RATIONALIZING FORMULA APPORTIONMENT

JOHANNES BECKER

WORKING PAPER
04/2023
Rationalizing formula apportionment*

Johannes Becker (U Münster)

February 15, 2024

Abstract

Many federal tax systems use formula apportionment to assign the taxable profit of large businesses to the federal subunits where typically the subunits’ own tax rates are applied. The formulas (including the one recently proposed by the EU Commission as well as the one agreed upon by the OECD/G20 Inclusive Framework) are remarkably similar, yet little is known how to rationalize them, i.e. from which normative criteria they are derived. To fill this gap, I take an axiomatic approach and derive a generalized system of formula apportionment from three criteria: fixed coverage, positive responsiveness and external independence. I demonstrate that any formula apportionment system that satisfies the three criteria suffers from the same distortion (unless all local tax rates are identical). The generalized system comprises existing real-world systems (as observed in federations like the US, Canada and Germany) as special cases, but allows for a degree of flexibility that has the (so far unrealized) potential for surplus increasing reforms.

JEL classification: H25, H71, F23

Keywords: Corporate taxation, business taxation, federalism, formula apportionment

*I thank Reuven Avi-Yonah, Kristoffer Berg, James Hines jr., Rick Krever, Marco Runkel, Dirk Schindler, Esteban Muñoz Sobrado and participants at the 2022 MATAX conference as well as the 2023 Oxford University Centre for Business Taxation Symposium for helpful comments. The usual disclaimer applies.
1 Introduction

In October 2021, the OECD/G20 Inclusive Framework agreed on a “two-pillar solution to address the tax challenges arising from the digitalization of the economy” (OECD 2021). Pillar One of the agreement implies a reallocation of taxing rights with regard to the profit of multinational firms and includes an element of formula apportionment, i.e. part of the profit is to be allocated in proportion to the multinational firm’s sales in each country.\footnote{The Pillar One agreement stipulates that very large firms use formula apportionment for part (25 percent) of their ‘residual’ profit (i.e. profit that exceeds a certain return).} In 2023, the EU Commission launched the BEFIT\footnote{BEFIT is an acronym for “Business in Europe: Framework for Income Taxation”.} proposal including – among others – a renewal of its 2011 plan to introduce formula apportionment as a means of allocating multinational firm profits within the EU.\footnote{The original 2011 proposal by the Commission for a common consolidated corporate tax base (CCCTB) with formula apportionment for large multinational firms in the EU failed to gain sufficient political support.}

Formula apportionment is a method of allocating tax base across jurisdictions whenever a firm is active in more than one jurisdiction and at least two jurisdictions claim the right to tax the firm’s income ‘at source’.\footnote{To be precise, the level of activity usually has to surpass a certain minimum threshold. Then, the firm is said to have ‘nexus’ to the jurisdiction under consideration.} In federations like the US, Canada and Germany, where the federal subunits (i.e. US States, Canadian provinces, German municipalities) have the right to tax business income, total business income within the federation is consolidated and then allocated to the subunits based on a formula. Then, each of the subunits applies its own tax rate to its share in the total tax base. The formula is typically based on tangible capital (or property), payroll (sometimes number of employees) and/or sales. In contrast, the predominant international method to allocate multinational firms’ income across countries is to tax each of the firm’s affiliates separately based on the arm’s length principle (\textit{separate accounting}).

Fueled by growing discontent with the international tax system (specifically with regard to mounting evidence of multinational firms engaging in base erosion and profit shifting), there have been initiatives to reform the international system towards a formula-based system of tax base allocation, e.g. the above mentioned OECD/G20 agreement and the EU Commission’s BEFIT proposal.

The apportionment formulas used in real-world federations and proposed in the EU Commission’s plans as well as in the Two-Pillar Reform Agreement are of an intriguingly similar shape. Formulas that use only one indicator are of the
Formulas that use more than one indicator and allocate $b\cdot 100$ percent according to a specific indicator take the form: “allocate $bx$ percent of the tax base where $x$ percent of the indicator is”. The indicators used in the formula and their weights differ, but the general shape of the formula is remarkably consistent across systems.

In this paper, I seek to rationalize the general formula shape. I argue that, while the existing formulas may have been introduced in an ad-hoc manner, the consistency in shape (and the fact that similar formulas have been introduced simultaneously in different parts of the world) suggests that it appeals to some common sense. The paper sets out to, first, make this common sense explicit, i.e. it asks for the implicit normative criteria from which the formula can be derived. It thus provides theoretical studies with a proper foundation for their analyses; instead of motivating a study of formula apportionment just with “because it exists in real-world tax systems”, I provide the motivation “because it is an optimal solution to a specific problem”. Second, the paper explores the potential of a generalized formula derived from these common sense principles. It proposes an axiomatic foundation of a class of apportionment systems which satisfy a number of desirable features – and comprises most (if not all) of the existing formulas. In other words, it demonstrates that, if the tax base allocation system is supposed to adhere to these desirable features, the formula is close to those observed in real-world tax systems. In fact, if the formula is based on only one indicator (e.g. sales), the formulas observed (assigning tax base to a region in proportion to its share in sales) is the only one that satisfies these features. Third, the paper shows that, for two or more indicators used in the formula, the generalized formula allows for more flexibility that, under certain circumstances, improves the efficiency of the tax system (albeit to a limited extent).

To be specific, I consider the class of apportionment systems that are characterized by (1) fixed coverage, (2) positive responsiveness and (3) external independence. **Fixed coverage** means that the fraction of the total tax base to be allocated is fixed. This includes the case in which all of the tax base is allocated according to formula (full coverage). **Positive responsiveness** means that an increase in the firm’s activity (be it input use, sales or the location of firm functions) in a given region weakly increases the tax base fraction allocated to this region, whereas an increase in activity outside the region weakly decreases it. **External independence** means that the tax base fraction assigned
to a given location is insensitive to where exactly the activity outside of this location is happening. E.g., if tangible assets are shifted from one location to the other (without affecting the firm’s total stock of assets), both of them outside of location $i$, the tax base fraction assigned to $i$ stays constant.

I derive a general formula for tax base apportionment that satisfies these three properties. The formula contains the existing formulas as special cases. I demonstrate that such systems generally suffer from two potential kinds of biases. First, with fixed coverage and strict positive responsiveness, a re-allocation of activity from low-taxed to high-taxed locations within the firm increases total surplus (i.e. economic pre-tax profit). I show that this dispersion bias cannot be avoided. Second, depending on the firm-specific geographical allocation of activity, the aggregate activity level within the firm may be too low (or too high) due to the incentives provided by the apportionment formula. That is, an increase (or a decrease) in activity in all locations of the firm would increase surplus. I demonstrate that, under certain conditions, a general form of the formula (derived from the three properties above) has the potential to mitigate the distortion of the aggregate activity level. This effect cannot be attained by using the simplified type of formula used so far in real world tax systems.

As pointed out above, formula apportionment is only one way to solve the problem of profit allocation across jurisdictions for tax purposes. Who invented this method? Krever and Mellor (2020) report that formula apportionment simultaneously emerged at the end of the 19th Century in several locations (like in the German State of Baden, in France and in Australia and in several US States).

“The intuitive solution adopted almost universally for the earliest income and predecessor taxes when an enterprise operated across borders was to allocate a share of the profits to each jurisdiction that contributed to the total profits by reference to formulas incorporating the factors that were presumed to contribute to the generation of those profits.” (Krever and Mellor 2020, p. 9)

In the US, individual States unilaterally applied this “intuitive solution” to specific industrial sectors like railways (where the track length in the state provided the indicator for the railway company’s nexus and tax liability) and telegraphs. In fact, as Krever and Mellor (2020) report, the very early formulas were used to calculate the property taxes that out-of-state businesses owed to the state. Later on, the same formula-based method was adopted for levying
the ‘capital tax’ that effectively required assigning a share of the distributed profit to each state.

Over time, formula apportionment became the preferred method of profit allocation in federations like the US and Australia (between states), Canada (between provinces), Germany (between communities) and others. In contrast, the international system of allocating profit across countries did not adopt formula apportionment but established a system with separate accounting (based on the arm’s length principle and governed by double tax treaties, see Weiner 1999).

The coexistence of two distinct systems of tax base allocation has kept alive the debate on potential reforms, specifically of the international tax system. Each of the two systems has its flaws. Tax base allocation based on formulas is likely to distort the firm’s decision where to invest and to hire labor, see McLure (1981) as well as Gordon and Wilson (1986). In contrast, the current international system of business taxation is prone to base erosion and profit shifting, as a large literature starting with Hines and Rice (1994) shows, and may have revenue implication in violation of inter-nation equity. Moreover, the growing awareness of tax competition between the EU member led to a number of studies that considered the incentives for (non-cooperative) tax policy under competition under the two alternative systems, separate accounting and formula apportionment, with some studies focusing on the choice of the formula weights.

---

5The literature on this topic is large, see Hellerstein and McLure (2004), Weiner (2005), Avi-Yonah andClausing (2008), Altshuler and Grubert (2009), Bettendorf et al. (2010), Clausing (2016) for more recent contributions.


7See Dharmapala (2014) and Riedel (2018) for literature reviews.

8In the early 1970s, Musgrave and Musgrave (1972) pointed out major disadvantages of the separate accounting for international profit allocation specifically for developing countries and promoted unitary taxation with formula apportionment as an alternative. These arguments have been adopted nowadays by tax justice activists.

9See e.g. Pethig and Wagener (2003), Eggert and Schjelderup (2003), Nielsen et al. (2003, 2010), Riedel and Runkel (2010), Gresik (2010) and Becker and Fuest (2010). Mintz and Smart (2004) show that using formula apportionment reduces the tax base elasticity in Canadian provinces, as compared to a system with separate accounting.

10Eichner and Runkel (2008) argue in favor of local sales as a formula factor. Runkel and Schjelderup (2011) discuss the choice between capital, labor and sales and show that the
An obvious candidate for adopting a (US style) system with formula apportionment is the EU (or its predecessor, respectively), as observers like McLure (1989) and Daly and Weiner (1993) have noted early on. In 2004, the EU Commission proposed for a Common Consolidated Corporate Tax Base (CCCTB) with formula apportionment.\textsuperscript{11} So far, the proposal failed to gain the necessary political support among EU Member States, but it was nevertheless renewed in the 2023 BEFIT proposal. Recently, the OECD/G20 Inclusive Framework agreed on a fundamental reform of the international system that introduces formula apportionment of profits according to sales (Navarro 2021), however only for large multinational firms and only with regard to part of their profit.

To the best of my knowledge, all of the above mentioned studies and literature branches have in common that they take the general formula as given, i.e. they accept its general shape as an exogenously given policy constraint.\textsuperscript{12} Given its shape, formula’s weights are varied and its effects are criticized for causing inefficiency cost. However, with regard to its general form, the formula itself is never rationalized nor put into question. This paper sets out to fill this gap.

The remainder of the paper is organized as follows. The following section provides the model analysis. Section 3 concludes.

2 Model analysis

Consider a federation with $I$ regions and a multiregional firm, which is active in all $I$ regions. Each region levies a linear profit tax at source with a rate of $t_i$ with $i = 1, \ldots, I$. The total tax base is the firm’s consolidated pre-tax profit $\pi$.\textsuperscript{13} Region $i$’s share in the total tax base is denoted as $b_i$ and the vector of regional tax base fractions is $b = (b_1, \ldots, b_I)$, which can be understood as the federation’s apportionment system. For instance, in a federation with two regions, the apportionment system $b = \left( \frac{1}{3}, \frac{2}{3} \right)$ implies that one third of the tax

\textsuperscript{11}See e.g. Fuest et al. (2007), Devereux and Loretz (2008), Cobham and Loretz (2014), de Mooij et al. (2021) for impact assessments and tax revenue estimations.

\textsuperscript{12}In some rare cases, alternative shapes are discussed, e.g. Gordon and Wilson (1986), without reference to deeper-level principles, though. I will discuss two of these alternatives below.

\textsuperscript{13}I ignore any differences between economic profit and accounting profit.
base is allocated to region 1 and the remainder to region 2.

Note that, so far, I did not impose any restrictions on $b_i$. Specifically, it may be that the sum $\sum b_i$ is smaller or larger than one (e.g. it is possible that only part of the tax base is allocated to the regions and the remainder is taxed by the federal government). Let $\tilde{T} = T + T^F$ denote the total effective (average) tax rate on firm profits, which is the weighted sum of the effective tax on allocated profit $T$, equal to

$$T = \sum b_i t_i$$

and the tax on non-allocated profit $T^F = (1 - \sum b_i)t^F$, possibly levied by the federal government, with $t^F$ denoting the associated tax rate.

The firm’s consolidated pre-tax profit $\pi$ depends on a range of choices made by the firm. For instance, the firm chooses a level of tangible assets and employment in each region (input choice), it decides how much of the total product to sell in a given region (sales shares) and where to locate the marketing department, the risk management etc. (firm functions). I summarize these choices under the label activities $x$ that have an impact on total profit, i.e. $\pi = \pi(x)$.\textsuperscript{14}

Let $a = 1, ..., A$ be the index for the activities and $x_i = (x_{i1}, ..., x_{iA})$ the vector of the $A$ activities in region $i$. Then, $x = (x_1, ..., x_I)$ denotes the vector of activities in regions $i = 1, ..., I$. Although some of the activity considered here imply discrete choices (e.g. the marketing department being located either in $i$ or in $j$, but not split-up and located in both), I will treat all activities as continuous variables.\textsuperscript{15}

Activities are called bound, if their level is naturally limited. For instance, a firm cannot sell more than 100 percent of its product; i.e. when the firm chooses the fraction of total sales for a given region $i$, an increase of this share requires a reduction in some other region by the same amount. Similarly, some firm functions (e.g. headquarters services) may be considered a constant for a given level of output. An example for an unbound activity is employment. An increase in employment in one location does not necessitate a reduction in employment somewhere else (although it may affect the return to employment in other locations). Any activity that can be measured and aggregated across regions can, in principle, be used for purpose of allocation of the tax base.

\textsuperscript{14}In the Appendix, I present a slightly more specific model with explicit expressions for inputs, sales shares and functions.

\textsuperscript{15}Adapting the model to account for discrete choices is straightforward, but involves some notational effort. In any way, it does not alter the results in important ways.
With this in mind, the optimization problem of the firm is given by
\[
\max_{x} \pi(x) \left(1 - \tilde{T}\right) \quad \text{s.t.} \quad \sum x_{ia} = \bar{x}_a \quad \text{for all} \quad a \in \bar{A} \tag{2}
\]
where \(\bar{A}\) is the set of bound activities. The first order conditions for the \(x_{ia}\) are
\[
\pi_{x_{ia}} \left(1 - \tilde{T}\right) - \pi \frac{dT}{dx_{ia}} + \lambda^{x_{ia}} \geq 0 \quad \forall i, a \tag{3}
\]
where \(\lambda^{x_{ia}}\) is the Lagrangian parameter with \(\lambda^{x_{ia}} \neq 0\) if \(a \in \bar{A}\) and \(\lambda^{x_{ia}} = 0\) if \(a \notin \bar{A}\). Moreover, \(\frac{dT}{dx_{ia}} = \frac{dT}{dx_{ia}} + \frac{dT}{dx_{ia}}\) with \(\frac{dT}{dx_{ia}} = \sum_j \frac{db_j}{dx_{ia}} t_j\). The first order condition holds with a strict less-than sign, if no activity \(a\) in region \(i\) is the optimal choice. It holds with a strict greater-than sign, of all of activity \(a\) is optimally located in \(i\), which may happen with bound activities. An inner solution emerges, if the first order condition holds with a strict equality sign.

The following discussion refers to a setting with inner solutions.

If \(\frac{dT}{dx_{ia}} = 0\), the firm’s activity choice is unbiased, i.e. the optimum condition is \(\pi_{x_{ia}} = 0\) for an unbound activity, and \(\pi_{x_{ia'}} = \pi_{x_{ia''}}\) for any bound activities \(a', a''\). In contrast, if the tax payment \(\tilde{T}\) responds to a marginal increase in activity in \(i\), there is a behavioral bias due to the tax system. The magnitude and the sign of this bias depends on the apportionment system chosen by the federation. In the following, I will assume that \(T^F = 0\) and, thus, \(\tilde{T} = T\), for simplicity.

I now introduce the first two formula properties, in order to characterize the formula effects on firm behavior.

\textbf{Properties 1} (i) Fixed coverage: A fixed fraction \(\bar{b} \in (0, 1]\) of total profit \(\pi\) is allocated to the regions:
\[
\sum_i b_i = \bar{b} \tag{4}
\]
with \(\bar{b} = 1\) implying full coverage.

(ii) Positive responsiveness:
\[
\frac{\partial b_i}{\partial x_{ia}} \geq 0 \quad \text{and} \quad \frac{\partial b_i}{\partial x_{-ia}} \leq 0 \tag{5}
\]
with the stronger form (strictly positive responsiveness) requiring \(\frac{\partial b_i}{\partial x_{ia}} > 0\) and \(\frac{\partial b_i}{\partial x_{-ia}} < 0\).

With fixed coverage, the total fraction of tax base allocated to the regions
does not respond to an increase in activity in some region \( i \), i.e. \( \sum_{j \neq i} \frac{db_{ij}}{dx_{ia}} = 0 \) and, thus, \( \frac{db_{i}}{dx_{ia}} = - \sum_{j \neq i} \frac{db_{ij}}{dx_{ia}} \). Using this, I can write the above first order condition (3) as follows

\[
\pi_{x_{ia}} + \frac{\lambda x_{ia}}{1 - T} = \bar{b} \left( t_{i} - \tilde{t}_{-i} \right) \frac{db_{i}}{dx_{ia}} \frac{\pi}{1 - T} \tag{6}
\]

where \( \tilde{t}_{-i} = \sum_{j \neq i} \sigma_{ij} t_{j} \) is a weighted average of the tax rates outside of \( i \) with \( \sum_{j \neq i} \sigma_{ij} = 1 \) and \( \sigma_{ij} = - \frac{db_{ij}}{dx_{ia}} / \frac{db_{i}}{dx_{ia}} \). The above equation (6) is the equivalent to eq. (4) in Gordon and Wilson (1986).

In what follows, I will use the term distorted or biased, if an activity is higher or lower than its Pareto efficient level for a given level of all other activities. In order to make statements on Pareto efficiency, I assume that \( \pi (x) \) is the total surplus.\(^{16}\)

**Proposition 1** With fixed coverage and strictly positive responsiveness, the choice of activity \( x_{ia} \) is biased downwards if \( t_{i} > \tilde{t}_{-i} \) and upwards if \( t_{i} < \tilde{t}_{-i} \).

**Proof.** Consider a small increase of an unbound activity \( a_{i} \). The effect on total surplus is \( \pi_{x_{ia}} dx_{ia} \) which equals \( T_{a_{i}} \frac{\pi}{1 - T} dx_{ia} \), which is equal to the right hand side of the first order condition in eq. (6). With strictly positive responsiveness and \( t_{i} > \tilde{t}_{-i} \), the effect is unambiguously positive. Now, consider a small increase of a bound activity \( a_{i} \) and an associated decrease of \( a_{j} \) with \( j \neq i \). Furthermore, let \( t_{i} > \tilde{t}_{-i} \) and \( t_{j} < \tilde{t}_{-j} \). With \( dx_{ia} = - dx_{ja} \), the effect on total surplus is \( (T_{a_{i}} - T_{a_{j}}) \frac{\pi}{1 - T} dx_{ia} \) which equals \( \left( t_{i} - \tilde{t}_{-i} \right) \frac{db_{i}}{dx_{ia}} - \left( t_{j} - \tilde{t}_{-j} \right) \frac{db_{j}}{dx_{ja}} \). Thus, eq. (6) implies that activity \( a \) in region \( i \) is distorted downwards if \( t_{i} > \tilde{t}_{-i} \) and vice versa.

The above Proposition generalizes a notion that occurs repeatedly in the literature (e.g. Gordon and Wilson 1986): Formula apportionment has an efficiency cost. The firm has an incentive to shift activity to locations with below ‘average’ (i.e. below \( \tilde{t}_{-i} \)) tax rates. This incentive is closely linked to the formula itself, and the shape and the weights in the formula affect this incentive. However, the above Proposition tells us that, independent of its specific shape (which I have not specified yet), any formula characterized by fixed coverage and strictly positive responsiveness causes a behavioral distortion that cannot be avoided (unless all local tax rates are identical).

\(^{16}\)This is the case if one assumes that input remuneration just covers the opportunity cost of input providers and that there are no other externalities.
It is helpful to differentiate two aspects of the tax distortion. First, efficient production requires that the tax-inclusive marginal cost of activity (e.g. the cost of capital) is equal in all regions. However, with different tax rates across regions, strictly positive responsiveness implies dispersion of marginal tax cost (the right hand side of eq. (6)), captured by the term \((t_i - \bar{t})\) in the above formula. Second, for unbound activities, \(\lambda x = 0\), the aggregate level of activity may be too low (or too high). Note that, if the tax rates are the same across locations, the dispersion of tax rates is zero and, following eq. (6), the tax on the marginal unit of activity is zero in each region. In this case, the aggregate activity is at its efficient level as well. This implies that, in the absence of a dispersion distortion, there cannot be a distortion of the aggregate activity level. However, in the presence of dispersion, the aggregate activity level may be too high or too low – in the following sense: An increase in the marginal tax on activity in each region (i.e. a uniform increase of the right hand side of eq. 6 for all \(i\)), increases or decreases total pre-tax profit.

The following analysis will yield that a generalized formula may mitigate the aggregate level distortion, but not the dispersion distortion. Before, I illustrate both distortions with an example.

**Example 1**

Consider a simple apportionment formula based on assets \(k_i\) and employment \(l_i\) (both of which are unbound activities). Region \(i\)'s share of the tax base is given by

\[
b_i = \bar{z} \frac{k_i}{K} + (1 - \bar{z}) \frac{l_i}{L} \quad \text{with} \quad (7)
\]

where \(\bar{z} \in [0, 1]\) is a fixed weight, \(K = \sum_i k_i\) and \(L = \sum_i l_i\).

After tax profit is \(\pi (k, l) (1 - T) - \tau K\) where \(\tau\) is set by the federal government. The latter’s sole purpose is to analyze the aggregate level distortion. The first order condition with respect to a regional capital stock \(k_i\) can be expressed as

\[
\pi_{k_i} = (t_i - \bar{t}) \frac{\bar{z}}{K} \frac{\pi}{1 - T} + \frac{\tau}{1 - T}
\]

where \(\bar{t} = \sum_j b_j \frac{k_j}{K}\) denotes the capital-weighted average tax rate.

The distortion due to dispersion is obvious. The difference between the marginal tax cost levels of the two locations with the highest and lowest local tax rates, \(t_{\max}\) and \(t_{\min}\), is \((t_{\max} - t_{\min}) \frac{\bar{z}}{K} \frac{\pi}{1 - T}\). The distortion of the aggregate level of capital inputs can be illustrated by considering an increase of \(\tau\) starting...
from $\tau = 0$. For simplicity, assume that production is separable in labor and capital. Then, with $\sum \frac{dk_i}{d\tau} = \frac{dK}{d\tau}$, an increase in $\tau$ has the following effect on pretax profit $\pi$:

$$\sum \pi_k \frac{dk_i}{d\tau} d\tau = \frac{\bar{z}}{K} \frac{\pi}{1 - T} d\tau \sum \left( (t_i - \bar{t}^K) \frac{dk_i}{d\tau} \right) + \frac{\tau}{1 - T} \frac{dK}{d\tau} d\tau$$

(9)

Starting at $\tau = 0$, the effect of a small increase in $\tau$ on pretax profit is

$$\frac{\bar{z}}{K} \frac{\pi}{1 - T} (\bar{t}^\Delta - \bar{t}^K) \frac{dK}{d\tau} d\tau$$

(10)

where $\bar{t}^\Delta = \sum t_i \frac{dk_i}{d\tau}$ is an average tax rate weighted by the effects of $\tau$ on local capital stocks.

For the sake of the argument, consider the case in which the $dk_i/d\tau$ are equal for all $i$ (which may be the case with quadratic production functions). Then, $\bar{t}^\Delta$ equals the average tax rate $\bar{t}$ and the above eq. (10) can be interpreted as follows. If the $k_i$ are larger in low-tax regions, the average tax rate is larger than the capital-weighted one, $\bar{t} > \bar{t}^K$. In this case, an increase in $\tau$ decreases the firm’s pretax profit since $dK/d\tau < 0$, implying that the aggregate level of capital inputs is too low. Similarly, if the firm’s capital stocks are evenly distributed across regions and the $dk_i/d\tau$ are larger (i.e. less negative) for low-tax regions, the expression on the right hand side of eq. 9 is negative, which – again – means that the aggregate level of capital inputs is too low.

These admittedly stylized examples show that there may be an aggregate level distortion that needs to be distinguished from the dispersion distortion.

A generalized apportionment formula

I will restrict my attention to apportionment formulas characterized by fixed coverage, positive responsiveness and a third property:

**Properties 2** (iii) External independence: With constant total activity, $b_i$ does not depend on in which specific region $j \neq i$ activity $x_j$ is located.

Let $x_{-i}$ denote the vector of all $x_j$ outside of region $i$, i.e. $j \neq i$. That is, $x = (x_i, x_{-i})$. Then, for a given $x_i$ and any two $x_{-i}$ and $x'_{-i}$, external independence implies

$$b_i(x_i, x_{-i}) = b_i(x_i, x'_{-i}) \text{ if the } X_{-i} \text{ remain constant}$$

(11)
with \( \mathbf{X}_{-i} = (X_{-i1}, \ldots, X_{-iA}) \) and \( X_{-ia} = \sum_{j \neq i} x_{ja} \), with the latter denoting the aggregate activity \( a \) outside of \( i \), i.e., \( X_a - x_{ia} \).

External independence implies that, if some activity outside of \( i \) is shifted to another place outside of \( i \), the tax base fraction allocated to \( i \) does not change. For instance, in an EU apportionment system, if a production unit moves from Portugal to Spain, the tax base fraction allocated to Germany stays the same.

The external independence property can be interpreted as a pragmatic approach to save on complexity and information cost. Although plausible, this property may be violated under certain circumstances. For instance, a formula designed to account for the firm’s presence in tax havens would not have this property. Another example is provided in Gordon and Wilson (1986) who propose a system of allocation based on the firm’s capital (or property) where a region’s tax base is (in the notation used above): 

\[
 b_i = \frac{k_i/(1-t_i)}{\sum_j (k_j/(1-t_j))},
\]

i.e., a higher tax ceteris paribus implies a higher tax base share. Obviously, such a system satisfies the properties fixed coverage and strictly positive responsiveness. However, it does not have the property external independence. If capital is shifted from a low-tax to a high-tax region outside (both outside of region \( i \)), the denominator increases and region \( i \)'s tax base share is reduced.\(^{17}\)

I may now characterize the class of apportionment systems that satisfy the properties (i), (ii) and (iii). For this purpose, the following three Lemmas are useful.

**Lemma 1** *Due to external independence,*

\[
 b_i (\mathbf{x}) = b_i (x_i, \mathbf{X}_{-i})
\]  

**Proof.** Note first that the above is trivially true if \( I = 2 \). For \( I > 2 \), external independence implies that region \( i \)'s tax base share \( b_i \) is, for a given level of locational activity \( x_i \), the same for any two \( \mathbf{x}_{-i} \) and \( \mathbf{x}_{-i}' \) as long as the total activity level \( \mathbf{X}_{-i} \) remains constant, see eq. (11). This includes the \( \mathbf{x}_{-i} \) in which all activity outside of \( i \) takes place in one single region. If the latter is region \( j \), region \( i \)'s share is \( b_i (0, \ldots, x_i, \ldots, x_j, \ldots, 0) \) with \( x_j = \mathbf{X}_{-i} \). Thus, \( x_i \) and \( \mathbf{X}_{-i} \) is sufficient information to determine \( b_i (\cdot) \). ■

**Lemma 2** \( \frac{\partial b_i (\mathbf{x}, \mathbf{X}_{-i})}{\partial x_{ia}} - \frac{\partial b_i (\mathbf{x}, \mathbf{X}_{-i})}{\partial X_{-ia}} \) is the same for all \( i \).

\(^{17}\)Hines (1990) proposes a sophisticated (and, in terms of data requirements, demanding) system of formula apportionment, based on the notion of hypothetical average local profit. While having favorable efficiency properties, such a system does not satisfy the positive responsiveness property.
Proof. Fixed coverage implies $\sum b_i = \bar{b}$ and, thus, $\frac{\partial b_i(x_i X_{-i})}{\partial x_{ia}} + \sum_{j \neq i} \frac{\partial b_i(x_i X_{-i})}{\partial X_{ia}} = 0$. Due to external independence (and following the proof of Lemma 1), $\frac{\partial b_i(x_i X_{-i})}{\partial x_{ia}}$ is, for a given $j$, equal for all $i$. Thus, $\frac{\partial b_i(x_i X_{-i})}{\partial x_{ia}} = \frac{\partial b_i(x_i X_{-i})}{\partial X_{ia}}$ for $i \neq j$. It thus follows that $\frac{\partial b_i(x_i X_{-i})}{\partial x_{ia}} - \frac{\partial b_i(x_i X_{-i})}{\partial X_{ia}} = - \sum_j \frac{\partial b_i(x_i X_{-i})}{\partial X_{ia}}$ where the right hand side is equal for all $i$. ■

The above Lemma implies that an increase of $x_{ia}$ with an associated decrease of $X_{-ia}$ of the same size (i.e., with a constant overall level of activity, $X_a$) has an effect on $b_i$ that is the same for all $i$. In order to use this property, it is convenient to redefine the tax base fraction $b_i(\cdot)$ as a function of the regional activities in $i$ and the total activity levels, i.e., redefine $b_i(x_i, X_{-i})$ as $b_i(x_i, X)$. While $\frac{\partial b_i(x_i, X_{-i})}{\partial x_{ia}}$ holds the level of $X_{-ia}$ constant and allows for an adjustment in $X_a$, $\frac{\partial b_i(x_i, X)}{\partial x_{ia}}$ holds $X_a$ constant and allows for an adjustment in $X_{-ia}$; to be specific, $d x_{ia} = -d X_{-ia}$. Formally speaking, $\frac{\partial b_i(x_i, X)}{\partial x_{ia}} = \frac{\partial b_i(x_i, X_{-i})}{\partial x_{ia}} - \frac{\partial b_i(x_i, X_{-i})}{\partial X_{-ia}}$. With Lemma 2, I can derive the following statement.

Lemma 3 $\frac{\partial b_i(x_i, X)}{\partial x_{ia}}$ is the same for all $i$.

Proof. See above. ■

The consequence from the above Lemma is that $\frac{\partial b_i(x_i, X)}{\partial x_{ia}}$ cannot specifically depend on $x_{ia}$ or any other region-specific activity. This means that, for a given $X_{-ia}$, the tax base share $b_i$ is linear in $x_{ia}$.

The following Proposition describes the general form of allocation systems with properties (i) through (iii).

Proposition 2 A system of tax base allocation system with fixed coverage, positive responsiveness and external independence has the following general form:

$$b_i = z_i^0(\cdot) + \sum_a z_i^a(\cdot) \frac{x_{ia}}{X_a} \tag{13}$$

where $\sum_i z_i^0(\cdot) + \sum_a z_i^a(\cdot) = \bar{b}$. The weights $z_i^0(\cdot)$ and $z_i^a(\cdot)$ may be functions of the firm’s characteristics, e.g., its aggregate activity levels $X_a$ and $X_{a+1}$, but only $z_i^0(\cdot)$ may depend on the region’s identity or its characteristics (e.g., its GDP, population size).

Proof. Due to Lemma 2, $\frac{\partial b_i(x_i, X)}{\partial x_{ia}}$ is the same for all $i$, i.e., it depends neither on $x_i$ nor on $i$ in general. Define $\zeta_a := \frac{\partial b_i(x_i, X)}{\partial x_{ia}}$, implying that the share of the tax base allocated according to activity $a$ is $\sum_i \zeta_a x_{ia}$. Define as $z_i^0(\cdot)$ the share
of the tax base allocated to region $i$ independent of any activity location. Note that $z^0_i(.)$ may be region-specific while $\zeta^a(.)$ may not. The sum of all $b_i$ is $\bar{b}$ which implies $\sum_i z^0_i(.) + \sum_a \zeta^a(.) X_a = \bar{b}$. With $z^a(.) = \zeta^a(.) X_a$ follows the above Proposition. 

So far, the list of properties does not include anonymity, i.e. the requirement that the formula must be the same for all regions. However, the part of the above formula that allocates tax base according to activity satisfies an anonymity criterion (since the $z^a$ do not depend on $i$).

The total effective tax rate on the tax base allocated to the regions is

$$T = \sum_i b_i t_i = \bar{t}^0 + \sum_a z^a(.) \bar{t}^a$$

where $\bar{t}^0 = \sum_i z^0_i t_i$ denotes the $z^0_i$-weighted average tax rate (with $\bar{t}^0 = \bar{t}$ if $z^0_i(.) = z^0(.)$ for all $i$) and $\bar{t}^a = \sum t_i z^a_i X_a$ the activity-$a$-weighted average tax rate.

The generalized formula allows for the special case of $b_i = z^0_i(.)$, i.e. each region receives an individual fraction of the tax base independent of the activity located in the region, or, in a simpler version, $b_i = z^0(.) = \bar{b}$, i.e. the fraction is the same for all regions. In both cases, region $i$’s tax base satisfies the positive responsiveness property, however only in its weaker form. Specifically, the firm’s activity choices do not affect the regional tax base fraction $b_i$. In fact, a region would be entitled to a fraction of the tax base, even if the firm were not active in this region at all. In order to avoid this kind of degenerate formula, I strengthen the positive responsiveness property by adding a nexus condition, which requires that the tax base fraction in a given region is zero, if there is no activity in this region.

Properties 3 (ii’) Positive responsiveness or strictly positive responsiveness with a nexus condition, where the latter requires

$$b_i(\mathbf{0}, \mathbf{X}) = 0$$

where $\mathbf{0}$ is a vector of zeros. With a nexus condition, $z^0_i = 0$ for all $i$, and the formula is fully anonymous (i.e. two regions with the same activity profile have the same tax base share).
The formula in eq. (7) has the properties (i), (ii’) and (iii). Note that the anonymity property has not been imposed by assumption; it rather emerges endogenously from the properties (i), (ii’) and (iii). The following Corollary provides a strong result for the special case in which the tax base is to be allocated based on only one variable.

**Corollary 1** Assume that the tax base is allocated based on only one activity, the quantity of which is \( x_i \) in region \( i \) and \( X \) in total, i.e. \( b_i = b_i (x_i, X) \). Then, with full coverage \( \bar{b} = 1 \), the only formula that satisfies properties (i), (ii’) and (iii) is

\[
b_i (x_i, X) = \frac{x_i}{X}
\]  

(16)

With less-than-full coverage, the formula is given by \( b_i (x_i, X) = \bar{b} \frac{x_i}{X} \).

**Proof.** The Corollary follows directly from Prop. 2 with \( z_i^0 = 0 \) and \( \bar{b} = 1 \) implying \( z^a = 1 \). ■

The above Corollary shows that if the allocation of tax base uses only one factor in its formula (e.g. sales), the formula above is unique in the sense that it is the only one that satisfies the criteria (i), (ii’) and (iii).

**Firm behavior under generalized apportionment**

In what follows, I analyze the effects of a generalized system of apportionment satisfying properties (i), (ii’) and (iii) on firm behavior. In order to save on notation, I will assume that \( \bar{b} = 1 \) (full coverage), i.e. that the apportionment system covers the total firm profit. The findings below can easily be generalized to the case of \( \bar{b} < 1 \).

With a formula of the shape \( b_i = \sum_a z^a (. \cdot \frac{X_m}{X_m}) \), see eq. (13) with \( z_i^0(.) = 0 \) for all \( i \), the first order condition with respect to some activity \( m \) can be expressed as

\[
\pi_{x,m} + \frac{X_m}{1 - T} = \left( \frac{X_m}{X_m} (t_i - \bar{t}^m) + \sum_a z^a \frac{X_m}{1 - T} \right)
\]  

(17)

where \( z^a_{X_m} = \frac{\partial z^a}{\partial X_m} \) denotes the marginal response of the weight \( z^a \) to an increase in total activity level \( X_m \). Compare the above equation to the simple formula in equation (7) with \( \tau = 0 \) that gives rise to the first order condition in (8), again with \( \tau = 0 \). The generalized formula above has an additional term reflecting that the weights \( z^a \) may respond to the choice of total activity, i.e. \( z^a_{X_m} > 0 \)
or $z^a_{Xm} < 0$, which can be used manipulate the firm’s behavior. The example below shows how this could work.

However, it immediately becomes clear that even the general formula is unable to mitigate the dispersion bias (captured by the first term in round brackets on the right-hand side of eq. 17). The new term, $\sum a z^a_{Xm} \bar{t}^a$, is not region-specific (i.e., it does not depend on $i$), which makes it unsuitable to deal with the dispersion distortion.

The fact that the new term in eq. (17) is common across regions suggests, though, that the $z^a(.)$ functions could be chosen in order to affect the distortion of the aggregate level of activity. While this is true in some cases, see the example below, it turns out that the potential gains of a formula manipulation depend on the geographical structure of the firm’s activity allocation. With $\sum a z^a(.) = 1$ and, thus, $\sum a z^a_{Xm} = 0$ follows that $\sum a z^a_{Xm} \bar{t}^a = 0$ if the $\bar{t}^a$ do not differ across activities. What does equal $\bar{t}^a$ across activities mean? Recall that $\bar{t}^a = \frac{1}{a} \sum_i t^a_i \frac{X^a_i}{X^a_n}$. That is, the $\bar{t}^a$ are the same for all $a$, if the $\frac{X^a_i}{X^a_n}$ are the same (though this is not a necessary condition) across activities (not across locations). Thus, even if the activity within the firm is unequally distributed across locations, flexible weights are toothless if all the activities are distributed in the same way (i.e., the $\bar{t}^a$ are the same).

The following Proposition summarizes these results.

**Proposition 3** Consider a generalized formula for apportionment of tax base of the following form: $b_i = \sum a z^a(.) \cdot \frac{X^a_i}{X^a_n}$.

(1) The dispersion of tax biases across locations under formulas adhering to properties (i), (ii’), and (iii), measured as the difference between the highest and the lowest marginal tax cost of activity is:

$$\left(t^{\text{max}} - t^{\text{min}}\right) \cdot \frac{z^m}{1 - T} \cdot \frac{\pi}{X^m}$$

(18)

which is the same as with formulas using constant weights.

(2) The level of tax biases is the same as under formulas with constant weights if the $\bar{t}^a$ are the same for all $a$.

The above Proposition is kind of a negative result and tells us that, under certain circumstances, the generalized formula does not have advantages over the simplified one. There are cases, however, in which the additional flexibility of
the generalized formula may be used to increase surplus. The following example illustrates such a case.

**Example 2**

Consider the example 1 from above with $\pi(k,l)$ and with $b_i = zK_i + (1 - z)L_i$ for all $i$ where $z$ is a function of $K$ and $L$, i.e.

\[ z(K, L) = \alpha_0 + \alpha_1 K + \alpha_2 L \tag{19} \]

and $\alpha_0$, $\alpha_1$ and $\alpha_2$ are constants. Note that the above formula satisfies properties (i) with full coverage, (ii') due to $b_i = 0$ if $k_i = l_i = 0$ and (iii).

The first order conditions with respect to $k_i$ and $l_i$ are given by

\[
\pi_{k_i} = \left[ \frac{\alpha_0 + \alpha_1 K + \alpha_2 L}{K} (i_i - \bar{i}_K) + \alpha_1 (i_i - \bar{i}_K) \right] \frac{\pi}{1 - T} \tag{20}
\]

\[
\pi_{l_i} = \left[ \frac{1 - \alpha_0 - \alpha_1 K - \alpha_2 L}{L} (i_i - \bar{i}_L) + \alpha_2 (i_i - \bar{i}_L) \right] \frac{\pi}{1 - T} \tag{21}
\]

An increase of $\alpha_0$ by $d\alpha_0$ and an associated decrease of $d\alpha_1 = -\frac{1}{K}d\alpha_0$ leaves $z$ unaffected. However, the tax cost of capital use, $\pi_{k_i}$, is reduced by $-\frac{1}{K} (i_i - \bar{i}_K) \frac{\pi}{1 - T} d\alpha_0$, while the marginal tax cost of labor remains unaffected. That is, without revenue loss, the firm can be incentivized to increase all capital stocks, which – as is shown in Example 1 above – is able to increase surplus under certain conditions. Similarly, an adequate variation in $\alpha_0$ and $\alpha_2$ may increase (or decrease, if desired) all labor inputs across firm locations.

Example 2 shows that a generalized formula could, in principle, be used to mitigate the distortion of the aggregate activity level and increase surplus. This finding should be taken with caution, though. Since the aggregate level distortion (just like the dispersion distortion) is firm-specific, a surplus-increasing reform is possible, if the specific activity profile of the firms under consideration is known. It is, however, impossible to design a formula that improves the efficiency of the allocation independent of where (and how) the firms are active.

### 3 Discussion

This paper shows that formula apportionment systems that satisfy two plausible properties, *fixed coverage* and *strictly positive responsiveness*, necessarily suffer
from a dispersion bias (due to differences in local tax rates). The dispersion bias is unavoidable as long as local tax rates are not uniform. Depending on the firm’s regional activity structure, there may also be distortion of the aggregate activity level, i.e. the overall activity level is too high or too low.

Adding another criterion allows deriving a generalized formula that may serve for the apportionment of tax base in the federation: external independence. The formula derived from these three criteria is a little more general than the formulas used in real-world tax systems or envisaged in existing plans for reform and, thus, may be used to rationalize them. Using the full potential of the generalized formula may, in some cases, mitigate the aggregate level distortion and yield a gain in surplus. However, this can only happen in special cases depending on the geographical allocation of activity within the firm. The dispersion bias, however, cannot be mitigated; the paper thus shows that, in allocation systems adhering to the three criteria discussed above, the only way to minimize the distortions due to the formula is to optimally choose the weights (as in Eichner and Runkel 2008 as well as Runkel and Schjelderup 2011).

4 Appendix

A more specific model of the firm

In this Appendix, I present a slightly more explicit model of the multiregional firm, which chooses inputs, sales shares and function locations. In each region $i$, the firm chooses a vector of $N$ inputs, $\mathbf{k}_i = (k_{i1}, \ldots, k_{iN})$. Let $\mathbf{k} = (\mathbf{k}_1, \ldots, \mathbf{k}_I)$ denote the vector of regional input vectors. The inputs determine an output $y = f(\mathbf{k})$ and an associated input cost $c(\mathbf{k}, \cdot)$. A fraction $s_i$ of the output is sold in region $i$ at a price of $p_i$ with $\sum s_i \leq 1$. The selling price may depend on quantity sold, i.e. $p_i = p_i(s_i, y)$ with $p_i' \leq 0$. In addition, there may be a cost to expanding sales in a given region, i.e. $c = c(\mathbf{k}, \mathbf{s}, y, \cdot)$ and $\frac{\partial c(\cdot)}{\partial s_i} = c_{s_i, y}$. Finally, the firm has different functions (e.g. risk management, marketing etc.), which can be codified in a vector $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_I)$ where $\mathbf{q}_i$ denotes the vector of functions located in region $i$. I assume that the relocation of functions has, for a given level of inputs, no effect on output. However, there may be a cost involved, i.e. $c = c(\mathbf{k}, \mathbf{s}, \mathbf{q})$. The relocation of firm functions

---

18A natural addendum (here neglected) to the list of desired properties is tax payer anonymity implying that, for a given activity level and structure and location, $b$ is the same for all firms.
will usually involve a discrete jump in activity in two locations. For notational convenience, I treat functions as a continuous variable and assume that the cost function is twice differentiable in functions \( q \). For simplicity, I assume that functions are measurable on a one-dimensional scale, which gives rise to a vector \( q = (q_1, \ldots, q_I) \).

Net profit is given by

\[
\pi (1 - T) = (\bar{p} f (k) - c (k, s \cdot f (k), q)) (1 - T)
\]

with \( \bar{p} = \sum s_t p_t \). Let \( R_i f (\cdot) \equiv p_t s_t f (\cdot) \) denote regional revenue and \( R'_i f (\cdot) = (p_t + p'_t s_t f (\cdot)) f (\cdot) \) marginal regional revenue.

The first-order conditions with respect to \( s_t \) and \( k_{in} \) are:

\[
\begin{align*}
    s_t &: \quad (R'_i - c_{s,y}) f (k) (1 - T) - \frac{dT}{ds_t} + \lambda^s = 0 \\
    k_{in} &: \quad \left( \sum s_j R'_j f_{k_{in}} - c_{k_{in}} \right) (1 - T) - \frac{dT}{dk_{in}} + \lambda^k = 0 \\
    q_i &: \quad - c_{q_i} (1 - T) - \frac{dT}{dq_i} + \lambda^q = 0
\end{align*}
\]

In the absence of taxes, the firm equates the marginal regional revenues (net of marginal sales cost) across locations: \( R'_i - c_{s,y} = R'_j - c_{s,y} \) for all \( i, j \).

Moreover, with unbound inputs, \( \lambda^k = 0 \), it sets the marginal return to input equal to marginal input cost (weighted by the average marginal regional revenue, \( \sum s_j R'_j = \sum s_j p_j \left( 1 + \frac{1}{\varepsilon_j} \right) \)), where \( \varepsilon_j = \frac{dF_j}{dp_j} p_j s_j f \) denotes the price-elasticity of demand in region \( j \), i.e. \( f_{k_{in}} = \frac{c_{k_{in}}}{\sum s_j p_j (1 + \frac{1}{\varepsilon_j})} \). If the firm is a price-taker, this condition is \( f_{k_{in}} = c_{k_{in}} / \bar{p} \). With limited inputs, \( \lambda^k > 0 \), the gap between marginal return and weighted marginal cost is equalized across locations. Finally, it minimizes the cost by allocating functions such that the marginal cost of function location is equalized across locations.

With taxes, it depends on how the tax burden is affected by marginal variations in the choice variables. Thus, it depends on the formula.

References


