overline{B}, \overline{K})$ are necessary to achieve an arithmetic mean of Y^0.

An example of a production function that is modular in *B* and *K* is one for which the functions involving inputs *B* and *K* enter additively, i.e.,

$$F(B,K,L) = g(B)h(L) + i(K)j(L) + m(B) + o(K) + p(L),$$

where all the individual functions are increasing in their inputs. By comparison, a function of the form

$$F(B,K,L) = g(B)h(L)i(K)j(L) + m(B) + o(K) + p(L)$$

is supermodular.

Log modularity, in contrast, compares relative marginal products rather than the absolute marginal products of inputs. That is, log modularity in *B* and *K* holds if and only if

$$F(\underline{B},\underline{K})F(\overline{B},\overline{K}) = F(\underline{B},\overline{K})F(\overline{B},\underline{K}) \ . \qquad \text{(A3)}$$

As in the discussion of (sub/super) modularity, another way to think about the implications of log modularity is by choosing the input combinations on the RHS of equation (A3) so that (A2) holds. For log modularity, the *geometric* mean of $F(\underline{B},\underline{K})$ and $F(\overline{B},\overline{K})$ must equal Y^0. For log super (sub) modularity, the geometric mean of $F(\underline{B},\underline{K})$ and $F(\overline{B},\overline{K})$ must be greater (smaller) than Y^0. For example, in the case of log submodularity, total output of Y^0 can be produced with the geometric mean of $F(\underline{B},\underline{K})$ and some other output level generated with higher amounts of both inputs, but the output level $F(\overline{B},\overline{K})$ is not large enough to achieve this; instead, inputs larger than $(\overline{B},\overline{K})$ are necessary to achieve a geometric mean of Y^0.

An example of a production function that is log modular in *B* and *K* is one for which the inputs enter multiplicatively, e.g.,

$$F(B,K,L) = g(B)h(K)i(L),$$

where all the individual functions are increasing in their inputs. Note that such functions are simultaneously supermodular in *B* and *K*. An example of a log supermodular function is

$$F(B,K,L) = g(B)h(K)i(L) + j(L),$$

which is simultaneously super modular in *B* and *K*.

An example of a log submodular function is

$$F(B,K,L) = g(B) + h(K) + i(L),$$

which is simultaneously modular in *B* and *K*. However, log submodular functions can also be simultaneously supermodular; an example is

$$F(B,K,L) = [g(B) + h(K) + i(L)]^{\alpha},$$

where $\alpha > 0$.

A.2 *Proof of Lemma 1: Expressing Log (Sub/Super) Modularity in terms of elasticities*

Note that log modularity can be expressed as

$$\frac{F(\underline{B},\overline{K})}{F(\underline{B},\underline{K})} = \frac{F(\overline{B},\overline{K})}{F(\overline{B},\underline{K})}. \qquad \text{(A4)}$$

Hence a log modular production function must exhibit the same percentage change in output due to a given increase in *K*, regardless of the level of *B*, or

$$\frac{\partial \varepsilon_K}{\partial B} = 0, \qquad \text{(A5)}$$

where $\varepsilon_K = F_K K/F$, which holds if and only if

$$K\frac{F_{KB}F - F_B F_K}{F^2} = 0. \qquad \text{(A6)}$$

Since by Young's Theorem $F_{KB} = F_{BK}$, and $\varepsilon_{F_B,K} = F_{BK}K/F_B$, (A6) can be expressed as

$$\varepsilon_{F_B,K} = \varepsilon_K. \qquad \text{(A7)}$$

Similarly, log submodularity requires $\varepsilon_{F_B,K} < \varepsilon_K$ and log supermodularity requires $\varepsilon_{F_B,K} > \varepsilon_K$.