CESifo WORKING PAPERS

9329
2021

September 2021

Second-Best Source-Based Taxation of Multinational Firms

Johannes Becker

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9329*

Second-Best Source-Based Taxation of Multinational Firms

Abstract

I consider a continuum of multinational enterprises (MNEs), which differ in profitability. MNEs employ capital, shift profit to tax havens and may relocate their production facilities to other countries. Source countries provide public inputs and levy taxes. I derive optimal policy choices for different government objectives (to maximize tax revenue, national income or the representative household's utility) allowing for an unrestricted set of tax policy instruments — in contrast to most existing work on corporate taxation. With observable productivity types, source governments set type-dependent lump-sum taxes and attain the first-best allocation. With unobservable productivity types, the optimum source-based tax system consists of a small lump-sum tax (driving low-profit types out of the market) and positive marginal taxes on reported profit. Optimal marginal tax rates on capital inputs are positive if more profitable firms employ more capital. Optimal public inputs are lower than in the first best if they are of higher value to more profitable firm types. I use a sufficient statistics approach (following Saez 2001) to express optimal tax and input choices as functions of elasticities of observable choice variables. Finally, I use the model to evaluate tax policy measures, e.g. the introduction of an effective minimum tax on profits in tax havens, and to derive the welfare properties of tax competition with an unrestricted set of tax instruments.

JEL-Codes: H250, H710, F230.

Keywords: corporate taxation, multinational firms, optimum taxation, tax competition.

Johannes Becker
University of Münster / Germany
johannes.becker@wiwi.uni-muenster.de

September 28, 2021
We thank Manon Francois, Eckhard Janeba, Lars Persson, Kurt Schmidheiny, Daniel Waldenström and participants at research seminars in Basel and Stockholm and the Venice Summer Institute 2019 for helpful comments. The usual disclaimer applies.

1 Introduction

With the ongoing multilateral efforts to reform the international tax system and plans for major unilateral tax reforms in the US and Europe, there is a renewed interest in the question how to optimally tax multinational enterprises (MNEs). More precisely, what is the best way of taxing an entity that is part of a multinational firm which uses internationally mobile production factors, has some leeway in reporting profits and may, if necessary, migrate to other locations? Turning to economic theory for guidance, there is surprisingly little work that derives optimum taxes on MNEs from an unrestricted set of tax instruments. To be sure, the literature on international taxation is vast, but the largest part of it starts from a limited set of tax instruments (e.g. a linear tax on profit and some allowance proportional to corporate capital).

This paper sets out to fill this gap. It offers a model with a continuum of MNEs which differ in profitability. Each MNE consists of a production entity and another entity which holds the firm's IP and is located in a tax haven. The production entity uses mobile capital to produce an output good and it may shift profit to tax havens as well as relocate to other locations ('countries'). The government of the location that hosts the production entity (the 'source country') levies taxes and provides public inputs which increase the MNE's profit. The model is used to derive optimal tax rates and optimal public inputs for different government objectives (tax revenue, national income, domestic household utility). It is augmented to allow for type-dependent profit shifting opportunities and foreign firm ownership. Using the sufficient statistics approach by Saez (2001), I express the optimal marginal tax rates on profit and capital inputs in terms of elasticities based on observable variables.

The results are the following. With observable productivity types, the optimum source-based tax is a type-dependent lump-sum tax that accounts for the firm's outside option, but leaves capital inputs undistorted and does not trigger profit shifting to tax havens. With unobservable productivity types, the government optimally uses the firm's observable features, i.e. capital inputs and reported profits, to solve the screening problem. Then, the optimal tax system consists of a uniform lump-sum tax, which distorts market entry, and strictly positive marginal tax rates on reported profit, which induce profit shifting to tax havens. Optimal marginal taxes on capital inputs are positive (negative) if capital inputs are complements (substitutes) with profitability. The level of public input provision is distorted downwards (upwards), if more profitable firms benefit more (less) from public inputs than less profitable ones.

The model may be used to rationalize tax structures observed in real-world tax systems and it may serve as a tool to evaluate policy measures. To start with, the model demonstrates that, in many cases, capital input distortions are not unnecessary, as is sometimes suggested (e.g. by proponents of cash-flow taxes or ACE regimes[1]), and may not be eliminated without cost. There are special cases, though, in which capital distortions are optimally chosen to be zero (and only 'pure profit' is taxed). In general, however, the distortion of reported profit and capital inputs is inevitable (for similar reasons that the distortion of labour supply is unavoidable in the Mirrlees (1971) model) and even taxes on non-profitable firms (like the Alternative Corporate Minimum Tax in the US) can be rationalized. Note that, in contrast to the Mirrleesian optimum income tax, these optimal distortions are not resulting from an efficiency-equity tradeoff; they are rather driven by the desire to equate the marginal efficiency cost of taxation across decision margins.[2] Moreover, the model provides the optimal tax structures for specific settings in the economy. Referring to a frequently used description of the modern economy: If the economy consists of low-profitability 'brick-and-mortar' firms, which heavily rely on capital and public inputs ('infrastructure'), and high-profitability 'digital' firms, which are mostly independent of public inputs, the model shows that the optimal tax system subsidizes capital use and optimal public inputs are higher than the efficient level. In addition, policy measures like e.g. a crack-down on tax havens or the introduction of an effective minimum tax can be evaluated. As it turns out, it depends on the technological features of the concealment cost function whether an effective minimum tax increases or decreases the optimal tax rates on reported profit and mitigates or aggravates the distortions of capital inputs and public input provision.

While source-based taxation is a typical feature of the current system of international business taxation, it deserves some justification, since there are strong theoretical arguments against it. The production efficiency theorem (Diamond/Mirrlees 1971) states that source taxation with locally differentiating tax rates is suboptimal. Moreover, there is the general presumption that cap-

[1] Both, cash-flow tax systems and ACE systems (for 'allowance for corporate equity') are intended to reduce the tax cost of (equity) capital to zero. This paper shows that a tax cost of capital of zero may not be optimal.

[2] To be specific, with firms differing in profitability, uniform lump-sum taxes may deter market entry and, thus, have an efficiency cost. Therefore, positive marginal tax rates on reported profit and, potentially, capital inputs are optimal.

ital should not be taxed at all (Atkinson/Stiglitz 1976), at least not if pure profits can be fully taxed.[3] Gordon (1986) adds to this that a source tax on mobile capital inputs is borne by the immobile factors anyway and is, therefore, welfare dominated by directly taxing the latter.[4]

The arguably most important justification of source-based taxes is based on Tiebout (1956) who argued that, under interjurisdictional competition, taxes will reflect the utility from public goods provision. Richter/Wellisch (1996) translate this framework to a setting where jurisdictions compete for households and firms (benefitting from public inputs[5]) and demonstrate that the resulting equilibrium may equally be efficient.[6] This 'benefit principle' is explicitly mentioned in the reports underlying the 1920s decision in the League of Nations:[7] Profits shall be taxed where government provide public inputs, i.e. at source where firms benefit from local infrastructure, educated workers, the legal system and so on.[8] In fact, it may be argued that public inputs provided at source make source-based taxes indispensable – at least if contractual arrangements on public input provision between countries remain absent or imperfect.[9]

The property tax literature in the tradition of Tiebout (1956) gave rise to the classical tax competition literature (Zodrow/Mieszkowski 1986, Wilson 1986) which started with the assumption that lump-sum taxes ('head taxes') are not available and which, instead, considered unit taxes on capital inputs. Later work shifted the focus from taxing property to taxing firms and from taxing capital inputs to taxing profit. Sinn (1990) points out that taxing pure profit at source by making capital inputs fully deductible immunizes countries against tax com-

[3] The other seminal argument against capital income taxation, based on Chamley (1986) and Judd (1985) originally refers to a closed economy setting, but extends to a multi-country open economy framework (Gross et al. forthcoming).

[4] Later work has shown that, if pure profits cannot be perfectly identified, a marginal tax on capital inputs may be justified (e.g. Keen/Piekkola 1997).

[5] Productive public inputs are analyzed by Zodrow/Mieszkowski (1986), Bayindir-Upmann (1998), Matsumoto (1998, 2000), Matsumoto/Sugahara (2017).

[6] The idea that public inputs create location-specific rents for the firm, which can be taxed even under competition, also plays a role in the bidding-for-firms literature (Black/Hoyt 1989) and the new economic geography (Baldwin/Krugman 2004). Under perfect competition among locations, countries are only able to tax the location-specific rent.

[7] The committee that recommended taxing rights for source countries argued that "foreigners, whose activities reach some minimum threshold, should contribute to the costs of services provided by the host government" (Graetz 2001, p. 1396).

[8] In addition, there is the trivial fact that source countries may just feel entitled to tax the income that has been generated on their territory. "The claim of source countries to tax income produced within their borders is analogous to a nation's long-recognized claim of sovereignty over natural resources within its boundaries." (Graetz 2001, p. 1396).

[9] See Keen/Marchand (1997) for a model with government spending on public goods and public inputs under fiscal competition.

petition. In the mid-1990s, the focus shifted again from capital mobility to firm mobility (Devereux/Griffith 1998) and to profit shifting (Hines/Rice 1994), both making Sinn's (1990) immunization strategy impossible. For both decision margins, it has been shown in reduced form models (which assume, among others, proportional tax rates on corporate profits) that constrained optimality requires a tax on capital inputs (see e.g. Keen/Piekkola 1997, Haufler/Schjelderup 2000, Becker/Fuest 2011). There is no justification in any of these models, though, of why more sophisticated tax instruments, e.g. lump-sum taxes with non-linear taxes on profit and capital inputs, are not used despite being superior in efficiency terms. The Keen/Konrad (2013) handbook article on tax competition and coordination almost exclusively cites contributions that consider a constrained set of tax instruments.[10]

This 'constrained' approach has obvious advantages like tractability and capturing typical features of real-world corporate tax systems (which often have uniform statutory tax rates in large parts of the tax schedule). However, it would be helpful to know what a source based tax on multinational firms is able to achieve if it were designed from scratch without any restrictions on the choice of instruments. Put differently, what is missing so far, is a Mirrleesian (1971) approach to business taxation. Mirrlees (1971) showed that, in a second-best environment, a tradeoff between equity (more redistribution) and efficiency (less distorted labour supply) cannot be avoided; and that optimum taxation is actually income taxation (since income signals skill). I argue in this paper that a similar approach to international business taxation is worthwhile since it reveals the deep sources of inefficiency inherent to the international tax system. In an international business tax setting, the trade-off is not between equity and efficiency, but rather between raising revenue to finance public input goods on the one hand, and minimizing distortions of business decisions (e.g. investment) on the other hand. Such a more comprehensive approach does not, to be sure, overthrow the existing literature; in many ways, it confirms and generalizes propositions derived in settings with constrained sets of tax instruments. There are some novel insights, though, including the finding that a lump-sum tax on non-profitable firms is part of the optimum tax system and the derivation of sufficient conditions for taxing capital at source.

The distortions inherent to the second best allocation are due to informa-

[10] In contrast, the survey article by Gresik (1998) devotes a whole section to information asymmetries and discusses a number of studies that use mechanism design theory in an international tax context. I will discuss these studies and how they relate to this study below.

tion asymmetries. A number of papers (some of them surveyed by Gresik 1998) considers information asymmetries in corporate tax competition. In an early contribution, Gresik/Nelson (1994) consider the regulation of an MNE which has private information on its cost and needs to apply a transfer price for tax purposes. Osmundsen et al. (1998) analyze a setting in which the MNE's mobility is private information; these authors show that the capital inputs of (relatively) immobile firms is subsidized in the second-best allocation.[11] Becker/Schneider (2019) consider countries bidding for large firms when the firms' productivity is unknown and show that the second-best bidding equilibrium implies inefficiently high employment targets. These contributions have in common that there is a single government (the principal) dealing with the firm (the agent) or, in a setting with mobile firms, many principals competing for the exclusive service of an agent (see, e.g., Biglaiser/Mezzetti 1993, 2000).

An alternative setting, that is specifically relevant for multinational firms, is where where multiple principals (here: countries/governments) deal with the agent (here: the MNE).[12] Bond/Gresik (1996) apply this approach to the transfer pricing regulation problem within MNEs. Olsen/Osmundsen (2001, 2003) present a two-region model where the firm is simultaneously affected by the policy choices in these regions. They analyze tax competition for capital allocated within the firm when the firm has private information on its productivity. The screening equilibrium implies a downward distortion of investment.

The remainder of the paper is organized as follows. In the next section, I set out the model and derive the optimum tax for immobile firms. Section 3 augments the model to allow for firm mobility. Section 4 discusses the results and concludes.

2 Model with immobile firms

The model analysis proceeds in two steps. In this section, I assume that firms have a fixed (i.e. type-independent) outside option. This includes the case of immobility, in which the outside option is zero, i.e. the participation constraint for all types is to have non-negative profit. In Sect. 3, I introduce heterogeneous firm mobility.

[11] Becker/Schneider (2017) consider tax policy with unknown firm mobility in an intertemporal setting.

[12] Both of these settings (multiple principals competing for the exclusive services of one agent and an agent serving multiple principals simultaneously) are covered in the theory of common agency (Bernheim/Whinston 1986, Martimort/Stole 2002).

2.1 Setup

Consider a small open economy hosting production entities of multinational enterprises (MNEs). Each MNE consists of a production entity and another entity. The latter is located in a tax haven, i.e. a jurisdiction with low or zero tax rates, and holds the firm's IP.[13] I assume that the IP is the result of previous R&D effort and fixed in size and economic value. For now, I assume that intra-firm dividends are tax exempt, such that the question which entity is the headquarters of the firm is not relevant. I also ignore decisions on the sources of finance (debt vs. equity) and the level of R&D. Importantly, I assume that the firm cannot be split up or costlessly merged with other firms.[14]

I assume that there is a continuum of MNEs with a mass of one. MNEs differ in a parameter $\theta > 0$ which is the firm's private information.[15] θ is distributed on the interval $\left[\underline{\theta}, \bar{\theta}\right]$ with a c.d.f. of $F(\theta)$ and an associated density $f(\theta) = F'(\theta)$.

Each production entity employs capital k to produce an output good. It benefits from a public input good g provided by the source country. The entity's profit function is given by $\pi(k, g, \theta)$. Profits increase in k, g, and θ. Furthermore, I assume that $\pi_{kk} < 0$ and that the second derivatives are symmetric, i.e. $\pi_{kg} = \pi_{gk}$ etc. With $\pi_\theta > 0$, I will refer to θ as a *profitability* parameter. θ may be interpreted as the firm's intellectual property (IP), some other fixed factor (like a productive firm culture, skilled managers etc.), access to cheap financing sources etc.

The below analysis will show that optimal taxes crucially depend on how the profitability parameter interacts with capital inputs and public inputs. Capital inputs k and θ will be called *complements* if $\pi_{k\theta} > 0$ and *substitutes* if $\pi_{k\theta} < 0$. Then, in the absence of tax distortions, the k chosen by the firm increases or decreases in θ, respectively. Similarly, public inputs g and θ will be called *complements* (or *substitutes*) if $\pi_{g\theta} > 0$ (or $\pi_{g\theta} < 0$), respectively. Then, firms with high θ benefit more (or less) from public inputs than firms with low θ.

The source location provides g at a cost of $c(g, n)$ where n is the number of production entities. I assume that $c_g(g, n) > 0$ and $c_n(g, n) \geq 0$. If $c_n > 0$, there is a congestion cost of public input provision, i.e. the cost depends on the

[13] The two-entity MNE is the simplest MNE structure possible. However, as long as other entities in the firm do not interact with the production entity, the results below are not affected.

[14] Allowing for split ups and mergers would impose a new non-trivial constraint into the optimal tax problem.

[15] An example considered in more detail below is $\pi(k, g, \theta) = \theta k^\psi g - k$ with $\psi \in (0, 1)$.

number of firms using the public input.

In addition, the source country government levies taxes T that may be conditioned on observable firm characteristics. The tax levied by the haven government is denoted as T^h. For tax purposes, the firm has to report a profit to the tax authorities in the source location and the tax haven. It declares a profit $\hat{\pi}$ in the source country and a profit $\hat{\pi}^h$ in the tax haven, with the restriction that $\hat{\pi} + \hat{\pi}^h \geq \pi$ – thus, there is no opportunity for tax evasion. Declaring a profit $\hat{\pi} < \pi$ comes at a resource cost of concealment that is strictly convex in $\Delta = \pi - \hat{\pi}$ and equal to zero if $\Delta = 0$. In other words, 'true' locational profits is π in the production entity and 0 in the IP entity. This assumption is made for simplicity, none of the below results depend on the level of 'true' profit.

For the sake of brevity, I denote as $\kappa(\Delta)$ the total cost of reporting profit abroad, i.e. the sum of the resource cost of concealment, denoted by $\tilde{\kappa}(\Delta)$ and the haven's tax $T^h(\Delta)$, i.e. $\kappa(\Delta) = \tilde{\kappa}(\Delta) + T^h(\Delta)$. I assume that $\kappa(\Delta)$ is strictly convex in Δ and that $\kappa(\Delta) = 0$.[16]

After-tax profits $\Pi(\theta)$ are thus

$$\Pi(\theta) = \pi(k, g, \theta) - \kappa(\Delta) - T(\pi - \Delta, k) \tag{1}$$

The firm chooses k and Δ in order to maximize $\Pi(\theta)$. The first order conditions are

$$\pi_k = \tilde{T}_k(\theta) \quad \text{and} \quad \kappa' = T_{\hat{\pi}}(\theta) \tag{2}$$

where $T_{\hat{\pi}}(\theta)$ and $\tilde{T}_k(\theta)$ is short for $T_{\hat{\pi}}(\pi(k, g, \theta) - \Delta(\theta), k(\theta))$ and $\tilde{T}_k(\pi(k, g, \theta) - \Delta(\theta), k(\theta))$, respectively, and $\tilde{T}_k \equiv \frac{T_k}{1 - T_{\hat{\pi}}}$.

I can now define the Pareto-efficient or first-best allocation. For this purpose, I assume that the marginal value of public funds equals 1, i.e. taxation and public inputs are used to maximize income. At a later stage, I will allow for a higher marginal value of public funds. Let Θ denote the set of all types that enter the market and $n = n(\Theta)$ the number of firms in that set. Assuming that the outside option, $\Pi^o \geq 0$, does not generate tax revenue or any other externalities, the first-best allocation solves

$$\max_{k, \Delta, g, \Theta} \left[\int_{\theta \notin \Theta} \Pi^o f(\theta) \, d\theta + \int_{\theta \in \Theta} (\pi(k, g, \theta) - \kappa(\Delta)) f(\theta) \, d\theta - c(g, n) \right] \tag{3}$$

Definition 1 *The first-best allocation* $(k^*, \Delta^*, \Theta^*, g^*)$ *is defined by:*

[16] Strict convexity requires that $T^{h\prime\prime}$ is not strongly negative.

(i) undistorted capital demand: $\pi_k^\theta = 0$ *for all types* $\theta \in \Theta$.
(ii) truthful profit reporting: $\hat{\pi}^* = \pi$ *and, thus,* $\hat{\pi}^{h*} = \Delta^* = 0$ *for all types* $\theta \in \Theta$.
(iii) $n^* = n(\Theta^*)$ *being such that all types* $\theta \in \Theta^*$ *have a surplus of at least the marginal cost of public input provision* $\pi(k, g, \theta) \geq c_n(g, n^*)$.
(iv) efficient public input provision $(g = g^*)$: $\int_{\theta \in \Theta} \pi_g f(\theta)\, d\theta = c_g(g^*, n^*)$

With respect to timing, all policy parameter (i.e. the tax system and the level of public input provision) are set before the firm decides on (i) staying in the market or exit the market, (ii) the level of reported profit, (iii) capital inputs. I start with considering as the government's objective function the goal of maximizing tax revenue net of cost of public input provision:

$$\int_{\theta \in \Theta} T(\theta) f(\theta)\, d\theta - c(g, n) \tag{4}$$

Below, I consider the alternative goals of maximizing national income (taking into account the domestic ownership share of firms) as well as maximizing the representative household's utility (which can be expressed as a weighted average of tax revenue and national income).

2.2 Source taxation without profit shifting

As a benchmark, assume that $\kappa'(\Delta)$ is prohibitively high for all Δ such that profit shifting is ruled out. Thus, $\hat{\pi} = \pi^\theta$ and $\hat{\pi}^h = 0$.[17] The government has access to uniform lump-sum taxes and non-linear taxes conditioning on the two observable features, profit and capital, i.e. $T = T(k, \pi)$. In this case, the following Proposition holds.

Proposition 1 (no profit shifting) *In the absence of profit shifting, the revenue-maximizing source tax system is profit tax based, investment neutral* ($T_k = 0$) *and has a lump-sum element if* $c_n > \Pi^o$. *The resulting allocation is the first-best allocation (see Def. 1).*

A formal proof is straightforward and is omitted here. Instead, I provide some intuition. Without profit shifting, reported profit equals true profit. Marginal taxes on profit with zero marginal taxes on capital inputs, have no ef-

[17] The firm may decide to not realize its full profit potential, which is not optimal from firm's viewpoint, though, as long as $1 - T_\pi \geq 0$ in the relevant range. It is straightforward to show that the government has no incentive to set $T_\pi > 1$.

ficiency cost. That is, any revenue can be raised that accounts for the participation constraint of the firms, i.e. $\Pi(\theta) \geq \Pi^o$ for all θ. For purpose of understanding, assume that marginal tax rates on profits above Π^o are 100 percent. The source country thus captures the whole (location-specific) surplus and is, thus, in the same position as the benevolent planner who maximizes total income (see Def. 1). In the presence of congestion cost of public input provision, the government ensures that only firms with a surplus above the congestion cost enter the market. This condition is satisfied if $\Pi^o \geq c_n(n^*, g^*)$. If $\Pi^o < c_n(n^*, g^*)$, a lump-sum payment (an 'alternative minimum tax') is levied to drive out firms with surplus below the marginal congestion cost. Note that the first best allocation is also attained if the government maximizes national income (or a weighted average between revenue and national income).

The above Proposition states that source based taxation in the presence of mobile capital and unobservable types may be efficient (and, thus, replicates a well-known feature of international tax theory, see e.g. Sinn 1990). The optimal source tax system is profit tax based, since profit taxes are efficient and superior to uniform lump-sum taxes,[18] and investment-neutral (Gordon 1986). And it contains a lump-sum minimum tax that internalizes the congestion cost of public (input) good provision (Tiebout 1956). Public input provision is efficient under these circumstances.

2.3 Tax revenue maximization

Now, assume that the true profit level π^θ is unobservable to source countries' governments.[19] As a consequence, the source country government can no longer condition its taxes on π^θ. Instead, it has to condition tax payments on the two observable entity features, the declared profit $\hat{\pi}$ and inputs k, i.e. $T = T(\hat{\pi}, k)$.

The source country government maximizes tax revenue net of input provision cost:

$$\max_{T(\hat{\pi},k),g} \int_{\theta\in\Theta} T(\hat{\pi}, k) f(\theta) d\theta - c(g, n) \tag{5}$$

[18] Uniform lump-sum taxes would drive out low productivity types. Profit based taxes are, however, equivalent to individualized lump-sum taxes.

[19] In the simple framework discussed here, this assumption may seem stark since a basic information exchange between the source country and the haven would suffice to know about the true level of profits. However, with a more complex firm structure, where more than just one production entity shifts profits to the haven, information about the true level of profits would be hard to extract. This may provide some intuition for the assumption made here (an alternative would be to allow for two-dimensional heterogeneity, as in Lockwood/Weinzierl (2015) that gives rise to behaviorally indistinguishable types, i.e. to firms that have the same total profit but differ in true local profits π^θ).

After setting $T(\hat{\pi}, k)$ and g, firms choose to enter the market or not, choose their capital input k and the reported profit.

To solve for the optimal tax schedule, I make use of the taxation principle (Rochet 1996).[20] According to the taxation principle, the principal (here: the government) may offer to the agent (here: the firm) a menu of allocations to choose from. An allocation consists of a capital stock k, a level of profit shifting Δ, a tax payment T as well as an entry decision. An allocation is implementable through a non-linear tax schedule if it satisfies the first-order conditions in (2) (*incentive compatibility constraint*) and if the maximized after-tax profit after entering is at least as high as the net income from choosing the outside option (*participation constraint*).

If incentive compatibility holds, i.e. if $k(\theta)$ and $\Delta(\theta)$ maximize the firm's after-tax profit, the after-tax profit increases in the type according to

$$\Pi'(\theta) = \pi_\theta(k, g, \theta) \cdot (1 - \kappa') \tag{6}$$

In order to ensure incentive compatibility, after-tax profits must increase at the rate given in (6) – which may be interpreted as the information rent. This 'law of motion' captures the constraints due to firms choosing capital inputs and reported profit. Equation (6) therefore serves as the *incentive compatibility constraint.*[21] If incentive compatibility is satisfied, the government may directly 'choose' an allocation. This approach is used below.

The fact that the after-tax profit needs to monotonically increase in the type implies that, if at all, only the lower types will choose the outside option. Let $\underline{\theta}^T \geq \underline{\theta}$ denote the lowest type willing to enter the market (instead of choosing the outside option). The number of firms entering the market is thus $n = 1 - F\left(\underline{\theta}^T\right)$.

The government chooses the optimal allocation subject to the law of motion

[20] In the closed economy context, where the firm only interacts with one government, the taxation principle yields the same results as the revelation principle (Rochet 1996). Since the model is later augmented with a migration decision (i.e. the firm interacts with more than one government), the revelation principle does not generally hold (Martimort/Stole 2002). Therefore, I refer to the taxation principle which holds since, even with a migration decision, the firm interacts with only one government at a time. The introduction cites references of work where the firm simultaneously interacts with more than one government.

[21] Note that, while (6) is necessary for the first order conditions in (2) to hold, it is not sufficient. To see this, note that condition (6) can be written as $\Pi'(\theta) = \pi_\theta^\theta(k, g, \theta) \cdot (1 - \kappa') - (\kappa'(\Delta) - T_{\hat{\pi}}) \frac{d\Delta}{d\theta} + (\pi_k(1 - T_{\hat{\pi}}) - T_k) \frac{dk}{d\theta}$. The two latter terms are zero if first order conditions in (2) hold (necessity), but they may also be zero if the two latter terms add up to zero (no sufficiency). This implies that there are potentially more solutions to the optimization problem considered here than those implementable through a non-linear tax schedule.

in (6) and the participation constraints, $\Pi(\theta) \geq \Pi^o$ for all θ:

$$\max_{\underline{\theta}^T, k(\theta), \Delta(\theta), g} \int_{\underline{\theta}^T}^{\bar{\theta}} \left(\pi(k, g, \theta) - \kappa(\Delta) - \Pi(\theta)\right) f(\theta)\, d\theta - c(g, n) \tag{7}$$

As indicated above, the available information constrains the government to a certain tax structure: The tax on an individual type may have a lump-sum element (using the entity as information), a profit-based element (using reported profit as information) and a capital-based element (using the observable stock of capital as information). Tax revenue is thus given by

$$\int T(\theta) f(\theta)\, d\theta = \int \left[\pi\left(\underline{\theta}^T\right) - \Pi^o + \int_{\pi\left(\underline{\theta}^T\right)}^{\hat{\pi}(\theta)} T_{\hat{\pi}}(\theta)\, d\hat{\pi} + \int_{k\left(\underline{\theta}^T\right)}^{k(\theta)} T_k(\theta)\, dk\right] f(\theta)\, d\theta \tag{8}$$

where $\pi\left(\underline{\theta}^T\right) - \Pi^o = T\left(\underline{\theta}^T\right)$ is the lump-sum element.

I can now state the following Proposition:

Proposition 2 *The revenue-maximizing system of source based taxation has the following properties:*
(i) The marginal tax rates on reported profit $\hat{\pi}(\theta)$ and capital input $k(\theta)$ are

$$T_{\hat{\pi}}(\theta) = \frac{1 - F(\theta)}{f(\theta)} \cdot \pi_\theta \cdot \kappa'' \tag{9}$$

$$\tilde{T}_k(\theta) = \frac{1 - F(\theta)}{f(\theta)} \cdot \pi_{\theta k} \cdot (1 - \kappa') \tag{10}$$

for all $\theta \in \left(\underline{\theta}, \bar{\theta}\right)$ and zero for the lowest and the highest types, $\underline{\theta}$ and $\bar{\theta}$.
(ii) The lump-sum tax element equals $T\left(\underline{\theta}^T\right) = \pi\left(\underline{\theta}^T\right) - \Pi^o$ with

$$T\left(\underline{\theta}^T\right) = \pi_\theta\left(k, g, \underline{\theta}^T\right) \frac{1 - F\left(\underline{\theta}^T\right)}{f\left(\underline{\theta}^T\right)} + c_n \tag{11}$$

i.e. even if $c_n = 0$, the lump-sum tax element is strictly positive.
(iii) Public inputs are provided according to

$$\int \pi_g f(\theta)\, d\theta - c_g = \int \frac{1 - F(\theta)}{f(\theta)} \pi_{\theta g} (1 - \kappa') f(\theta)\, d\theta \tag{12}$$

Proof. See the Appendix. ■

Consider first the optimal marginal tax on reported profit. With the assumptions made above, the optimal marginal tax on reported profit is strictly positive for all $\theta \in \left(\underline{\theta}^T, \bar{\theta}\right)$. The optimal tax rate expression has two distinct parts. The first part, $\frac{1-F(\theta)}{f(\theta)}$, is well-known from the optimal tax rate formula in income taxation. It depends entirely on the distribution of types and reflects that an increase of tax at some profit level increases the tax payments for all firms, $1 - F(\theta)$, with a reported profit or capital above that level. This is weighed against the efficiency cost due to taxation at exactly that profit level (with a weight of $f(\theta)$). The second part, $\pi_\theta \kappa''$, reflects how the firm's choice of reported profit affects the incentive compatility constraint (6), whereas κ'' is a measure of how an increase in profit shifted abroad (i.e. a decrease in reported profit) increases the marginal cost of profit shifting and, thus, an inverse measure of how elastically profit shifting responds to tax rate changes.[22]

The optimal marginal tax on capital[23] has a similar structure as $T_{\hat{\pi}}(\theta)$, but its sign is not clearly determined by the assumptions made above. The sign of the tax rate depends on whether θ and k are complements or substitutes. In case of complements ($\pi_{\theta k} > 0$), the ICC is tightened by an increase in capital use, and the optimal tax rate is positive and implies a downward distortion of capital inputs.[24] This is the case for a production function θk^ψ with $\psi \in (0, 1)$ or a profit function of the form $k^\psi - r\frac{k}{\theta}$ (with r being the interest rate). In case of substitutes, the ICC is relaxed by an increase in k and the optimal tax implies a subsidy of capital use. This is the case e.g. in a production function of the form $(k + \theta)^\psi$, where θ may be interpreted as intellectual property (which captures the case in which more IP dependent firms need less tangible capital). In this example, firms with low θ are 'brick-and-mortar' firms, firms with high θ are IP based firms. Finally, if capital inputs are not affected by θ, capital use

[22] Expressed in terms of behavioral elasticities, the optimal tax on reported profit is

$$\frac{T_{\hat{\pi}}(\theta)}{1 - T_{\hat{\pi}}(\theta)} = \frac{1 - F(\theta)}{f(\theta)} \cdot \frac{1}{\theta} \cdot \frac{\varepsilon_{\pi,\theta}}{E(\theta)}$$

where $\varepsilon_{\pi,\theta}$ is the type-elasticity of profit π and $E(\theta) = \varepsilon_{\hat{\pi},1-T_{\hat{\pi}}} + \frac{k}{\pi}\tilde{T}_k \varepsilon_{k,\tilde{T}_k}$ is the weighted sum of the tax rate elasticity of reported profit and the user cost elasticity of capital input. The latter is only relevant, if $\tilde{T}_k \neq 0$.

[23] In terms of behavioral elasticities,

$$\tilde{T}_k(\theta) = \frac{1 - F(\theta)}{f(\theta)} \cdot \frac{\pi}{k} \frac{\varepsilon_{\pi,\theta}\varepsilon_{\pi_\theta,k}}{\theta} \cdot (1 - \kappa')$$

with $\varepsilon_{\pi,\theta}$ defined above and $\varepsilon_{\pi_\theta,k}$ being the elasticity of π_θ with respect to k.

[24] A downward distortion of capital use may be attained by denying full deductibility. The business tax reforms in the 1990s and 2000s that lowered statutory tax rates and broadened the tax base (the so-called tax rate cut cum base broadening type of reforms, see e.g. Becker/Fuest 2011 for a documentation) may, thus, be interpreted as a response to profit shifting activities.

is left undistorted (e.g. in case of a production function $k^{\psi} + \theta$). In this case, firms of equal capital size differ in profits, and a tax on capital use cannot be used to discriminate between firms.

As discussed above, the finding that capital use is optimally distorted is not novel, e.g. Haufler/Schjelderup (2000) show that capital use is taxed in the presence of profit shifting. The above Proposition, however, does not simply replicate the findings derived from optimization problems with limited set of tax instruments. It adds general conditions under which circumstances capital is taxed or subsidized and it shows that capital distortions may be unavoidable – even if lump-sum taxes are allowed for.

A novel finding (to the best of my knowledge) is the lump-sum tax element of the optimal tax system in part (ii) of the above Proposition. A lump-sum tax element is known from the literature on Tiebout (1956) style competition to internalize the congestion cost of public good provision. In the model presented here, however, it is optimal to use such a tax even in the absence of a marginal congestion cost. That is, in the optimal tax system, a kind of an "alternative minimum tax" drives drive out the least profitable firms out of the market in order to implement higher lump-sum taxes on the remaining ones. As will be shown, this finding does not depend on the Leviathan assumption (revenue maximization).

The level of public input provision g depends on the complementarity between θ and g, see the right hand side of eq. (12). The sign and the size of $\pi_{\theta g}$ determine whether (and how much) the public input makes the profit-type slope steeper or flatter (see Boadway/Keen 1993 for similar considerations in a two-type model). In addition, optimal public input provision may be affected by the optimal distortion of the firm's capital stock whenever $\pi_{gk} \neq 0$, see the left hand side of (12).

Corollary 1 *(i) Public inputs and capital inputs are undistorted if* $\pi^{\theta}_{\theta k} = \pi^{\theta}_{\theta g} = 0$ *for all* θ.
(ii) Public inputs and capital inputs are lower than in the first best if both inputs are complements with θ *and* $\pi^{\theta}_{gk} \geq 0$.
(iii) Public inputs and capital inputs are higher than in the first best if both inputs are substitutes with θ *and* $\pi^{\theta}_{gk} \geq 0$.

An example for case (i) is a production technology of $\theta k^{\psi} + g$ with $\psi \in (0, 1)$. An example for case (ii) is production technology of $\theta g k^{\psi}$. Then, public inputs make capital use more productive and more so for higher productivity types. An example for case (iii) is a production function of $(gk+\theta)^{\psi}$. Then, public inputs make capital use more productive, but capital is a substitute for θ. By providing inputs, the government make conventional brick-and-mortar industries more productive and – as a side effect – facilitates taxation by relaxing the ICC.

Discussion

It is the double type of mobility of both capital and taxable profit that creates the distortions necessary for the second best allocation. If taxable profit is observable (and therefore immobile), the first best can be implemented as shown above. If capital is immobile, efficient taxation is feasible with the tax system completely conditioning on capital inputs (and, thus, ignoring mobile profit). The latter is only possible, though, if capital use is correlated with profitability.

It may be instructive to briefly compare the above model of second best source taxation to the Mirrlees (1971) model of income taxation. In the above model, the equivalent to the Mirrleesian unobservable skill level is the firm's unobservable profitability. The equivalent to the disutility of work is the net return to profit shifting, both are not directly taxable.[25] Note that, if the government had access to taxes on income shifted to tax havens, the net return to profit shifting would be taxable and, therefore, efficient taxation would be possible.

In the above model, there is no redistribution motive. In contrast, revenue is maximized which implies that lower type firms may be pushed out of the market (an effect that does not occur in the Mirrlees model since individuals do not vanish when they are pushed out of the market and, therefore, remain part of the welfare function).

An interesting question is whether optimal tax rates on reported profit can be constant across types – as they are in many real-world tax systems (at least if one ignores the implicit tax due to type-dependent enforcement). The optimality condition for $T_{\hat{\pi}}(\theta)$ under a Pareto c.d.f. (with a Pareto parameter of ρ and assuming $\bar{\theta}\to\infty$) and quadratic concealment cost (implying a constant $\kappa''(\Delta)=\bar{\kappa}''$) is

$$T_{\hat{\pi}}(\theta)=\pi_{\theta}\cdot\bar{\kappa}''\cdot\frac{\theta}{\rho} \tag{13}$$

That is, to be independent of type, π_{θ} must equal $\frac{c}{\theta}$ (where c denotes a constant). An example for $\pi_{\theta}=\frac{c}{\theta}$ is a production function of the type $\ln k\theta g$ that implies

[25] In contrast, both the return and cost of capital investment is observable (and therefore fully taxable by making the cost of capital deductible).

$\pi_\theta = \frac{1}{\theta}$. Note that, in this case, the optimal marginal tax on capital must be zero.

2.4 National welfare maximization

The policy goal considered above, tax revenue maximization, implies that the government puts a zero welfare weight on firm profits. This may be a plausible assumption if (i) all firms are held by foreigners or (ii) all profits accrue to the rich and the government pursues (some kind) of Rawlsian welfare maximization goal.

In this subsection, I consider maximizing national welfare as a policy goal. National welfare depends on domestic firms' profits and tax revenue. The latter has a welfare weight of $\lambda \geq 1$ which reflects that public funds from alternative source have a (constant) marginal efficiency cost.[26] Domestic ownership of firms in the source country is captured by $\alpha(\theta)$, the fraction of domestic ownership of firm type θ. The government's objective function is thus given by

$$\int (\alpha(\theta)\Pi(\theta) + \lambda[\pi(k,g,\theta) - \kappa(\Delta) - \Pi(\theta)])\, f(\theta)\, d\theta - \lambda c(g) \tag{14}$$

With $\alpha(\theta) = 0$ for all θ, the case of revenue maximization is modelled (and λ is redundant). With $\lambda = 1$, the case of national income maximization is captured.

Proposition 3 *Assume that the government maximizes national welfare given in (14). Then, optimal tax rates are*

$$T_{\hat{\pi}}(\theta) = \frac{\lambda - \bar{\alpha}_\theta}{\lambda} \frac{1 - F(\theta)}{f(\theta)} \pi_\theta \kappa'' \tag{15}$$

$$\tilde{T}_k(\theta) = \frac{\lambda - \bar{\alpha}_\theta}{\lambda} \frac{1 - F(\theta)}{f(\theta)} \pi_{\theta k}(1 - \kappa') \tag{16}$$

where $\bar{\alpha}_\theta$ where is the average domestic firm ownership for firms of types higher than θ. Again, tax rates $\tilde{T}_k(\theta)$ and $T_{\hat{\pi}}(\theta)$ are zero for $\theta \in \{\underline{\theta}, \bar{\theta}\}$. The lump-sum

[26] National welfare maximization is equivalent to maximizing the representative household's utility if the latter owns all domestic firms and has a utility function that linearly depends on private consumption (out of dividend income) and λ times the public goods financed by tax revenue.

tax element is given by $T\left(\underline{\theta}^T\right) = \pi^{\underline{\theta}^T} - \Pi^o$ *with*

$$T\left(\underline{\theta}^T\right) = \frac{\lambda - \bar{\alpha}_\theta}{\lambda} \frac{1 - F\left(\underline{\theta}^T\right)}{f\left(\underline{\theta}^T\right)} \pi_\theta \left(k, g, \underline{\theta}^T\right) + c_n = 0 \qquad (17)$$

The optimality condition for public input provision is

$$\int_{\underline{\theta}^T}^{\bar{\theta}} \pi_g f(\theta)\, d\theta - c_g = \int_{\underline{\theta}^T}^{\bar{\theta}} \frac{\lambda - \bar{\alpha}_\theta}{\lambda} \frac{1 - F(\theta)}{f(\theta)} \pi_{\theta g} \left(1 - \kappa'\right) f(\theta)\, d\theta \qquad (18)$$

Proof. See the Appendix. ■

National welfare maximization differs from tax revenue maximization if $\frac{\lambda - \bar{\alpha}_\theta}{\lambda} \neq 1$. With $\lambda \in [1, \infty)$ and $\bar{\alpha}_\theta \in [0, 1]$ and, thus,

$$\frac{\lambda - \bar{\alpha}_\theta}{\lambda} = 1 - \frac{\bar{\alpha}_\theta}{\lambda} \in [0, 1] \qquad (19)$$

it follows that, as long as $\bar{\alpha}_\theta > 0$, optimal marginal tax rates are closer to zero than under revenue maximization, the lump-sum tax is smaller and the distortion of public input provision is smaller. With full domestic ownership, $\bar{\alpha}_\theta = 1$ for all θ, the revenue maximizing tax rates are scaled by a factor $\frac{\lambda - 1}{1}$ that reflects the spending need for public inputs and the loss in private income. As a consequence, if there is an efficient revenue-raising (tax) instrument and, thus, $\lambda = 1$, optimal tax rates on reported profit and capital inputs are zero if $\bar{\alpha}_\theta = 1$ for all θ.

Both parameters, λ and $\alpha(\theta)$ are relevant for understanding potential deviations between the *national* welfare maximization and *global* welfare maximization (an argument that has been put forward by Keen/Wildasin 2004). This can be illustrated for the case in which all countries have a zero-cost revenue source (not explicitly modelled here) that implies $\lambda = 0$. In this case, global welfare maximization implies marginal tax rates on reported profit and capital inputs of zero. However, even slight deviations from full domestic ownership justify a non-zero tax on both mobile tax bases and, thus, some efficiency cost (note, though, that the efficiency cost capital input taxation may be zero if tax rates are the same on all types and all countries). In other words, global welfare maximization requires the government to set tax policy as if $\bar{\alpha}_\theta = 1$ for all θ.

2.5 A simple policy experiment

The above model can be used to analyze the effects of policy measure on tax rate setting and welfare. For purpose of illustration, I provide a simple example.

A whole class of recent policy measures aims at reducing profit shifting to tax havens. Let $\kappa(\Delta) = \beta\Delta + \tilde{\kappa}(\Delta)$ denote a specific concealment cost function where β is a policy variable and $\tilde{\kappa}(\Delta)$ is strictly convex in Δ (the haven's tax rate is assumed to be zero). By increasing β, the government increases the marginal cost of shifting profit to the tax haven – which may be interpreted e.g. as a tightening of transfer pricing rules.

How does a change in β affect tax policy? The optimal tax on reported profit is given by $T_{\hat{\pi}}(\theta) = \frac{\lambda - \bar{\alpha}_\theta}{\lambda}\frac{1-F(\theta)}{f(\theta)}\pi_\theta^\theta\kappa''$. For a given tax on capital inputs, only κ'' responds to changes in Δ or $T_{\hat{\pi}}(\theta)$. Using $T_{\hat{\pi}}(\theta) = \kappa'(\Delta)$, it follows that

$$\frac{d\Delta}{d\beta} = -\frac{1}{\tilde{\kappa}'' - \frac{\lambda - \bar{\alpha}_\theta}{\lambda}\frac{1-F(\theta)}{f(\theta)}\pi_\theta^\theta\kappa'''} \tag{20}$$

With $T_{\hat{\pi}}(\theta) = \kappa'(\Delta)$, this translates into a change in the tax rate according to

$$\frac{dT_{\hat{\pi}}(\theta)}{d\beta} = 1 - \frac{\tilde{\kappa}''}{\tilde{\kappa}'' - \frac{\lambda - \bar{\alpha}_\theta}{\lambda}\frac{1-F(\theta)}{f(\theta)}\pi_\theta^\theta\tilde{\kappa}'''} \tag{21}$$

If $\tilde{\kappa}''' = 0$, the tax rate stays the same. If $\tilde{\kappa}''' > 0$, the tax rate increases, if $\tilde{\kappa}''' < 0$ decreases.[27]

For purpose of illustration, consider first the case of $\tilde{\kappa}''' = 0$.[28] Then, $\tilde{\kappa}''$ is a constant independent of Δ. Accordingly, β does not affect the optimal level of $T_{\hat{\pi}}(\theta)$. With $T_{\hat{\pi}}(\theta) = \kappa'(\Delta) = \beta + \tilde{\kappa}'$, it follows that an increase in β by one unit reduces Δ by $1/\tilde{\kappa}''$. Since $T_{\hat{\pi}}(\theta)$ stays the same, $\tilde{T}_k(\theta)$ stays the same and so does optimal public input provision. That is, the only policy effect is a reduction of profit shifting. With regard to welfare, a small increase of β by $d\beta$ reduces profit by $d\beta/\tilde{\kappa}''$ and increases tax revenue by $T_{\hat{\pi}}(\theta)\,d\beta/\tilde{\kappa}''$. Whether this is a welfare improvement, depends on the marginal value of public funds, λ. However, if β were an additional payment to the government, e.g. an expected fine, the revenue increase would be $(1 + T_{\hat{\pi}}(\theta))\,d\beta/\tilde{\kappa}''$ and, therefore, unambiguously a welfare improvement.

Now, consider the case of $\tilde{\kappa}''' > 0$. An increase in β increases the tax on

[27] Note that the denominator, $\tilde{\kappa}'' - \frac{\lambda - \bar{\alpha}_\theta}{\lambda}\frac{1-F(\theta)}{f(\theta)}\pi_\theta^\theta\tilde{\kappa}'''$, is always positive since it corresponds to the second derivative of the government's objective function.

[28] This case occurs if the concealment cost function is e.g. $\kappa(\Delta) = \beta\Delta + \frac{1}{2}\Delta^2$.

reported profit, $T_{\hat{\pi}}(\theta)$. It follows from optimality conditions (16) and (18) that the distortions of capital use and public inputs are reduced, at least if either both inputs are positively correlated with θ or both are negatively correlated with θ.[29] If, however, $\tilde{\kappa}''' < 0$, the allocation becomes more distorted.

The above considerations may be used to shed light on the effects of minimum taxes on tax haven profits. The tax inclusive cost of profit shifting is then given by $t^h \Delta + \kappa(\Delta)$ where t^h is a proportional tax rate on tax haven income Δ. By setting $\beta = t^h$, the above analysis can be interpreted in terms of minimum taxes. Depending on the third derivative of the concealment cost function, minimum taxes will lead to more efficient or less efficient capital use and public input provision.

2.6 Type-dependent profit shifting

In the above model, the concealment cost function is identical across types. This assumption may be too strong since empirical evidence (e.g. Davies et al. 2018) suggests that the propensity to shift profit varies across firm types. I will therefore now assume that concealment cost is type-dependent: $\kappa = \kappa(\Delta, \theta)$.

The ICC needs to be modified to

$$\Pi(\theta)' = \pi_\theta^\theta (1 - \kappa_\Delta) - \kappa_\theta \tag{22}$$

Proposition 4 *With type-dependent concealment cost $\kappa = \kappa(\Delta, \theta)$, the optimal tax rate on reported profit is given by*

$$T_{\hat{\pi}}(\theta) = \frac{1 - F(\theta)}{f(\theta)} \left[\pi_\theta^\theta \kappa_{\Delta\Delta} + \kappa_{\theta\Delta} \right] \tag{23}$$

The optimal tax rate on capital input is, as before, given in eq. (10).

Proof. See the Appendix. ■

The above Prop. implies that optimal marginal tax rates on reported profit account for the fact that some firm types are more prone to profit shifting. The optimal marginal tax rates differ from those in Prop. 2 in $\kappa_{\Delta\Delta}$ and in $\kappa_{\theta\Delta}$. The former is a measure of how elastic reported profit responds to a change in the marginal incentive to shift profit; low levels of $\kappa_{\Delta\Delta}$ require low marginal tax

[29] If they are both positively correlated with profitability, both capital and public inputs are increased. If they are negatively correlated (and, thus, there is an upward distortion to capital and public input use), both factors are reduced. In any case, the 'crack-down' on tax havens improves the efficiency of the resource allocation within the economy.

rates (ceteris paribus). The latter measures how the profit shifting elasticity changes with the type. If $\kappa_{\theta\Delta} > 0$, higher types have higher profit shifting cost; accordingly, optimal marginal tax rates are higher (ceteris paribus).

2.7 Sufficient statistics approach

The above derived optimality conditions for marginal tax rates on reported profits and capital are not easily translated into empirical applications. I will therefore use Saez' (2001) method to express optimal tax rates as functions of the underlying behavioral elasticities for observable reported profit levels and capital stocks (instead of types).

Since much of the empirical literature uses firm size as a differentiating variable, I use the firm's capital stock as the equivalent to type.

Consider a population of firms that differs in capital stocks k. Let $\hat{\pi}(k)$ denote the reported profit of a firm with a capital stock k. I will denote as $f(k)$ the density of firms with a capital stock k and as $f(\hat{\pi}|k)$ the density of firms with a reported profit of $\hat{\pi}$ for a given level of capital k. In the Appendix, I outline how $f(k)$ relates to the type density $f(\theta)$.

Tax revenue is given by

$$\int_0^{\infty}\left[\int_{-\infty}^{\infty} T(k,\hat{\pi}(k))\, f(\hat{\pi}|k)\, d\hat{\pi}\right] f(k)\, dk \tag{24}$$

First, consider the small increase in $T_{\hat{\pi}}$ denoted as $d\tau_{\hat{\pi}}$ that affects the schedule between $\hat{\pi}^*$ and $\hat{\pi}^* + d\hat{\pi}$. With a revenue-maximizing tax schedule, the total marginal effect on tax revenue is zero. Using this property, optimal tax rates can be derived (see the Appendix for a formal derivation).

Let c denote the non-tax cost of capital and $\varepsilon_{k,c+\tilde{T}_k}$ the user cost elasticity of capital. The optimal tax rates on reported profit and capital are implied by

$$\frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}} = \frac{1-F(\hat{\pi}^*)+\int_0^{\infty}\varepsilon_{k,c+\tilde{T}_k}\frac{\tilde{T}_k^2}{c+\tilde{T}_k}\frac{k}{(1-T_{\hat{\pi}})}f(\hat{\pi}^*|k)\,f(k)\,dk}{\int_0^{\infty}\left[\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}}\hat{\pi}\right]f(\hat{\pi}^*|k)\,f(k)\,dk} \tag{25}$$

$$\frac{\tilde{T}_k}{c+\tilde{T}_k} = \frac{1-F(k^*)}{f(k^*)}\frac{1}{k^*\cdot\int_{-\infty}^{\infty}\left(-\varepsilon_{k,c+\tilde{T}_k}\right)\left[1+\frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}}\frac{d\hat{\pi}}{d\pi}\right]f(\hat{\pi}|k^*)\,d\hat{\pi}} \tag{26}$$

where $F(\hat{\pi}^*)$ is the fraction of firms with a reported profit of $\hat{\pi}^*$ or less.

For purpose of illustration, assume for a moment that there is a perfect mapping of capital stocks and reported profit, i.e. there is no variance of reported profit within each capital stock class. Then, the optimal tax rate on reported profit is implied by

$$\frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}} = \frac{1-F\left(\hat{\pi}^{*}\right)}{f\left(\hat{\pi}^{*}\right)} \frac{1}{\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}}\hat{\pi}} + \frac{\tilde{T}_k^2}{c+\tilde{T}_k} \frac{\varepsilon_{k,c+\tilde{T}_k}}{\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}}\hat{\pi}} \frac{k}{1-T_{\hat{\pi}}} \tag{27}$$

If capital use is undistorted, the optimal marginal tax on reported profit is simply given by $\frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}} = \frac{1-F(\hat{\pi}^{*})}{f(\hat{\pi}^{*})} \frac{1}{\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}}\hat{\pi}}$, which is trade-off between the revenue effect of taxing a fraction $1-F\left(\hat{\pi}^{*}\right)$ of all firms and the behavioral distortion that affects $f\left(\hat{\pi}^{*}\right)$ of all firms and that is quantified by $\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}}\hat{\pi}$. With distorted capital use, $\tilde{T}_k \neq 0$, the efficiency cost of taxing reported profit is increased (independent of the sign of $\tilde{T}_k$) and optimal tax rates are lower. The more general expression above captures the variance of reported profit within each capital class.

With a perfect mapping of capital stocks and reported profit, the optimal tax on capital can be expressed as

$$\frac{\tilde{T}_k}{c+\tilde{T}_k} = \frac{1-F\left(k^{*}\right)}{f\left(k^{*}\right)} \cdot \frac{1}{k^{*}\left(-\varepsilon_{k,c+\tilde{T}_k}\right)\left[1+\frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}}\frac{d\hat{\pi}}{d\pi}\right]} \tag{28}$$

As before, the optimal tax rate on capital input is governed by (i) the mass of firms that increase their tax payments, $1-F\left(k^{*}\right)$, (ii) the mass of firms suffering from behavioral distortions, $f\left(k^{*}\right)$, and (iii) the (weighted) behavioral elasticity $\varepsilon_{k,c+\tilde{T}_k}$. In addition, $\frac{d\hat{\pi}}{d\pi}$ captures how much reported profit is affected if true profit varies (due to a change in the capital tax). If $\frac{d\hat{\pi}}{d\pi} > 0$, the optimal tax on capital is lower because distorting capital input has an 'externality' on another tax base (here: the base of the tax on reported profit).

3 Model with mobile firms

In this section, I introduce firm mobility. As a preliminary step, I consider an improvement of the outside option Π^{o}. A variation in the outside option value provides some (albeit rudimentary) insights in the effects of firm mobility. The main contribution of this section is the analysis of heterogeneous mobility cost for firms within each type (as in Lehmann et al. 2014).

3.1 Uniform outside option

In the model in Sect. 2, each firm type earns $\Pi^o \geq 0$ if it opts out of the market. Since after-tax profit increases in the type (according to the ICC in (6)), only the lowest type's participation condition is binding. Here, I will consider an increase in the uniform outside option. There are several scenarios where such an assumption makes sense. First, think of outside investors consider withdrawing their capital and, instead, invest in some other asset with a type-independent return. Second, the firm type may be location-specific; outside of the location, each firm has the same profit. Third, the outside option may, in fact, be type-dependent but it increases at a lower rate in the type than after-tax profit with a revenue-maximizing tax system.

Making the outside option more attractive increases welfare 'mechanically' by making all firms choosing the outside option better off. An increase in welfare due to a more attractive outside option is therefore not meaningful as a welfare criterion. Instead, I will check whether an increase in Π^o *raises welfare by more the income increase of firms that choose* Π^o *in the status quo.*

The following Corollary considers a setting where the government maximizes national welfare with full domestic ownership and $\lambda > 1$ (recall that $\lambda = 1$ would imply zero taxation). The Corollary can thus be understood as referring to Prop. 3.

Corollary 2 *Consider the model in Sect. 2.4 with full domestic ownership, $\alpha(\theta) = 1$ for all θ, and a welfare weight of $\lambda > 1$ on public revenue. A small increase in Π^o by $d\Pi^o$ increases welfare by less than $\int_{\underline{\theta}}^{\underline{\theta}^T} d\Pi^o f(\theta)\, d\theta$ (i.e. by less than the profit increase of firms choosing the outside option in the status quo).*

Proof. See the Appendix. ■

The intuition behind the above Corollary can be explained as follows. The marginal firm type $\underline{\theta}^T$ ignores that, by choosing the outside option, the government would lose $T\left(\underline{\theta}^T\right) - c_n > 0$ (the strictly positive sign is due to eq. (17)). Thus, by making the outside option more attractive, more firms choose the Π^o and, thus, reduce the government's budget. Since all other variables are optimally chosen, this is the only direct effect (apart from the revenue increase for all firms choosing the outside option).

The above Proposition implies that taxing the outside option would increase welfare even if the resulting revenue were given back to the households in lump-sum manner, i.e. if $d\Pi^o < 0$, and households get, in return, a lump-sum transfer of $-d\Pi^o$ for each firm that chooses the outside option in the status quo.

Note that, by construction of the model, there is no externality on other countries – which implies that the above Corollary may be understood in terms of global welfare.

3.2 Heterogeity in mobility cost

I now augment the above model following the example of Lehmann et al. (2014) and assume that, within each type θ, there is a continuum of firms with heterogeneous migration cost $m \in \mathbb{R}_+$. Depending on taxes and public input levels, the source country under consideration (labelled 'home') may lose some of its firms to the rest of the world or attract some from it. For notational simplicity, I will refer to the rest of the world as a country on its own (labelled 'abroad'). Let $f(\theta)$ and $f^a(\theta)$ denote the density of firms at home and abroad if no migration takes place. Let $\gamma(m|\theta)$ denote the conditional density of m for a given type θ and $\Gamma(m|\theta) = \int \gamma(m|\theta)\, dm$ the associated cumulated distribution function. $\gamma^a(m|\theta)$ and $\Gamma^a(m|\theta)$ denote the equivalent variables for firms abroad.

A home firm of type θ with migration cost m will choose to stay at home if

$$\Pi(\theta) \geq \Pi^a(\theta) - m \tag{29}$$

where $\Pi^a(\theta)$ is the net-of-tax profit of locating abroad. That is, if m' denotes the level of migration cost at which the above equation holds with strict equality, I can conclude that a fraction $1 - \Gamma(m'|\theta)$ of type θ firms stays in the current location. Equivalently, a firm from abroad is indifferent between staying and moving to 'home' if $\Pi(\theta) - m^{a\prime} = \Pi^a(\theta)$ and, thus, $\Gamma^a(m^{a\prime}|\theta)$ is the fraction of abroad firms to locate at home.

The density of firms locating at home is thus given by

$$\varphi(\Lambda, \theta) = \begin{cases} (1 - \Gamma(m|\theta))\, f(\theta) & \text{if } \Delta \leq 0 \\ f(\theta) + \Gamma^a(m|\theta)\, f^a(\theta) & \text{if } \Delta > 0 \end{cases} \tag{30}$$

where $\Lambda = \Pi(\theta) - \Pi^a(\theta)$. An increase in $\Pi(\theta) - \Pi^a(\theta)$ thus increases the density, $\varphi_\Lambda > 0$. Let $\Phi(\Lambda, \theta) = \int \varphi(\Lambda, \theta)\, d\theta$ denote the migration adjusted fraction of firms in the home country for a type θ or lower.

Thus, each individual firm is characterized by three variables: its type θ, its migration cost m and its original location. Optimal tax rate setting and public input provision responds to (the threat of) firm migration. For the following Proposition, let $\eta(\theta) = \frac{\varphi_\Lambda}{\varphi(\Lambda,\theta)} > 0$ be the semi-elasticity of migration with respect to a marginal change in Λ, and let $\overline{T\eta}^{\theta} = \frac{1}{1-\Phi(\Lambda,\theta)} \int_\theta^\infty [T(\theta)\eta(\theta)]\varphi(\Lambda,\theta)\,d\theta$ be the average tax-revenue-weighted migration semi-elasticity for types above θ.

Proposition 5 *With heterogeneous migration cost, second best marginal tax rates are given by*

$$T_{\hat{\pi}}(\theta) = \frac{1-\Phi(\Lambda,\theta)}{\varphi(\Lambda,\theta)} \cdot \pi_\theta \cdot \kappa'' \cdot \left(1 - \overline{T\eta}^{\theta}\right) \tag{31}$$

$$\tilde{T}_k(\theta) = \frac{1-\Phi(\Lambda,\theta)}{\varphi(\Lambda,\theta)} \cdot \pi_{\theta k} \cdot (1-\kappa') \cdot \left(1 - \overline{T\eta}^{\theta}\right) \tag{32}$$

for all $\theta \in \left(\underline{\theta}, \bar{\theta}\right)$ *and zero for* $\theta \in \left\{\underline{\theta}, \bar{\theta}\right\}$. *The lump-sum tax element equals* $\pi\left(\underline{\theta}^T\right) - \tilde{\Pi}^o$ *with*

$$\pi\left(\underline{\theta}^T\right) - \tilde{\Pi}^o = \pi_\theta\left(k, g, \underline{\theta}^T\right) \frac{1-\Phi\left(\Lambda, \underline{\theta}^T\right)}{\varphi\left(\Lambda, \underline{\theta}^T\right)} \left(1 - \overline{T\eta}^{\underline{\theta}^T}\right) + c_n(g,n) \tag{33}$$

where $\tilde{\Pi}^o \geq 0$ *denotes the best available outside option. Optimal public input provision is implied by*

$$\int \pi_g \varphi(\Lambda,\theta)\,d\theta - c_g(g,n) \tag{34}$$

$$= \int \left[\frac{1-\Phi(\Lambda,\theta)}{\varphi(\Lambda,\theta)} \pi_{\theta g}(1-\kappa')\left(1 - \overline{T\eta}^{\theta}\right)\right] \varphi(\Lambda,\theta)\,d\theta \tag{35}$$

Proof. See the Appendix. ■

I can now outline the properties of the symmetric competitive equilibrium. For this purpose, I consider two identical source countries and a zero-rate tax haven.

Corollary 3 *In the symmetric tax and public input competition equilibrium, both identical countries levy the taxes outlined in Prop. 5. All firms stay in their original location (i.e. no moving cost is incurred). Compared to the case of immobile firms, tax rates are closer to zero, and public input provision is closer to its efficient level.*

Proof. Omitted. ■

How does firm mobility affect welfare? The following Corollary refers to the case of symmetric countries and full domestic ownership.

Corollary 4 *Assume a symmetric tax and public input competition equilibrium with firm migration. Then, the following holds.*
(i) A coordinated tax increase on reported profit and a coordinated tax variation on capital use increases welfare.
(ii) The tax coordination optimum is described by Prop. 3.
(iii) Allowing for firm migration reduces welfare.

Proof. Omitted. ■

Instead of a formal proof, I briefly outline the intuition behind the Corollary for the case of two symmetric countries. A coordinated increase of the tax on reported profit for each type has no migration effect. Starting from the uncoordinated Nash equilibrium, the net effect on welfare is therefore positive. With regard to the tax on capital, it depends on whether capital use is taxed ($\pi^{\theta}_{\theta k} > 0$) or subsidized ($\pi^{\theta}_{\theta k} < 0$) in equilibrium. In the former case, a tax increase increases welfare, in the latter case, a tax reduction does. Since there are no externalities between production locations in the absence of firm migration, the coordination optimum is the optimum without firm migration as described in Prop. 3. It follows that allowing for firm migration reduces welfare.

4 Conclusion

In this paper, I analyze a setting where firms differ in profitability, employ mobile capital, shift profit to tax havens and may migrate to other locations. Source locations provide public inputs and levy taxes. Even if source country governments maximize global welfare, some distortions are inevitable. For an intuitive understanding, note that the two obvious candidates for efficient taxation have some crucial disadvantages: first, uniform lump-sum taxes distort market entry (since they may exceed some firm types' profit) and, second, proportional profit taxes induce costly profit shifting. Therefore, the second-best tax system has positive marginal tax rates on reported profit and, in many cases, a non-zero tax on capital inputs. It includes a lump-sum element which implies that even non-profitable firms pay a strictly positive tax. Such an 'alternative minimum tax' is not motivated by congestion costs of public (input) provision. In addition, the level of public inputs may be optimally distorted in order to alleviate distortions on the tax side.

With benevolent governments, (symmetric) tax competition for mobile capital, profits and production facilities reduces welfare, at least if foreign firm ownership is sufficiently small. A crack-down on tax havens, e.g. by implementing a minimum tax on profits located in tax havens, potentially increases welfare, but may further distort capital use and public input provision (depending on the technological features of the concealment cost function).

The model can be used to build an understanding of typical features of real world tax systems as well as to compare theoretically optimal tax systems with real-world tax systems (which do not only account for behavioral distortions, but need to take into account equal treatment provisions and are shaped by other than purely efficiency driven considerations). While the above analysis makes a first step to derive optimum business taxation from a unconstrained set of tax instruments, the generality of the analysis is constrained in (at least) two important ways.

First, the above analysis neglects residence-based taxation of corporate profits (apart from the minimum tax on profits located in tax havens). This assumption may be justified by the fact that most corporate tax systems are now based on a territorial system. However, residence-based taxes may interact with source-based taxes in an important way and, as I have argued above for non-linear source-based taxes on profit and capital, this type of tax should not just be 'assumed away'.

Second, a number of important decision margins of the MNE are ignored. In a more general model, firms choose between different sources of finance (equity vs. debt). They may choose to split up or to merge with other firms, if this lowers their tax payments, and they may restructure the chain of production. Especially the last decision margin seems to be of importance for the analysis of multinational firms. With production taking place in different parts of the firm, transfer pricing rules are of central importance. In the above paper, I assumed that the rules for transfer prices (used to shift profit to the tax haven) are given and cannot be affected by the source country government. In a more complicated model, governments could be assumed to not only optimize over tax rates and the base (for a given taxable local income), but also to negotiate with other source countries over transfer pricing rules. This, however, requires a more complicated model, since there are, at least, two governments involved – which makes it a model of common agency (see Gresik/Nelson 1994 for an early contribution). While this is clearly beyond the scope of this paper, the question how optimum source taxes look like if governments coordinate over transfer pricing (and other accounting rules), but not over tax rates, seems an interesting and fruitful one.

5 Appendix

5.1 Proof of Prop. 2

Proof. The Hamiltonian of the government's optimization problem is

$$\begin{aligned} H &= [\pi(k,g,\theta) - \kappa(\Delta) - \Pi(\theta)] f(\theta) \\ &\quad + \mu(\theta)\pi_\theta(k,g,\theta)(1-\kappa') + \nu(\theta)\Pi(\theta) f(\theta) \end{aligned}$$

For a given firm population (i.e. a given $\underline{\theta}^T$) and a given level of g, the government solves $\max_{k(\theta),\Delta(\theta)} H$ for each θ. The first order conditions with respect to k and Δ are

$$\begin{aligned} \pi_k f(\theta) + \mu\pi_{\theta k}(1-\kappa') &= 0 \\ -\kappa' f(\theta) - \mu\pi_\theta\kappa'' &= 0 \end{aligned}$$

as well as $\frac{\partial H}{\partial \Pi(\theta)} = \nu(\theta) f(\theta) - f(\theta) = -\mu'(\theta)$ for all θ. The transversality conditions are $\mu\left(\underline{\theta}^T\right) = 0$ and $\mu(\bar{\theta}) = 0$. Since the ICC implies that net profit increases in θ, the participation constraint is only binding for the lowest type, i.e. $\nu\left(\underline{\theta}^T\right) > 0$ and $\nu(\theta) = 0$ for all $\theta > \underline{\theta}^T$. Integrating $\mu'(\theta) = f(\theta)$ over types gives $\mu(\theta) = \int_{\underline{\theta}^T}^{\theta} f(\theta)\, d\theta = F(\theta) - F\left(\underline{\theta}^T\right)$. With $\mu'\left(\underline{\theta}^T\right) = -\nu\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) + f\left(\underline{\theta}^T\right)$ and $\mu'(\theta) = f(\theta)$, $\mu(\theta)$ can be expressed as $\mu(\theta) = -\nu\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) + \int_{\underline{\theta}^T}^{\theta} f(\theta)\, d\theta$. Since $\mu(\bar{\theta}) = 0$, it must be that $\nu\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) = \int_{\underline{\theta}^T}^{\bar{\theta}} f(\theta)\, d\theta$. It follows

$$\mu(\theta) = -\int_{\underline{\theta}^T}^{\bar{\theta}} f(\theta)\, d\theta + \int_{\underline{\theta}^T}^{\theta} f(\theta)\, d\theta = -(1 - F(\theta)) < 0$$

The first optimality condition (optimizing over k) above can be expressed as

$$\tilde{T}_k(\theta) = \frac{1 - F(\theta)}{f(\theta)} \pi_{\theta k}(1 - \kappa')$$

where I used $\frac{1-\tilde{F}(\theta)}{\tilde{f}(\theta)} = \frac{1-F(\theta)}{f(\theta)}$ (see above). And the second optimality condition (optimizing over $\hat{\pi}$) can now be expressed as

$$T_{\hat{\pi}}(\theta) = \frac{1-F(\theta)}{f(\theta)} \pi_\theta \kappa''$$

Now, turn to the provision of public input goods. The government sets g to maximizes $\int_{\underline{\theta}}^{\bar{\theta}} H d\theta - c(g)$. The first order condition with respect to g is

$$\int (\pi_g f(\theta) + \mu(\theta) \pi_{\theta g} (1-\kappa')) d\theta - c_g = 0$$

from which follows

$$\int \pi_g f(\theta) d\theta - c' = \int \frac{1-F(\theta)}{f(\theta)} \pi_{\theta g} (1-\kappa') f(\theta) d\theta$$

Finally, consider the choice of $\underline{\theta}^T$. The choice of $\underline{\theta}^T$ does not affect the intra-marginal types' optimality conditions for marginal tax rates, i.e. the ICC is not affected. Accordingly, the government chooses $\underline{\theta}^T$ to maximize its objective function, $\int_{\underline{\theta}^T}^{\bar{\theta}} T(\theta) d\theta - c(g, n)$. With the above, tax revenue can be expressed as the sum of a lump-sum tax element and the integrals over taxing reported profit and capital. The lump-sum tax element is equal to the difference between the $\underline{\theta}^T$-type pre-tax profit and its outside option. Thus, tax revenue is given by

$$\int_{\underline{\theta}^T}^{\bar{\theta}} T(\theta) f(\theta) d\theta = \int_{\underline{\theta}^T}^{\bar{\theta}} \left[\pi\left(k, g, \underline{\theta}^T\right) - \Pi^o + \int_{\pi\left(\underline{\theta}^T\right)}^{\hat{\pi}(\theta)} T_{\hat{\pi}}(\theta) d\hat{\pi} + \int_{k\left(\underline{\theta}^T\right)}^{k(\theta)} T_k(\theta) dk \right] f(\theta) d\theta$$

The first-order condition is

$$\int \pi_\theta \left(k, g, \underline{\theta}^T\right) f(\theta) d\theta - T\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) + f(\underline{\theta}) c_n = 0$$

from which follows the optimality condition in (11). ■

5.2 Proof of Prop. 3

Proof. For a given firm population and a given level of g, the Hamiltonian is

$$\begin{aligned} H &= [\alpha(\theta) \Pi(\theta) + \lambda [\pi(k, g, \theta) - \kappa(\Delta) - \Pi(\theta)]] f(\theta) \\ &\quad + \mu(\theta) \pi_\theta (1-\kappa') + \nu(\theta) \Pi(\theta) f(\theta) \end{aligned}$$

The government solves $\max_{k(\theta),\Delta(\theta)} H$ for each θ. The first order conditions with respect to k and $\hat{\pi}$ are

$$\begin{aligned}\lambda \pi_k f(\theta) + \mu(\theta)\pi_{\theta k}(1-\kappa') &= 0 \\ -\lambda\kappa' f(\theta) - \mu(\theta)\pi_\theta \kappa'' &= 0\end{aligned}$$

as well as $\frac{\partial H}{\partial \Pi(\theta)} = \nu(\theta) f(\theta) - (\lambda - \alpha(\theta)) f(\theta) = -\mu'(\theta)$ for all θ. The transversality conditions are $\mu\left(\underline{\theta}^T\right) = 0$ and $\mu(\bar{\theta}) = 0$. The ICC states that net profit increases in θ implying that the participation constraint can only be binding for the lowest type, $\nu\left(\underline{\theta}^T\right) > 0$, with $\nu(\theta) = 0$ for all $\theta > \underline{\theta}$. With $\mu'\left(\underline{\theta}^T\right) = -\nu\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) + \left(\lambda - \alpha\left(\underline{\theta}^T\right)\right) f\left(\underline{\theta}^T\right)$ and $\mu'(\theta) = (\lambda - \alpha(\theta)) f(\theta)$ otherwise, $\mu(\theta)$ can be expressed as $\mu(\theta) = -\nu\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) + \int_{\underline{\theta}^T}^{\theta} (\lambda - \alpha(\theta)) f(\theta)\, d\theta$. Since $\mu(\bar{\theta}) = 0$, it must be that $\nu\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) = \int_{\underline{\theta}^T}^{\bar{\theta}} (\lambda - \alpha(\theta)) f(\theta)\, d\theta$. I can therefore express $\mu(\theta)$ as

$$\mu(\theta) = -\int_{\underline{\theta}^T}^{\bar{\theta}} (\lambda - \alpha(\theta)) f(\theta)\, d\theta + \int_{\underline{\theta}^T}^{\theta} (\lambda - \alpha(\theta)) f(\theta)\, d\theta = -(1 - F(\theta))(\lambda - \bar{\alpha}_\theta)$$

where $\bar{\alpha}_\theta = \frac{1}{1-F(\theta)} \int_{\theta}^{\bar{\theta}} \alpha(\theta) f(\theta)\, d\theta$, i.e. the average domestic ownership share in the type distribution above θ (and $1 - \bar{\alpha}_\theta$ is the average foreign ownership share).

The first optimality condition (optimizing over $\hat{\pi}$) can now be expressed as

$$\tilde{T}_k(\theta) = \frac{\lambda - \bar{\alpha}_\theta}{\lambda} \frac{1 - F(\theta)}{f(\theta)} \pi_{\theta k} (1 - \kappa')$$

And the second optimality condition (optimizing over $\hat{\pi}$) can now be expressed as

$$T_{\hat{\pi}}(\theta) = \frac{\lambda - \bar{\alpha}_\theta}{\lambda} \frac{1 - F(\theta)}{f(\theta)} \pi_\theta \kappa''$$

Now, consider public input provision. The first order condition is given by

$$\int \left[\lambda \pi_g f(\theta) + \mu(\theta) \pi_{\theta g} (1 - \kappa')\right] d\theta - \lambda c_g = 0$$

from which follows

$$\int \pi_g f(\theta)\, d\theta - c_g = \int \left(\frac{\lambda - \bar{\alpha}_\theta}{\lambda}\right) \left(\frac{1 - F(\theta)}{f(\theta)}\right) \pi_{\theta g} (1 - \kappa') f(\theta)\, d\theta$$

Using

Finally, consider the choice of $\underline{\theta}^T$. Again, the choice of $\underline{\theta}^T$ does not affect the intra-marginal types' optimality conditions for marginal tax rates, i.e. the ICC is not affected. Accordingly, the government chooses $\underline{\theta}^T$ to maximize its objective function, $\int_{\underline{\theta}}^{\underline{\theta}^T} \alpha(\theta)\,\Pi^o d\theta + \int_{\underline{\theta}^T}^{\bar{\theta}} (\alpha(\theta)\,\Pi(\theta) + \lambda T(\theta))\,d\theta - \lambda c(g,n)$. The first-order condition is

$$\int (\lambda - \alpha(\theta))\,\pi_\theta\left(k, g, \underline{\theta}^T\right) f(\theta)\,d\theta - \lambda T\left(\underline{\theta}^T\right) f\left(\underline{\theta}^T\right) + \lambda f(\underline{\theta})\,c_n = 0$$

from which follows the optimality condition in (11). ■

5.3 Proof of Prop. 4

Proof. The Hamiltonian is adjusted to

$$\begin{aligned} H &= [\pi(k,g,\theta) - \kappa(\Delta) - \Pi(\theta)]\,f(\theta) \\ &\quad +\mu(\theta)\,[\pi_\theta(1-\kappa_\Delta) - \kappa_\theta(\Delta,\theta)] + \nu(\theta)\,\Pi(\theta)\,f(\theta) \end{aligned}$$

The government solves $\max_{k(\theta),\Delta(\theta)} H$ for each θ. The first order conditions with respect to k and $\hat{\pi}$ are

$$\begin{aligned} \pi_k f(\theta) + \mu(\theta)\,\pi_{\theta k}(1-\kappa_\Delta) &= 0 \\ -\kappa_\Delta f(\theta) - \mu(\theta)\,[\pi_\theta \kappa_{\Delta\Delta} + \kappa_{\theta\Delta}] &= 0 \end{aligned}$$

The first f.o.c. is not affected. The second f.o.c. can be expressed as

$$T_{\hat{\pi}}(\theta) = \frac{1-F(\theta)}{f(\theta)}\,[\pi_\theta \kappa_{\Delta\Delta} + \kappa_{\theta\Delta}]$$

■

5.4 Sufficient statistics approach

This Appendix derives the optimal tax rate expressions in Sect. 2.7.

As a first step, the relationship between $f(k)$ and the type density $f(\theta)$ is explained. With $F(\theta) = F(k^*(\theta)) = \int_0^{k^*(\theta)} f(k(\theta))\,dk$ follows $F'(\theta) = f(\theta) = f(k^*(\theta))\,k^{*\prime}(\theta)$ where $f(k^*(\theta)) = \int_{-\infty}^{\infty} f(\hat{\pi}|k^*)\,d\hat{\pi}$. Based on the first order condition $\pi_k^\theta = \tilde{T}_k$, $k^{*\prime}(\theta)$ can be expressed as

$$\frac{dk}{d\theta} = -\frac{\pi_{k\theta}^\theta}{\pi_{kk}^\theta - \tilde{T}_{kk}}$$

That is, due to non-linearities in the tax system, the observable density is 'deformed'. As Saez (2001), I will replace this deformed density by a 'virtual' density that corrects for this deformation (by locally 'replacing' the non-linear schedule by a linear schedule for which $\tilde{T}_{kk} = 0$ holds).

Tax revenue is given by

$$\int_0^\infty \left[\int_{-\infty}^\infty T\left(k, \hat{\pi}\left(k\right)\right) f\left(\hat{\pi}|k\right) d\hat{\pi} \right] f\left(k\right) dk$$

First, consider the small increase in $T_{\hat{\pi}}$ denoted as $d\tau_{\hat{\pi}}$ that affects the schedule between $\hat{\pi}^*$ and $\hat{\pi}^* + d\hat{\pi}$. The *mechanical* effect (M) captures the increase in tax revenue for all firms with a reported profit level at or above $\hat{\pi}^*$. Holding behavior constant, each of these firms pays an additional tax of $d\hat{\pi} d\tau_{\hat{\pi}}$. The impact on tax revenue is given by

$$M_{\hat{\pi}} = d\hat{\pi} d\tau_{\hat{\pi}} \int_0^\infty \left[\int_{\hat{\pi}^*}^\infty f\left(\hat{\pi}|k\right) d\hat{\pi} \right] f\left(k\right) dk = d\hat{\pi} d\tau_{\hat{\pi}} \left(1 - F\left(\hat{\pi}^*\right)\right)$$

Now, turn to the elasticity effect (E), that is in generalized form given by

$$E_{\hat{\pi}} = \int_0^\infty \left[T_{\hat{\pi}}\left(k, \hat{\pi}\left(k\right)\right) \frac{d\hat{\pi}\left(k\right)}{d\tau_{\hat{\pi}}} + T_k\left(k, \hat{\pi}\left(k\right)\right) \frac{dk}{d\tau_{\hat{\pi}}} \right] f\left(\hat{\pi}^*|k\right) f\left(k\right) dk \cdot d\hat{\pi}^* d\tau_{\hat{\pi}}$$

A change in T_π holding T_k constant affects (i) reported profit in the range between $\hat{\pi}^*$ and $\hat{\pi}^* + d\hat{\pi}^*$ and (ii) the capital stock in the range between $k\left(\hat{\pi}^*\right)$ and $k\left(\hat{\pi}^* + d\hat{\pi}^*\right)$ (which are the capital stocks of firms that have reported profit levels between $\hat{\pi}^*$ and $\hat{\pi}^* + d\hat{\pi}^*$.

The elasticity effect on reported profit is given by

$$d\hat{\pi} = -\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}} \frac{\hat{\pi}}{1 - T_{\hat{\pi}}} d\left(d\tau_{\hat{\pi}} + dT_{\hat{\pi}}\right)$$

where $\varepsilon_{\hat{\pi},1-T_{\hat{\pi}}}$ is the net-of-tax elasticity of reported profit, and on capital input by

$$dk\left(\hat{\pi}\right) = \varepsilon_{k,c+\tilde{T}_k} \frac{\tilde{T}_k}{c + \tilde{T}_k} \frac{k}{\left(1 - T_{\hat{\pi}}\right)^2} d\left(d\tau_{\hat{\pi}} + dT_{\hat{\pi}}\right)$$

where $\varepsilon_{k,c+\tilde{T}_k}$ is the user cost elasticity of capital (with c denoting the non-tax cost of capital and, thus, $\frac{\tilde{T}_k}{c+\tilde{T}_k}$ the tax wedge as percent of the total user cost).

The term $dT_{\hat{\pi}}$ captures a change in marginal tax rates due to non-linearities, $dT_{\hat{\pi}} = T_{\hat{\pi}\hat{\pi}} d\hat{\pi}$. As Saez (2001), I use an approximization by linearizing the tax schedule, i.e. $T_{\hat{\pi}\hat{\pi}} = 0$. The elasticity effect is thus given by

$$E_{\hat{\pi}} = \int_0^{\infty} \left[-\frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}} \varepsilon_{\hat{\pi},1-T_{\hat{\pi}}} \hat{\pi} + \varepsilon_{k,c+\tilde{T}_k} \frac{\tilde{T}_k^2}{c+\tilde{T}_k} \frac{k}{(1-T_{\hat{\pi}})} \right] f(\hat{\pi}^*|k) f(k) \, dk \cdot d\hat{\pi}^* d\tau_{\hat{\pi}}$$

Note that there is no income effect for all firms above $\hat{\pi}^* + d\hat{\pi}^*$. The optimal tax schedule is then found where $M + E = 0$ which provides the optimal tax rate expression in Sect. 2.7.

Optimal marginal tax rate on capital

Now, consider the mechanical effect for k:

$$M_k = dk^* d\tau_k \int_{k^*}^{\infty} f(k) \, dk = d\hat{\pi} d\tau_{\hat{\pi}} (1 - F(k^*))$$

Now, turn to the elasticity effect (E), that is in generalized form given by

$$E_k = \int_{-\infty}^{\infty} \left[T_{\hat{\pi}} (\hat{\pi}(k^*), k^*) \frac{d\hat{\pi}(k^*)}{d\tau_k} + T_k (\hat{\pi}(k^*), k^*) \frac{dk(\hat{\pi})}{d\tau_k} \right] f(\hat{\pi}|k^*) \, d\hat{\pi} \cdot f(k^*) \cdot dk^* d\tau_k$$

A change in T_k affects the stock of capital in the range between k^* and $k^* + dk^*$. The elasticity effect is $dk = -\frac{\partial k}{\partial \tilde{T}_k} (d\tau_k + dT_k)$. With $dT_k = 0$, see above, we have

$$\begin{aligned} dk(\hat{\pi}) &= \varepsilon_{k,c+\tilde{T}_k} \frac{k}{c+\tilde{T}_k} \frac{1}{1-T_{\hat{\pi}}} (d\tau_k + dT_k) \\ d\hat{\pi} &= \frac{d\hat{\pi}}{d\pi} \varepsilon_{k,c+\tilde{T}_k} \frac{\tilde{T}_k}{c+\tilde{T}_k} \frac{k}{1-T_{\hat{\pi}}} (d\tau_k + dT_k) \end{aligned}$$

The elasticity effect is therefore given by

$$E_k = k^* f(k^*) \cdot dk^* d\tau_k \cdot \frac{\tilde{T}_k}{c+\tilde{T}_k} \int_{-\infty}^{\infty} \left[1 + \frac{T_{\hat{\pi}}}{1-T_{\hat{\pi}}} \frac{d\hat{\pi}}{d\pi} \right] \varepsilon_{k,c+\tilde{T}_k} f(\hat{\pi}|k^*) \, d\hat{\pi}$$

Adding the mechanical and elasticity effects, $M_k + E_k = 0$, yields the optimal tax rate expression in Sect. 2.7.

5.5 Proof of Corollary 2

Proof. The government chooses $T\left(\underline{\theta}^T\right)$, $T_{\hat{\pi}}(\theta)$ and $T_k(\theta)$ for all $\theta \geq \underline{\theta}^T$ and g as to maximize

$$\int_{\underline{\theta}}^{\underline{\theta}^T} \Pi^o f(\theta)\, d\theta + \int_{\underline{\theta}^T}^{\bar{\theta}} \Pi(\theta) f(\theta)\, d\theta$$
$$+\lambda \int_{\underline{\theta}^T}^{\bar{\theta}} \left[T\left(\underline{\theta}^T\right) + \int_{\pi\left(\underline{\theta}^T\right)}^{\hat{\pi}(\theta)} T_{\hat{\pi}}(\theta)\, d\hat{\pi} + \int_{k\left(\underline{\theta}^T\right)}^{k(\theta)} T_k(\theta)\, dk \right] f(\theta)\, d\theta - \lambda c(g,n)$$

In the optimum, changes in any the policy variables has, thus, no first order effect on welfare. Furthermore, note that a change in Π^o has no impact on any decision of firms with types $\theta > \underline{\theta}^T$. Accordingly, a change in Π^o has the following effect on welfare:

$$\int_{\underline{\theta}}^{\underline{\theta}^T} d\Pi^o f(\theta)\, d\theta - \left[T\left(\underline{\theta}^T\right) - c_n \right] \lambda f\left(\underline{\theta}^T\right) \frac{d\underline{\theta}^T}{d\Pi^o}$$

From (17) follows that $T\left(\underline{\theta}^T\right) - c_n = \frac{\lambda-1}{\lambda} \frac{1-F\left(\underline{\theta}^T\right)}{f\left(\underline{\theta}^T\right)} \pi_\theta \left(k, g, \underline{\theta}^T\right) > 0$. Therefore, the second term of the equation above is strictly smaller than $\int_{\underline{\theta}}^{\underline{\theta}^T} d\Pi^o f(\theta)\, d\theta$.
■

5.6 Proof of Prop. 5

Proof. The Hamiltonian is given by

$$\begin{aligned} H &= [\pi(k,g,\theta) - \kappa(\Delta) - \Pi(\theta)]\, \varphi(\Lambda,\theta) \\ &\quad + \mu(\theta)\, \pi_\theta(k,g,\theta)\,(1-\kappa') + \nu(\theta)\, \Pi(\theta)\, \varphi(\Lambda,\theta) \end{aligned}$$

The first order conditions are given by

$$\begin{aligned} \pi_k \varphi(\Lambda,\theta) + \mu(\theta)\, \pi_{\theta k}(1-\kappa') &= 0 \\ -\kappa' \varphi(\Lambda,\theta) - \mu(\theta)\, \pi_\theta \kappa'' &= 0 \end{aligned}$$

as well as

$$\frac{\partial H}{\partial \Pi(\theta)} = \left[-1 + \nu(\theta) + [\pi - \kappa(\Delta) - \Pi(\theta) + \nu(\theta)\Pi(\theta)] \frac{\varphi_\Lambda}{\varphi(\Lambda,\theta)} \right] \varphi(\Lambda,\theta) = -\mu'(\theta)$$

from which follows:

$$\mu'(\theta) = [1 - \nu(\theta) - [\pi - \kappa(\Delta) - \Pi(\theta) + \nu(\theta)\Pi(\theta)]\eta(\theta)]\varphi(\Lambda, \theta)$$

with $\eta(\theta) = \frac{\varphi_\Lambda}{\varphi(\Lambda,\theta)} > 0$ the migration semi-elasticity. For all types but the lowest one, we have

$$\begin{aligned}
\mu'(\theta) &= [1 - [\pi - \kappa(\Delta) - \Pi(\theta)]\eta(\theta)]\varphi(\Lambda, \theta) \\
\mu'(\theta) &= [1 - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)
\end{aligned}$$

and $\mu'(\underline{\theta}) = [1 - \nu(\theta) - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)$. The transversality conditions are $\mu(\underline{\theta}) = 0$ and $\mu(\infty) = 0$. Integrating $\mu'(\theta) = [1 - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)$ over types gives

$$\mu(\theta) = \int_{\underline{\theta}^T}^{\theta} \mu'(\theta)\, d\theta$$

With $\mu'\left(\underline{\theta}^T\right) = [1 - \nu(\theta) - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)$ and $\mu'(\theta) = [1 - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)$ otherwise, $\mu(\theta)$ can be expressed as $\mu(\theta) = -\nu(\underline{\theta})\varphi\left(\Lambda, \underline{\theta}^T\right) + \int_{\underline{\theta}^T}^{\theta} [1 - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)\, d\theta$. Since $\mu(\bar{\theta}) = 0$, it must be that $\nu(\underline{\theta})\varphi(\Lambda, \underline{\theta}) = \int_{\underline{\theta}}^{\bar{\theta}} [1 - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)\, d\theta$. It follows

$$\mu(\theta) = -\int_{\theta}^{\bar{\theta}} [1 - T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)\, d\theta = -(1 - \Phi(\Lambda, \theta))\left(1 - \overline{T\eta}^{\theta}\right)$$

where $\overline{T\eta}^{\theta} = \frac{1}{1-\Phi(\Lambda,\theta)} \int_{\theta}^{\bar{\theta}} [T(\theta)\eta(\theta)]\varphi(\Lambda, \theta)\, d\theta$ is the average tax-revenue-weighted migration semi-elasticity for types above θ.

The derivation of optimal tax rates follows the proof of Prop. 2 with an adjusted $\mu(\theta)$ and density term ($\varphi(\Lambda, \theta)$ instead of $f(\theta)$):

$$\begin{aligned}
\tilde{T}_k(\theta) &= \frac{1 - \Phi(\Lambda, \theta)}{\varphi(\Lambda, \theta)} \pi_{\theta k}(1 - \kappa')\left(1 - \overline{T\eta}^{\theta}\right) \\
T_{\hat{\pi}}(\theta) &= \frac{1 - \Phi(\Lambda, \theta)}{\varphi(\Lambda, \theta)} \pi_{\theta}\kappa''\left(1 - \overline{T\eta}^{\theta}\right)
\end{aligned}$$

Now, turn to the provision of public input goods. The government sets g to maximizes $\int_{\underline{\theta}}^{\bar{\theta}} H d\theta - c(g)$. The first order condition with respect to g is

$$\int (\pi_g \varphi(\Lambda, \theta) + \mu(\theta)\pi_{\theta g}(1 - \kappa'))\, d\theta - c_g = 0$$

from which follows

$$\int \pi_g \varphi(\Lambda,\theta)\, d\theta - c' = \int \left[\frac{1-\Phi(\Lambda,\theta)}{\varphi(\Lambda,\theta)} \pi_{\theta g}(1-\kappa')\left(1-\overline{T\eta}^{\theta}\right)\right]\varphi(\Lambda,\theta)\, d\theta$$

Finally, consider the choice of $\underline{\theta}^T$. For given marginal tax rates on profit and capital (both unaffected by the lump-sum tax due to the absence of income effects on firm decisions) as well as a given level of public input goods, g, tax revenue is given by $\int \left[\pi\left(k,g,\underline{\theta}^T\right) - \tilde{\Pi}^o + \int_{\pi\left(\underline{\theta}^T\right)}^{\hat{\pi}(\theta)} T_{\hat{\pi}}(\theta)\, d\hat{\pi} + \int_{k\left(\underline{\theta}^T\right)}^{k(\theta)} T_k(\theta)\, dk\right]\varphi(\Lambda,\theta)\, d\theta$ $\tilde{\Pi}^o$ denotes the best available option. Optimizing over $\underline{\theta}^T$ gives the optimality condition in (11).

$$\begin{aligned} 0 &= -\left(\pi\left(k,g,\underline{\theta}^T\right) - \tilde{\Pi}^o\right)\varphi\left(\Lambda,\underline{\theta}^T\right) + \pi_\theta\left(k,g,\underline{\theta}^T\right)\int \varphi(.,\theta)\, d\theta \\ &\quad -\pi_\theta\left(k,g,\underline{\theta}^T\right)\int T(\theta)\,\varphi_\Lambda(\Lambda,\theta)\, d\theta + \varphi\left(\Lambda,\underline{\theta}^T\right) c_n \end{aligned}$$

from which follows

$$\pi\left(\underline{\theta}^T\right) - \tilde{\Pi}^o = \pi_\theta\left(k,g,\underline{\theta}^T\right)\frac{1-\Phi\left(\Lambda,\underline{\theta}^T\right)}{\varphi\left(\Lambda,\underline{\theta}^T\right)}\left(1-\overline{T\eta}^{\underline{\theta}^T}\right) + c_n(g,n)$$

using $\overline{T\eta}^{\theta} = \frac{1}{1-\Phi(\Lambda,\theta)}\int_\theta^\infty [T(\theta)\,\eta(\theta)]\,\varphi(\Lambda,\theta)\, d\theta$. ■

References

[1] Atkinson, A.B., Stiglitz, J.E. (1976). The design of tax structure: direct versus indirect taxation. Journal of Public Economics 6 (1–2): 55–75.

[2] Baldwin, R. E., Krugman, P. (2004). Agglomeration, integration and tax harmonisation, European Economic Review, 48(1): 1–23.

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

[4] Becker, J., Fuest, C. (2011). Optimal tax policy when firms are internationally mobile. International Tax and Public Finance, 18, 580–604.

[5] Becker, J. and Schneider, A. (2017). Taxation of Firms with Unknown Mobility, Journal of Public Economic Theory 20, 202–217.

[6] Becker, J. and Schneider, A. (2019). Bidding for firms with unknown characteristics, Scandinavian Journal of Economics 121(3): 1222-1243.

[7] Beer, S., de Mooij, R., Liu, L. (2020). International Corporate Tax Avoidance: A Review of the Channels, Magnitudes, and Blind Spots, Journal of Economic Surveys 34(3): 660-88.

[8] Bernheim, B.D., Whinston, M.D. (1986). Common Agency, Econometrica 54(4): 923-942.

[9] Biglaiser, G., Mezzetti, C. (1993). Principals Competing for an Agent in the Presence of Adverse Selection and Moral Hazard, Journal of Economic Theory 61: 302–330.

[10] Biglaiser, G., Mezzetti, C. (2000), Incentive Auctions and Information Revelation, RAND Journal of Economics 31: 145–164.

[11] Bilicka, K. (2019). Multinationals' Profit Response to Tax Differentials: Effect Size and Shifting Channels, American Economic Review 109(8): 2921–2953.

[12] Black, D. A., Hoyt, W. H. (1989). Bidding for firms. American Economic Review, 79(5): 1249–1256.

[13] Bond, E. W., Gresik, T.A. (1996). Regulation of Multinational Firms with Two Active Governments: A Common Agency Approach, Journal of Public Economics 59: 33-53.

[14] Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives, Econometrica 54: 607–22.

[15] Davies, R., J. Martin, M. Parenti, Toubal, F. (2018). Knocking on tax haven's door: Multinational firms and transfer pricing, Review of Economics and Statistics, forthcoming.

[16] Devereux, M. P. and R. Griffith (1998). Taxes and the location of production: evidence from a panel of US multinationals, Journal of Public Economics 68(3): 335-367.

[17] Dharmapala, Dhammika (2014). What Do We Know about Base Erosion and Profit Shifting? A Review of the Empirical Literature. Fiscal Studies 35(4): 421-448.

[18] Diamond, P.A. and J.A. Mirrlees (1971). Optimal taxation and public production, American Economic Review 61: 8-27 and 261-278.

[19] Gordon, R. H. (1986). Taxation of investment and savings in a world-economy. American Economic Review, 76(5): 1086–1102.

[20] Graetz, M., O'Hear, M. (1997). The "original intent" of U. S. international taxation, Duke Law Journal 46(5): 1022–109.

[21] Graetz, M. (2001). Taxing International Income: Inadequate Principles, Outdated Concepts, and Unsatisfactory Policies, Brooklyn Journal of International Law 26: 1357-1449.

[22] Gresik, T.A. (2001). The Taxing Task of Taxing Transnationals, Journal of Economic Literature 39: 800-838.

[23] Gross, T., Klein, P., Makris, M. (forthcoming) Residence- and source-based capital taxation in open economies with infinitely-lived consumers, Journal of International Economics.

[24] Guvenen, F., Mataloni, R., Rassier, D., Ruhl, K. (2019). Offshore Profit Shifting and Domestic Productivity Improvement, NBER Working paper 23324.

[25] Haufler, A. and G. Schjelderup (2000). Corporate tax systems and cross country profit shifting, Oxford Economic Papers 52: 306-325.

[26] Heckemeyer, J., Overesch, M. (2017). Multinationals' Profit Response to Tax Differentials: Effect Size and Shifting Channels, Canadian Journal of Economics/Revue Canadienne d'économique 50(4): 965–94.

[27] Hines Jr, J.R., and E.M. Rice (1994). Fiscal paradise: foreign tax havens and American business, Quarterly Journal of Economics 109: 149-182.

[28] Judd, K. (1985), Redistributive taxation in a simple perfect foresight model, Journal of Public Economics 28: 59–83.

[29] Keen, M., and K. Konrad (2013). The Theory of International Tax Competition and Coordination,in Alan J. Auerbach, Raj Chetty, Martin Feldstein und Emmanuel Saez (eds.) Handbook of Public Economics, Volume 5, Amsterdam & Oxford: Elsevier, 257-328.

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

[31] Keen, M., Piekkola, H. (1997). Simple rules for the optimal taxation of international capital income. Scandinavian Journal of Economics 99: 447–461.

[32] Keen, M., Wildasin, D. (2004). Pareto-efficient international taxation. American Economic Review, 94(1): 259–275.

[33] Lehmann, E., L. Simula und A. Trannoy (2014). Tax me if you can! Optimal nonlinear income tax between competing governments, Quarterly Journal of Economics 129(4): 1995–2030.

[34] Lockwood, B.B., Weinzierl, M. (2015). De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution, Journal of Public Economics 124: 74–80.

[35] Martimort, D., Stole, L. (2002). The Revelation and Delegation Principles in Common Agency Games, Econometrica 70: 1659–1673.

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

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

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

[39] Mirrlees, J.A. (1971). An exploration in the theory of optimum income taxation, Review of Economic Studies 38(2): 175–208.

[40] Olsen, T. E., Osmundsen, P. (2001). Strategic Tax Competition; Implications of National Ownership, Journal of Public Economics 81: 253–277.

[41] Olsen,T. E., Osmundsen, P. (2003). Spillovers and International Competition for Investments, Journal of International Economics 59: 211–238.

[42] Osmundsen, P., Hagen, K. P., Schjelderup, G. (1998). Internationally Mobile Firms and Tax Policy, Journal of International Economics 45: 97–113.

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

[44] Riedel, N. (2018). Quantifying International Tax Avoidance: A Review of Academic Literature, Review of Economics 69(2): 169-181.

[45] Rochet, J.-C. (1986). Le Contrôle des Equations aux Dérivées Partielles Issues des la Theéorie des Incitations, Thesis Université Paris IX.

[46] Saez, E. (2001), Using elasticities to derive optimal income tax rates, Review of Economic Studies 68: 205–29.

[47] Sinn, H.-W. (1990). Tax harmonization and tax competition in Europe, European Economic Review 34(2/3): 489-50.

[48] Tiebout, C. (1956). A Pure Theory of Local Expenditures, Journal of Political Economy 64(5): 416–424.

[49] Tørsløv, T., Wier, L., Zucman, G. (2021). The Missing Profits of Nations, Working Paper.

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

[51] 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.