Efficiency properties of labor taxation in a spatial model of restricted labor mobility

Patrik Guggenberger\(^ a, b, c \), Ashok Kaul\(^ b \), Martin Kolmar\(^ c \)

\(^ a \) Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268, USA
\(^ b \) Institute for the Study of Labor (IZA) and University of Bonn, Bonn, Germany
\(^ c \) Department of Economics, University of Konstanz, Konstanz, Germany

Received 15 February 1999; accepted 12 July 2001

Abstract

We examine the efficiency properties of labor taxation. A spatial model of an economy is introduced whose key feature is a new approach to restricted labor mobility. We characterize the efficient allocation of labor and properties of a decentralized equilibrium. An efficient allocation of labor can be compatible with marginal productivity differentials stemming from binding mobility restrictions. We investigate two scenarios of labor taxation (firm-specific taxes and country-specific taxes) and in both setups we give a complete characterization of the cases for which there is scope for redistribution without affecting efficiency. Finally, we discuss the applicability of our model in the context of the place of employment and place of residence principle of taxation. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Restricted labor mobility; Efficient allocation of labor; Labor taxation

JEL classification: H7; R5; J61

1. Introduction

An important strand of the literature on fiscal externalities, tax competition and redistribution models the interaction between competing local jurisdictions in a
game theoretical approach. We introduce a two-dimensional model of an economy with a new concept of restricted labor mobility and characterize conditions, under which tax policies of local jurisdictions do not affect the efficient allocation of labor. Our results can serve as a guideline for efficient redistribution. But we abstract from strategic considerations and do not ask whether and how the scope for efficient taxation and redistribution is in fact exploited.

In contrast to previous models in the literature, which assumed either unrestricted labor mobility (e.g., Myers, 1990), or labor mobility restricted by either a fixed cost of migration (e.g., Hercowitz and Pines, 1991; Myers and Papageorgiou, 1997) or a cost proportional to the migration distance (e.g., Mansoorian and Myers, 1993; Hindriks, 1999, 2001), our model focuses on prohibitive impediments to mobility. We assume that the individual’s qualification profile and her willingness to move determine a subset of firms for which she may be employed. This subset defines the individual’s effective area of mobility. While we assume that the individual can costlessly move to the firms in her area of mobility, she incurs an infinite cost if she chooses employment outside this area. In contrast to the above literature, which typically describes labor mobility as an interregional phenomenon, mobility in our paper encompasses both interregional and intraregional mobility.

Our approach of modeling mobility can generate an efficient equilibrium with different marginal productivities across firms, because mobility restrictions possibly prevent individuals from working for the firm with the highest marginal productivity. These productivity differentials can allow for efficient taxation with tax instruments that vary across jurisdiction. Different tax rates across jurisdictions can be rationalized by the goal of income redistribution across different regions. In our model, local authorities levy lump-sum taxes on wages earned by workers in firms located within the jurisdiction. Taxes can cause inefficient decentralized equilibria, since individuals decide to work for the firm that pays the highest net wage within their respective mobility areas. However, efficiency requires that individuals work for the firm with the highest marginal productivity (= gross wage) within their mobility areas. The main intuition for efficient taxation is that taxes can sometimes skim off some of the differences in marginal productivities without creating an incentive to individuals to move to another firm.

For the case of two local authorities (either two countries or two regions), we show that tax rates are efficient if the tax differential between the two countries is not higher than the minimum productivity differential of ‘directly connected’ firms across the border. Two firms are directly connected if there are workers whose area of mobility contains both firms. This finding can serve as a guideline for redistributive programs, for example between countries in the European Union (Regional Funds) or between regions in a country (e.g., the ‘Länderfinanzausgleich’ in Germany). We discuss the applicability of our results and policy implications in the context of different principles of taxation (place-of-employment principle versus place-of-residence principle) and different forms of mobility.
(migration versus commuting). Furthermore, we give some insight into the distortion of the labor allocation that results from inefficient tax rates in the case where the two countries form an equal productivity area. Intuitively, one would think that the distortion of the labor allocation is restricted to a stripe along the border. We show that the distortion will affect the whole economy.

Of a more theoretical nature is the comparable study for the case where the number of local tax instruments equals the number of firms. We show that there always exist efficient tax rates, such that in the implied decentralized equilibrium net wages are the same across all firms. This result constitutes an important benchmark: while efficient equalization of net wages can most likely not be achieved in the case of only two jurisdictions, it can always be achieved if we refine the tax code to firm-specific instruments.

An appealing feature of our model lies in its spatial dimension. We model the economy as a subset of the two-dimensional space of real numbers and assume that individuals are distributed across the economy according to some continuous distribution function. This generalizes the setup of Hindriks (1999, 2001) and Mansoorian and Myers (1993), who assume a uniform distribution of the population over the unit interval forming their economic union, and of other models in the literature that do not include spatial dimensions at all. An obvious advantage of our two-dimensional approach is the straightforward translation of our results to the real world geographical situation.

The ongoing integration of markets and removal of legal restrictions implies greater mobility of labor. For example, the Treaty of Rome allows citizens of the European Union to seek employment in any member state. Studies by Gyourko and Tracy (1989) provide evidence that labor in the United States is very mobile and responsive to tax-related changes (see also Inman and Rubinfeld, 1996, pp. 315–316, for more literature on related studies). On the other hand, Wildasin (2000) examines the degree of labor mobility in the European Union. He concludes that while labor mobility is in fact substantial, it is too low to be compatible with the assumption of unrestricted labor mobility. As extreme cases, our model can account for completely restricted mobility (if the mobility area of each individual consists of only one firm) and free mobility (if each individual can work for all the firms).

We believe that there are multiple explanations why de facto labor mobility is still far from unrestricted although de jure it is not. We can categorize them into pecuniary (direct migration cost and indirect costs, for example the loss of clientele or networks) and non-pecuniary (language and other qualification, culture, attachment to home) reasons. In our approach, we focus on what we believe to be prohibitive impediments to mobility, at least in the short run. For example, knowledge of the language in the country of employment can be expected to be a necessary prerequisite for most white-collar jobs. Cultural factors may also be interpreted as an impediment to international mobility. Specific professional qualification requirements exclude many individuals from working for
certain firms. The impediments lead to an effective area of mobility, which at least in the short run can be viewed as exogenous. It turns out that our way of modeling mobility restrictions may lead to an equilibrium with discrete jumps in wages and productivities within one institutionally integrated area. We think that these discrete jumps do indeed exist, not only between countries, but also within one country, for example within Germany (West versus East, South versus North), France (Paris region versus the surrounding areas) or Italy (North versus South).

The paper is organized as follows. In Section 2, we introduce a spatial model of restricted labor mobility. In Section 3, we characterize the efficient allocation of labor. In Section 4, we establish the relationship between an efficient allocation of labor and a decentralized equilibrium. In Section 5.1 we introduce firm-specific wage taxes to the model. We completely characterize the cases where tax differentials do not lead to inefficiencies. In Section 5.2, we look at the special case of a two-country model. Instead of firm-specific wage taxes we now allow for only two different tax rates on wages, one for each country, and again we completely characterize the conditions under which tax differentials do not lead to inefficiencies. When the two countries constitute an equal productivity area, we present additional insight into how tax differentials distort the allocation of labor. In Section 5.3, policy implications are discussed. Section 6 summarizes and concludes the paper. All proofs are given in the Appendix.

2. The model

We consider an economy that is populated by individuals and firms. The economy is assumed to be a connected subset $U$ of the two-dimensional set of real numbers $\mathbb{R}^2$.

Individuals: The individuals are distributed on $U$ according to the continuous density function $\theta: U \rightarrow [0, \infty)$. This density function determines the original place of residence of the individuals. Every individual $i$ inelastically supplies one unit of labor and draws utility $u_i$ from consumption. We assume $u_i' > 0$ and thus among all the firms an individual can possibly work for, she always chooses to work for a firm that pays the highest wage.

Firms: There are $n$ firms that are located within $U$. Their location is exogenously given. Each firm produces the same homogeneous good using labor $l$ as the only input factor. We denote the firms by $t = 1, \ldots, n$. Firm $t$ produces according to the production function $F_t(l)$ where $F_t: [0, 1] \rightarrow [0, \infty)$ with $F_t'(l) > 0$, $F_t''(l) < 0$, $\lim_{l \to 0} F_t'(l) = \infty$ and $\lim_{l \to 1} F_t'(l) = 0$ (Inada conditions). In particular, $F_t$ is strictly concave.

Mobility: To each firm $t$ we assign an open set $K_t \subset U$, which we call the catchment area of firm $t$. Its interpretation is the following. Individuals located in $K_t$ satisfy two properties: they are willing to work for firm $t$ (and would choose to work for firm $t$ if it pays a competitive wage) and firm $t$ is willing to hire them.
Individuals located outside $K_t$ are either not willing to work for firm $t$ or firm $t$ is not willing to hire them (because they lack the qualification).

For simplicity we assume that $U = \bigcup_{t=1}^{n} K_t$. We can think of the catchment areas as overlapping circles around the firms’ locations. But the model allows for more complicated patterns of mobility. For example, the catchment areas do not have to be connected sets.

Furthermore we assume that an individual can commute costlessly to a firm if she belongs to the firm’s catchment area (but cannot work for the firm at all if she does not belong to the firm’s catchment area).

Alternatively, we could assign to each individual the set of those firms that would employ the individual and for which the individual is willing to work. These sets define each individual’s mobility restrictions. In our model we do not specify in more detail why individuals are not unrestrictedly mobile although there are no institutional barriers. We take it as an empirical fact that not all individuals are willing and qualified to move to any place in order to work. The justification for restricted mobility may lie in pecuniary or non-pecuniary costs of migration. Cultural or psychological reasons may keep people from working outside a certain range from home. Missing qualification or language skills may keep firms from hiring certain individuals.

The representation of the firms’ catchment areas allows us to define a partition of the union $U$ as follows: let $S$ be a nonempty subset of all firms $\{1, \ldots, n\}$, $\emptyset \neq S \subseteq \{1, \ldots, n\}$. We define

$$M_S := \{ x \in U : x \in \bigcap_{t \in S} K_t \text{ and } x \not\in K_t \text{ for } t \not\in S \}.$$ 

$M_S$ is the subset of $U$ which consists of those individuals which can work for every firm $t \in S$ but not for firms $t \not\in S$. Note that all individuals in $M_S$ where $S$ consists of only one element $t$ must work for the firm $t$ and cannot work anywhere else. We then have the following partition of $U$: $U = \bigcup_{S} M_S$. To simplify notation we will often omit brackets for sets: e.g., $M_{\{1,2,3\}}$ is written as $M_{123}$. Fig. 1 gives an example with three firms. In this figure the economic union $U$ is given by a rectangle. Individuals who can only work in either firm 1, 2, or 3 are indicated by $M_1$, $M_2$, and $M_3$, respectively. Individuals who can either work for firms 1 or 2, 1 or 3, or 2 or 3 are represented by $M_{12}$, $M_{13}$, and $M_{23}$, respectively. Finally, individuals who can work in each of the three firms are in the set $M_{123}$.

We then define $A_S := \int_{M_S} \theta(x) \, dx$. $A_S$ is the mass of workers in $M_S$, the size of the labor force in area $M_S$. Because the variables $A_S$ are derived from the $M_S$, which form a partition of the area $U$, the labor force is partitioned by the $A_S$.

A model with unrestricted mobility of labor where every individual can work in every firm is a special case of our model in which $M_S = \emptyset$, $\forall S \neq \{1, \ldots, n\}$ and $M_{\{1,\ldots,n\}} = U$ hold. On the other hand, complete immobility of individuals (each worker can only work for one firm) corresponds to $U = \bigcup M_t$ ($t = 1, \ldots, n$).

If for two firms $s$ and $t$ there exists a set $S \subseteq \{1, \ldots, n\}$ such that $s, t \in S$ and
\( A_i > 0 \) then we say that the firms are *directly connected*. In words, two firms are directly connected if there exists a positive mass of workers whose area of mobility contains both firms \( s \) and \( t \).

3. Efficient allocation of labor

In this section we will derive conditions for an efficient allocation of workers among the firms. By an efficient allocation we mean an allocation, that maximizes total output in the economy, subject to the mobility restrictions of the individuals. We show below, that an efficient allocation is equivalent to an allocation, in which — given the allocation of the other individuals to firms — each individual works for a firm within her mobility area that has highest marginal productivity. Therefore, each individual maximizes her wage and thus her utility, subject to her mobility restrictions and taking as given the decisions of the other individuals.

A social planner’s problem is to maximize output by allocating all individuals in \( U \) who can work for more than one firm, i.e., for all sets \( S \) where \( S \) contains at least two elements she has to assign a firm \( s \in S \) to each of the individuals in \( M_s \).

To handle this problem, we introduce variables \( M'_s \) and \( A'_s \) for all nonempty subsets \( S \subset \{1, \ldots, n\} \) and for all \( t \in S \) with the following meaning. We denote by \( M'_s \) the subset of individuals in \( M_s \) who work for firm \( t \). The variables \( M'_s \) thus refine the partition of \( U \) induced by the \( M_s \). Whereas \( M_s \) contains all the
individuals who can potentially work for each of the firms in $S$ (but for no other firms), the sets $M'_t$ consists of only those people in $M_t$ who actually work for firm $t \in S$. Furthermore, define $A'_t := \int_{M'_t} \theta(x) \, dx$. Thus, $A'_t \geq 0$ is the measure of workers in $M'_t$, i.e., the size of the labor force in $M_t$ that actually works for firm $t$.

By the definitions of $A$ and $A'$ the following relationship holds: $\sum_{t \in S} A'_t = A_t$, i.e., $A_t$ the size of the labor force that can work for exactly the firms in $S$ is partitioned among the firms in $S$.

To simplify notation, we introduce the following vector of variables: $(A') := (A'_t)_{t \in S, |S| \geq 2}$.

The aggregate output is given by the sum over the firms’ outputs. The argument in firm $t$’s production function is the measure of workers employed at this firm. This measure is the sum of the measures $A'_t$ (mass of workers that the firm $S$ employs from individuals in $M_t$) and $A_t$ (the mass of workers that can only work for this firm). Thus the aggregate output is:

$$F(A) := \sum_{t=1}^{n} F_t \left( A_t + \sum_{S, |S| \geq 2, j \in S} A'_j \right). \tag{1}$$

The social planner’s optimization problem (`Problem (P)`) can then be formulated as follows:

$$\max_{(A)} F(A) \text{ subject to the two sets of conditions:}$$

$$\sum_{t \in S} A'_t = A_t \quad \forall S \subset \{1, \ldots, n\} \quad \text{with} \quad |S| \geq 2$$ \tag{2}

$$A'_t \geq 0 \quad \forall S \subset \{1, \ldots, n\} \quad \text{with} \quad |S| \geq 2 \quad \text{and} \quad \forall t \in S. \tag{3}$$

Because $F$ is a continuous function and because the set of possible arguments under the restrictions is compact, a maximum exists. The set of restrictions in (3) is a natural nonnegativity constraint on the size of the labor force. The set of restrictions in (2) makes sure that all the workers who can work for several firms are partitioned between these firms, i.e., the employed labor force equals the available labor force.

There may be multiple allocations $(A)$ that lead to maximal output. However, we can still show a uniqueness result in the following sense.

**Lemma 1.** Assume $(A)$ and $(B)$ are two different efficient allocations. Then, for all $t = 1, \ldots, n$ we have $F'_t = F''_t$ where by $F'_t$ and $F''_t$ we denote the marginal productivity of firm $t$ in the allocation $(A)$ and $(B)$, respectively.

Lemma 1 proves that all allocations that lead to maximal output assign the same mass of workers to each firm. This does, however, not necessarily imply the uniqueness of a maximal allocation. Maximal allocations may differ in the way the total mass of workers assigned to firms is composed. Assume, for example, that in
the situation of Fig. 1 it turns out that in any maximal allocation (C) the workforce is distributed uniformly across the three firms, i.e., \( C_1 + C_1^1 + C_1^2 + C_1^1 + C_1^2 = C_2 + C_2^2 + C_2^1 + C_1^2 + C_1^1 + C_1^2 = 1/3 \). It may then be possible that two efficient allocations (A) and (B) are given by the following values:

\[
\begin{align*}
A_1 &= 1/10, A_1^1 = 2/10, A_1^2 = 0, A_2^1 = 0, A_2^2 = 1/10; \\
B_1 &= 1/10, B_1^1 = 2/10, B_1^2 = 1/10, B_2^1 = 1/10, B_2^2 = 2/10.
\end{align*}
\]

It is clear, that the efficiency of one allocation implies the efficiency of the other one.

The form of uniqueness of an efficient allocation established in Lemma 1 is sufficient for the purpose of this paper since the main results hinge on properties of the marginal productivities of the firms only.

To solve Problem (P), we form the Lagrangian function. Therefore we introduce one Lagrange multiplier \( \lambda \) and \( \mu_i \) for each constraint in the set of constraints (2) and (3), respectively. To simplify notation, we summarize all Lagrange multipliers in the vectors \( \lambda := (\lambda_i)_{S \subseteq \{1, \ldots, n\}, \mid S \mid \geq 2} \) and \( \mu := (\mu_i)_{S \subseteq \{1, \ldots, n\}, \mid S \mid \geq 2} \). The Lagrangian function is then given by:

\[
\mathcal{L}(A, \lambda, \mu) = F(A) - \sum_{S, \mid S \mid \geq 2} \lambda_i \sum_{t \in S} A_i^t - A_i^S + \sum_{S, \mid S \mid \geq 2} \mu_i A_i^S.
\]

The Kuhn–Tucker conditions are: \( \forall S \subseteq \{1, \ldots, n\}, \mid S \mid \geq 2 \) and \( \forall t \in S \).

\[
\begin{align*}
F_i^t - \lambda_i + \mu_i^t &= 0, \\
\mu_i^t A_i^t &= 0, \quad \mu_i^t \geq 0.
\end{align*}
\]

We will now present a first result that highlights the interplay between mobility restrictions in form of the catchment areas, technologies, and population distribution. The following lemma shows what constellations of marginal productivities are possible in an efficient allocation. Given a fixed distribution of the workforce and catchment areas for the firms (and therefore shares \( A_i \)), it is showed that arbitrary positive marginal productivities in the \( n \) firms (that is, any \( n \) positive numbers that stand for the marginal productivities of the firms once the workforce is allocated to the firms) are compatible with an efficient allocation of the workforce (A), if only one chooses appropriate production technologies for the firms.

**Lemma 2.** Let there be \( n \) firms with catchment areas \( K_i \) (\( i = 1, \ldots, n \)) and a distribution \( \theta \) of the workforce. Then for \( n \) arbitrary positive numbers \( f_i^t \) (\( i = 1, \ldots, n \)) there exist technologies of the firms [that is production functions \( F_i \) (\( i = 1, \ldots, n \)) such that in a solution to problem (P) [that is vectors (A) = (A_i^t), (\lambda), (\mu) satisfying conditions (2)–(5)]] we have \( F_i^t = f_i^t \).

The lemma proves in particular that in our model of restricted mobility,
marginal productivities do not have to be all equal in an efficient allocation. Note that Proposition 1 in the next section establishes that we can obtain each efficient allocation as a decentralized equilibrium. In the following discussion we can therefore interpret marginal productivities of an efficient allocation as wages. Let firms $s$ and $t$ be directly connected and assume firm $s$ has higher marginal productivity. Pick an individual who lives on the border of the catchment area $C$ of firm $s$ (we assume that $C$ is open and thus the individual does not belong to $C$) and works for firm $t$. He cannot work for firms $s$ because he is not included in $C$. However, any arbitrarily small neighborhood around the individual has a nonempty intersection with $C$. This means that arbitrarily close to the individual there exist other individuals being paid a higher wage or in other words, wage rates do not change continuously but may display discrete jumps. Because we think that these discrete jumps do indeed exist, for example within Germany (West versus East, South versus North), France (Paris region versus the surrounding areas) or Italy (North versus South), we consider this implication an important feature of our model. This feature cannot be obtained for example in a model where one assumes pecuniary mobility costs that are continuous in the travelling distance of the worker. In such a model individuals equate marginal gains from working for a firm and marginal costs of getting there, implying that individuals that live close to each other are paid approximately the same wage. In other words, changes in wages will only take place continuously. Such a model cannot provide an explanation for the observation of discrete wage jumps mentioned above.

In the unrestricted mobility literature, efficiency of an allocation implies equal marginal productivities across firms. In contrast to this observation, Lemma 2 proves that in our model only very little structure is imposed on efficient allocations a priori. This is done by demonstrating that for given catchment areas of the firms and a distribution of the workforce, any $n$ positive numbers can be realized as the marginal productivities (or wages, if wages are paid according to marginal productivities) of the firms in an efficient allocation. However, our model does impose more structure on efficient allocations once more information on catchment areas, population distribution and production functions is available. We will illustrate this with the following simple remark which follows immediately from the Kuhn–Tucker conditions.

**Remark.** Let $s$ and $t$ be two firms and $S \subseteq \{1, \ldots, n\}$ such that $s, t \in S$. 1. If $A'_s > 0$ and $A'_t > 0$, then $F'_s = F'_t$. 2. If $F'_s > F'_t$, then $A'_s \geq 0$ and $A'_t = 0$.

The first statement can be seen as follows. From (5) we have $\mu'_s = \mu'_t = 0$. Thus, from (4) we get the result. For the second statement note that by (3) it always holds that $A'_s \geq 0$. From (4) it follows that $F'_s - F'_t = \mu'_s - \mu'_t > 0$. Thus $\mu'_s > \mu'_t \geq 0$ and from (5) we must have $A'_s = 0$.

The interpretation of the remark is as follows. For two firms with a nonempty intersection of catchment areas it can be concluded that in an efficient allocation
marginal productivities have to be equal if among the workers who can work for both firms some workers actually work for the one and some workers for the other firm (first statement). Productivity differentials due to mobility restrictions can only be optimal if the less productive firm employs no workers from the intersection of their catchment areas (second statement). Otherwise one could increase total output by letting some of the individuals in the less productive firm work for the more productive firm. This is interesting from an empirical point of view. If we observe productivity differentials (or equivalently wage differentials) then we can conclude that mobility restrictions prevent the equalization of marginal productivities and therefore — as will be discussed later — different tax rates could be imposed without necessarily leading to inefficiency. Binding mobility restrictions are the key concept in this paper and we have to define carefully what we mean by it in the context of our model. We give a formal definition and then motivate it:

**Definition 1.** Let \((A) = (A_s)\) be an efficient allocation. We say that *mobility restrictions are binding* if (a) there are two firms \(s\) and \(t\) such that \(F'_s > F'_t\), or (b) the marginal productivities of all the firms are equal and there exists a set \(W \subset \{1, \ldots, n\}\) of firms such that for at least one set \(S \subset \{1, \ldots, n\}\) containing both elements of \(W\) and elements of \(CW\) (the complement of \(W\)) we have \(A_s > 0\), but for all such sets \(S\) satisfying these conditions we have \(A'_s = 0\) for all firms \(t\) in \(W\).

Note that negating the definition we get: *No mobility restriction is binding* if all firms have the same marginal productivities and for all sets \(W \subset \{1, \ldots, n\}\) for which there exists a set \(S \subset \{1, \ldots, n\}\) which contains both elements of \(W\) and elements of \(CW\) and for which \(A_s > 0\), it holds that there exists a set \(S' \subset \{1, \ldots, n\}\) which contains both elements of \(W\) and elements of \(CW\), for which \(A'_s > 0\) and \(A'_t > 0\) for a firm \(t\) in \(W\).

One should think of binding mobility restrictions in an allocation as productivity differentials between directly connected firms (part (a) of the definition). In an efficient allocation without restrictions on labor mobility this situation cannot occur. However, with restricted mobility it can. For example, take the simple case of just two firms \(s\) and \(t\). Imagine in equilibrium firm \(s\) employs all the workers who can work for both firms but still has higher marginal productivity than firm \(t\). A reason for that to happen is that the mass of individuals who can work for firm \(s\) only, is relatively small. Another reason could be a superior technology in firm \(s\). If mobility was unrestricted in this case, workers from firm \(t\) would switch to firm \(s\) until marginal productivities would be equalized.

Part (b) of the definition is more delicate, in that it deals with efficient allocations arising under restricted mobility that would still be efficient if we allowed for free mobility. For simplicity, take again the example of the two firms. In the notation of the definition we have \(W = \{t\}\) and \(S = \{s, t\}\). Assume that again
s employs all the workers who can work for both firms but this time marginal productivities are equalized. If we looked at the same scenario under free mobility, nothing would change. Why then do we define this situation as one of binding mobility restrictions? The rationale for including (b) in the definition of binding mobility is that any arbitrarily small upward change in the marginal productivity of firm s (induced, for example, by applying a new technology) will lead to a situation where total output is strictly smaller than in the case of free mobility because we could increase output by shifting some workers from t to s. However, due to mobility restrictions this is impossible. The case of several firms in part (b) is to be understood in the same way: all firms have the same marginal productivities. There exists a set W of firms that employs none of the individuals who can work for a firm in the complement of W although some individuals who work for firms in CW could also work for a firm in W. The equilibrium is not going to be affected by taking away mobility restrictions since marginal productivities of the firms are already equalized. However, any arbitrarily small increase in marginal productivity of a firm s ∈ W that is directly connected with a firm in W or a negative shock to the marginal productivities of all the firms in W will lead to different allocations under the two scenarios of restricted and unrestricted mobility. In the latter model, the relative increase in marginal productivity of the firm s (or equivalently the negative relative shock to firms in W) leads individuals to move from firms in W to s, in the former case this is impossible due to mobility restrictions.

In Section 4 we will introduce the concept of a decentralized equilibrium. Individuals take wages as given and move to the firm that pays the highest wage subject to their mobility restrictions. Thus, when it comes to the decentralized allocation what really matters are wages, which (in a more general setting introduced in Section 5) are equal to marginal productivities minus taxes. If what really affects the allocation are wages and not marginal productivities, then in the above explanation of part (b) of the definition the negative relative shocks to the firms in W can also be thought of as wage taxes on those firms.

We conclude this section with the following important lemma. It establishes that binding mobility restrictions are an intrinsic property of an efficient allocation. The definition does not depend on the special efficient allocation we pick. In other words:

**Lemma 3.** Let (A) and (B) be two efficient allocations. Then mobility restrictions in (A) are binding if and only if mobility restrictions in (B) are binding.

### 4. The decentralized equilibrium

A natural step is now to establish the relationship between the efficient allocation characterized by (2)–(5) and a decentralized solution. We assume a
competitive economy. In this economy there are $n + 1$ markets, one for the consumption good and $n$ for labor in the different firms. Without loss of generality we can fix the price of the consumption good to be equal to 1. For our purposes it suffices to characterize the economy by the $n$ wage rates of the firms. The wage rate of a firm equals its marginal productivity which in turn is determined by the size of the labor force working for this firm. Strictly speaking, the fact that firms pay marginal productivity wages requires that each individual is located within the catchment area of a large number of firms. We take a shortcut and assume that a single firm represents many firms with similar catchment area. In a decentralized market equilibrium individuals act as wage takers and maximize wage income subject to their mobility restrictions. We end up with a result in the spirit of the two welfare theorems.

**Definition 2.** A decentralized equilibrium is an allocation of the workforce $(A) = (A_i)$ and a set of induced wages $w_1, \ldots, w_n$ (induced by the marginal productivities of the firms) such that no individual $i$ can reach a firm with a strictly higher wage than $i$ earns in the allocation $(A)$.

Note that in the above definition and throughout the paper we say that a certain property holds for no individual, if whenever the property holds on a set $K$ of individuals, it follows that $\int_K \theta(x) \, dx = 0$. We say that a certain property holds for all individuals, if it holds on a set $K$ of individuals for which $\int_K \theta(x) \, dx = 1$.

We can now formulate the following important result:

**Proposition 1.** Let $(A)$ be an allocation. If $(A)$ constitutes a decentralized equilibrium, then it is efficient. If $(A)$ is efficient, then it can be obtained as a decentralized equilibrium.

The first part of the proposition obviously carries over to the case of uniform (union-wide) lump sum taxes. Thus, Proposition 1 constitutes an important benchmark when we introduce decentralized tax instruments.

**5. Differential wage taxation and the scope for redistribution**

In this section we introduce lump-sum taxes on wages. We do not specify what the intended purpose of this tax is. What we have in mind, however, is a redistributive tax used to generate a more equal income distribution across regions in a union. Equalization of income levels requires differences in net taxes. While tax differentials necessarily induce inefficiencies in models of unrestricted mobility of labor, there can be scenarios in our model when different tax rates can be
implemented without affecting the efficiency properties of the economy. In Sections 5.1 and 5.2 we will investigate and solve two scenarios: first, we will allow for a centralized, union-wide authority that can impose firm-specific taxes $T_t (t = 1, \ldots, n)$ on labor, that are possibly different for all firms. However, the notion ‘firm-specific taxes’ should not necessarily be taken literally. What is important is the fact that the tax code is not restricted to a uniform-lump sum tax but can condition on smaller units (for example regions) within the union. The smallest observable units in our model are firms. Under this assumption of firm-specific taxes, we will completely characterize the cases in our model in which any tax differential necessarily implies inefficiency, that is, the cases that are qualitatively equivalent to the results from the unrestricted mobility literature.

The second scenario is more realistic and of higher empirical relevance: we will divide the union in two countries and restrict the number of available tax instruments to only two country-specific taxes on wage income. Again, we present a complete characterization of the cases for which differing tax rates do not lead to inefficiency and thereby answer the question when redistribution in a federation can be done without affecting the efficiency of the economy. Finally, in Section 5.3, we will apply the results to wage taxation in the European Union under the place-of-employment versus the place-of-residence principle.

An important assumption implicitly made throughout this section is that each firm’s catchment area $K_t$, the set of workers a firm can attract, is given exogenously. In particular, it is independent of the tax rates. This assumption may seem to be questionable. If taxes become very high for some firm $t$ relative to the others, one might expect the catchment areas to change: workers who so far have not been willing or qualified to work for a firm $s$ (because, for example they did not speak the language of the country where $s$ is located) may now reevaluate the financial advantage of working for firm $s$ versus the multiple drawbacks to quit firm $t$ that had been predominant so far. Some workers may then indeed decide to quit $t$ and acquire the necessary skills to work for firms $s$. There is at least two ways to face this critique. First, we could interpret the model as a short-term model. After the tax is imposed on a firm workers will not quit immediately. Acquiring additional skills (a new language for example) will take some time. Second, we could interpret the mobility restrictions as real prohibitive impediments. Regardless of financial disadvantages individuals may, for example, not be willing to give up their cultural background.

5.1. Tax differentials and inefficiencies

We have mentioned above that tax differentials need not lead to inefficiencies in our model. Before we give a general characterization of when this can happen, we want to discuss a simple illustrative example:
Example. Let \((A)\) be an efficient allocation. Let \(s \in \{1, \ldots, n\}\) be a firm. Define \(T := \min\{F'_s - F'_t\}\) where the minimum is taken over all firms \(t\) that are directly connected with firm \(s\) and assume \(T > 0\). Then any tax \(T_s\) levied on the wages of the workers of firm \(s\) such that \(0 \leq T_s < T\) does not lead to inefficiency.

In the example we assumed that \(T > 0\). This case can occur in our model due to binding mobility restrictions. The intuition of the result is simple. If there are islands of high productivity, then these islands can be taxed without creating incentives to change the workplace as long as the tax is smaller or equal to the marginal productivity difference between this island and the neighbor with the next highest marginal productivity. It can be interpreted in favor of interregional transfer payments between regions with different standard of living, i.e., between rich and poor areas, that are intensively discussed in countries like Germany (‘Länderfinanzausgleich’) or in the European Union (Regional Funds).

We now tackle the general problem. Let us assume that a centralized authority can impose firm-specific taxes \(T_t (t = 1, \ldots, n)\). We are interested in describing all the cases in this setup in which redistribution always implies inefficiency. The remaining cases are then the ones where redistribution can be achieved without affecting efficiency. We thus build a bridge from our model of restricted labor mobility to models with free labor mobility (where differential taxes always imply inefficiency). To do so, the following definition will prove helpful:

**Definition 3.** Let \((A) = (A')\) be an efficient allocation. If there is no binding mobility restrictions in \((A)\) and any two firms \(s, t \in \{1, \ldots, n\}\) are linked by a chain of directly connected firms (that means for all firms \(s, t \\exists n = n(s, t) \\exists k_1, \ldots, k_n\) such that \(k_1 = s, k_n = t\) and \(\forall r = 1, \ldots, n - 1\) the firms \(k_r\) and \(k_{r+1}\) are directly connected) then we will call the economy an equal productivity area or region.

In particular, if the economy constitutes a region, marginal productivities throughout all the firms will be equal. Therefore, intuitively, one should think of a region as a connected area of equal standard of living. The concept of a region sheds light on the interpretation of a homogeneous good in the case of a spatial model. If one takes the Arrow–Debreu definition of a good literally, individuals living at different places supply different factors because they can be spatially distinguished. This definition would be inexpedient because it would lead to a total fragmentation of markets and it would be unclear how this concept could be related to the notion of homogeneous labor used in non-spatial models of markets. With the above definition we can define labor as being homogeneous if it is
employed within a region. Thus, it is not the locational difference per se that makes goods different, but the missing connection by market transactions.\footnote{The discussion is related to the notion of horizontal product differentiation from standard locational theory (see e.g., Gabszewicz and Thisse, 1992). Goods are said to be horizontally differentiated if consumers differ in demand even if the goods are sold at the same price, like for different flavors of ice cream. These goods would not be homogeneous according to the above definition. Ice cream with different flavors sold at the same price corresponds to labor supply of different equal-productivity areas that has the same price in equilibrium by chance. The economic rationale for this conclusion is as follows. In the language of Lancaster’s (1966) characteristics model, goods with different characteristics are different goods. Characteristics are, however, identified with locational distance together with a strictly monotone transportation-cost function that implies differences in utility net-of-transportation costs for goods with different distances from the own location. Hence, goods with different location but the same distance have the same characteristics from the point of view of the consumer and are therefore homogeneous. It is the strict monotonicity of the transportation-cost function that is missing in our model. This implies that the set of spatially differentiated but nevertheless homogeneous labor inputs may be larger than in the standard model. We would like to thank Konrad Stahl who pointed our attention to this similarity.}

Assumption: For simplicity, from now on we will assume that an individual who can work for firms $s$ and $t$ and actually works for $s$ changes her job to firm $t$ after the imposition of the lump sum tax, only if the wage in firm $t$ is going to be strictly higher than the wage in $s$.

We then have the following result:

**Proposition 2.** Let $(A)$ be an efficient allocation. Then the following two statements are equivalent:

1. The imposition of any set of tax rates $T_i \geq 0$ ($i = 1, \ldots, n$) such that $\exists t_1 \neq t_2$ with $T_{t_1} \neq T_{t_2}$ leads to an inefficient allocation. (In words: Any tax differential implies inefficiency.)
2. The economy constitutes a region.

The proposition completely characterizes the cases in our model which are qualitatively equivalent to models of unrestricted mobility when it comes to the impact of differential taxation. If and only if an allocation constitutes a region, any differential taxes will be inefficient. This result is to a certain degree intuitive but not trivial. Note that the characterization is based on the quite complex notion of ‘no binding mobility restriction’ as given in Definition 1. In particular, it is not obvious why there are no other cases beside regions where any differential taxes will be inefficient.

By generalizing the $1. \Rightarrow 2.$ direction of the proof we get the following corollary that states that in our model it is always possible to equalize wages efficiently by
firm-specific taxes. The idea is quite simple: one can inductively skim off marginal productivity differences between firms by firm-specific taxes without affecting the allocation.

**Corollary 1.** Let \((A)\) be an efficient allocation. Then there is firm-specific tax rates \(T_t \geq 0\) \((t = 1, \ldots, n)\), such that the resulting decentralized equilibrium after the imposition of the taxes is efficient and net wages \(F^-_t - T_t\) in all firms are equal.

The corollary is based on the assumption of firm-specific tax rates and is thus of only very limited empirical relevance. The corollary is still important because it establishes a theoretical benchmark and can be compared to the results in the next section, when more realistic assumptions on the tax code are made.

### 5.2. Differential wage taxes in the two country case

In this section we assume that the union is divided into two countries \(R\) and \(Q\): \(U = Q \cup R\). The authorities of the two countries impose country-specific wage taxes \(T_R\) and \(T_Q\) to be paid by individuals working for firms located in the respective country. Because we are only interested in the tax differential \(T_R - T_Q\), we can assume that only country \(R\) levies a tax, \(T := T_R\) and \(T_Q := 0\). In the next proposition we will characterize when differing tax rates lead to inefficiency.

Define \(T' := \min (F^-_s - F^-_t)\), where we take the minimum over all pairs \((s,t)\) such that firm \(s\) is located in \(R\) and firm \(t\) in \(Q\) and \(\exists S \subset \{1, \ldots, n\}\) with \(s, t \in S\) and \(A^s_s > 0\). If no such pair exists, set \(T' = 0\). Note that since \(A^s_s > 0\) implies \(F^-_s \geq F^-_t\) in the definition of \(T'\), it always holds that \(T' \geq 0\).

**Proposition 3.** Let \((A)\) be an efficient allocation. Assume a lump sum tax \(T \geq 0\) is levied on the wages of workers in region \(R\). A resulting decentralized equilibrium after the imposition of the tax is efficient if and only if one of the following conditions is met: 1. \(T \leq T'\); 2. for all \(s \in R\) and \(t \in Q\) and for all \(S \subset \{1, \ldots, n\}\) containing both \(s\) and \(t\) it is \(A^s_s = 0\).

In condition 1. \(T'\) serves as an upper bound for efficient taxation by country \(R\). \(T'\) is the minimum productivity differential between two directly connected firms \(s \in R\) and \(t \in Q\), such that there exist individuals who could work for \(t\) but do not work for \(s\). If we impose a higher tax than \(T'\), these individuals may have an incentive to change to firm \(t\). The intuition behind the first condition is that if country \(R\)'s border is uniformly more productive than the border of country \(Q\) then the minimum productivity differential can be skimmed off by a tax without affecting the efficient allocation.

The second condition is of rather pathological nature. It basically says that if the border region of country \(R\) was uniformly less productive before imposition of the tax (in the sense that in the original allocation before taxation each individual who
can work for a firm in $R$ and a firm in $Q$ works for the firm in $Q$) then country $R$ can be punished by a tax without affecting efficiency. All the workers in the border region who could possibly quit for a firm in $Q$ already work for a firm in $Q$. The others cannot evade from the tax. If the purpose of the tax is redistribution between ‘rich’ and ‘poor’ we should focus on the first condition.

Note how the scope for redistribution hinges on the degree of mobility. If, for example, we are in an environment of unrestricted mobility, then $T'$ will be zero. In general, $T'$ will gradually increase, if mobility is gradually restricted because binding mobility restrictions may lead to increased differences in marginal productivities.

The complexity of the proof of the proposition shows that the result is far from trivial. It is true that it seems to be quite intuitive that both conditions 1. and 2. do not affect efficiency. But note that even this apparently simple part of the proposition includes the non-trivial first part of Proposition 1 that a decentralized economy is efficient. (Take the case where $Q$ is the empty set. Then condition 2. holds. Therefore by Proposition 3 the decentralized equilibrium is efficient, which is the statement of the first part of Proposition 1.) Furthermore, it is not at all obvious to see the second part of Proposition 3. If we are given a tax that does not affect the efficiency of the allocation, why then has necessarily one of the conditions 1. or 2. of the proposition to be true?

As pointed out already, for empirical applications the first condition in the proposition is the relevant one. It pins down the scope for redistribution without efficiency losses of transfer payments between, for example, two countries or two regions in the EU. The co-existence of regions with high and low marginal productivities shows that mobility restrictions must be binding at the border areas of these regions. Wage taxes can be higher in the region with high marginal productivity without distorting the allocation of labor. Inter-jurisdictional transfers that might be legitimized by inter-jurisdictional inequality aversion can be efficiently administered by the means of high taxes on high productivity areas.

We will now investigate in more detail — for the special case where the two countries form an equal productivity area — how different tax rates affect the allocation of labor. The inefficiency after imposition of the tax is caused by workers who switch to firms in country $Q$ that pay higher wages (because wages of firms in $R$ are subject to the tax) and thereby causing a decrease in the marginal productivities of those firms. In general, there is no reason why the different tax rates in the two countries should have an impact on the workforce of all firms in $U$. In fact, the explanation for inefficiency just given, makes it seem that the effect of the tax is restricted to firms that are directly connected to firms of the other countries and that only the workers located in a stripe containing the frontier will change to another firm. As we will show now, this intuition does not hold in the case where the two countries form an equal productivity area. The tax will affect the amount of workers in all the firms in $U$. Reallocation of the workforce spreads from the frontier over the whole country similar to a domino effect.
Proposition 4. (Domino effect) Let $U$ be an equal productivity area. Introduction of a tax $T > 0$ in one country leads to a global change in the allocation of labor, i.e., if $s$ is any firm then after the introduction of the tax, $s$ will employ either strictly more or strictly less workers.

To some extent, Proposition 4 can be interpreted in defense of the results of that branch of the literature that assumes free labor mobility, namely that redistribution efforts based on differential tax rates necessarily have to imply inefficiency. Since, as discussed earlier, this assumption seems to be far too strong for real world applications, the results may seem to be of only modest empirical relevance. However, the proposition provides an example in the context of a model of restricted mobility — a union that forms an equal productivity area — that leads to the equivalent implication that different tax rates result in an inefficient distortion of the workforce.

5.3. Place-of-employment versus place-of-residence principle, commuting and migration

In this subsection we will focus on a multi-country setup and assume that each country can levy country-specific taxes on wages. We will discuss the empirical relevance of our model in the context of different tax principles and different types of mobility. We will start by defining the relevant concepts:

The two tax principles can be defined as follows. If workers pay taxes in the country in which they are employed, we call it a place-of-employment (PoE) principle. If they pay taxes in the country in which they have their residency, we call it a place-of-residence (PoR) principle.

We focus on two different types of mobility: commuting and migration. The former means that an individual who is a resident in country $Q$, does not move to country $R$ if she accepts a job in $R$. She remains a resident of country $Q$ and commutes on a regular (daily, weekly) basis from her home to her place of employment. The latter concept means that whenever an individual accepts a job in a country she is not resident of, she will move her place of residency to that country. Given the way we use the terms commuting and migration it would be more precise to talk about international commuting and migration.

Throughout the paper we have worked with the PoE principle. It can be argued that this is the prevailing principle in reality, for example in the European Union (EU):

• In most countries, compulsory health insurance is financed out of labor income. Workers have a right to benefits according to the principles of the country in which they are employed. This holds even true for commuters (European Court
of Justice, 1978, 825). Hence, the net burden of the health systems varies with the country of employment, a PoE principle applies.

- According to Regulation 1408/71 of the EU, workers pay contributions and accumulate benefit claims in the *public pension systems* according to the principles of the national systems. In most countries these systems are financed from contributions that are a fraction of an individual’s labor income. An exception is Denmark, which finances its pension system through general taxes. Benefits depend either on contributions (for example France or Germany), are fixed (for example UK or Sweden), or depend on the period of residency (Denmark).

- To avoid double taxation in the field of *wage-income taxes* as a consequence of cumulative taxation at the place of work and residence, double taxation conventions following the OECD Model Tax Convention (see OECD, 2000) have allocated the right to tax to one country. In practice, although Art. 15 of the OECD Model Tax Convention appears to make PoR taxation the general rule, the exceptions to that rule in Art. 15 II amount to a partial reversal of that rule. As a result, the PoE principle often holds (see Vogel, 1997, Art. 15, m.nos 6ff). However, for the case of commuters some countries in their treaty practice once again reverse the PoE principle and install the PoR principle (Vogel, 1997, Art 15, m.no 86c).

The phenomenon of international commuting has attracted considerable attention in the discussion about legal prerequisites for effective mobility (see for example regulations 1408/71, 574/72, and 1248/92 of the EU that settle problems of co-ordination in the field of social security). However, so far the formal literature on fiscal federalism had to say very little about it. On the other hand, our model — under the PoE principle — is naturally applicable under both forms of mobility. Therefore, whenever we observe the PoE principle, we can use the results of the previous section for labor tax policy advice. Take for example the insights of Proposition 4: consider a border area between countries with a relatively high population density and a high willingness to commute, such as — in the case of the European Union for example — the border areas of Belgium, The Netherlands, Luxembourg, France and Germany, the Spanish/Portuguese border or the German/Austrian border. In these areas, cultural and language barriers are relatively small. Assume the assumptions of the proposition hold, in particular we have the same marginal productivities of firms along the border line. Assume also we can apply the PoE which — as shown above — is reasonable in the case of the European Union. Then we can conclude that differences in tax rates will lead to an inefficient allocation of labor although labor is not unrestrictedly mobile. If migration is small compared to commuting the degree of efficiency loss could be reduced by switching from PoE to PoR. We can therefore conclude that even in the case of restricted labor mobility, policy coordination is relevant, particularly in border areas.
We have not dealt with the PoR principle in this paper. Under the PoR principle different tax rates may have an impact on efficiency properties only if migration takes place. Obviously, if there is only commuting and no migration different PoR tax rates do not distort the labor allocation. If in the context of our model we want to explore the impact of PoR in the case where both commuting and migration are possible, the analysis will become more difficult in that each firm does now have two (or in a multiple country setup multiple) net wages, one for each country of residency of a potential worker. We leave this analysis for future research.

Note that with unrestricted labor mobility, the distinction between the two principles of taxation would parallel the discussion in the literature on capital-tax competition (see Bucovetsky and Wilson, 1991) about the differences between the source principle (which parallels the PoE principle for wage taxes) and the residence principle (which parallels the PoR principle).

6. Summary and conclusion

This paper has considered a two-dimensional model of labor taxation. The key feature has been a new and quite general approach to partially restricted labor mobility. We introduce the concept of binding mobility restrictions which represents the driving force for our results. We can summarize our findings as follows.

1. We analyze the properties of an efficient allocation of labor in our model. Contrary to the unrestricted mobility literature, in our model an efficient allocation of labor is compatible with different marginal productivities of the firms. In fact, we show that given catchment areas of the firms and a distribution of the workforce, any \( n \) positive numbers can be realized as the marginal productivities of the firms in an efficient allocation. Furthermore we show, that an allocation is efficient if and only if it is a decentralized equilibrium.

2. We analyze the effects of labor taxation on efficiency. We investigate two scenarios of labor taxation and in both setups we give a complete characterization of the cases for which there is scope for redistribution. First, if an authority can impose firm-specific taxes, then any tax differential implies inefficiency iff the economy constitutes a region, i.e., a connected area of equal standard of living with no binding mobility restrictions. Second, in a two country setup with two national tax instruments, the imposition of a tax \( T \) on wage income in country \( R \) (and zero tax in the other country \( Q \)) is efficient iff essentially \( T \) is not higher than the minimum marginal productivity differential of directly connected firms in \( R \) and \( Q \). The findings give a basis for interjurisdictional financial adjustments between ‘rich’ and ‘poor’ countries or regions. In the case
of firm-specific taxes we show that there always exist efficient tax rates \( T_t \) for each firm, such that in the resulting decentralized equilibrium net wages are equal.

3. In a two-country setup we investigate in more detail the impact of different tax rates in both countries. If the economic union constitutes an equal productivity area, the tax will have a global effect on the allocation of labor. Every firm in the economy, not just the ones close to the borderline, will either employ less or more workers after the imposition of the tax.

4. We discuss the empirical relevance and applicability of our model in the context of two tax principles, namely the place-of-employment (PoE) principle and the place-of-residence (PoR) principle and two different notions of mobility, namely migration and commuting. We find that even with restricted mobility, tax policy coordination between countries can be relevant.

Acknowledgements

We would like to thank Anat Bracha, Friedrich Breyer, Nicola Fuchs, Geoffrey Garrett, Roland Ismer, Benoit Mercereau, John E. Roemer, Guenther G. Schulze, Konrad Stahl, and two anonymous referees for helpful comments. The first author thanks the Cowles Foundation for research support via a Cowles Prize.

Appendix A.

Proof of Lemma 1

Define the allocation \( (C) \) by \( C_s^i := 0.5A_s^i + 0.5B_s^i \) (for all \( S \subset \{1,...,n\}, |S| \geq 2 \) and for all \( t \in S \)) and \( C := B_t = A_t \) (for \( t = 1,\ldots,n \)). Note that (2) and (3) hold. Then \( F(C) = \sum_{s=1}^n F(c_s + \sum_{|S|\geq2,s\in S} ^ c_s) = \sum_{s=1}^n F(c_s + 0.5(A_s + \sum A_s^i) + 0.5(B_s + \sum B_s^i)) \geq \sum_{s=1}^n (0.5F(c_s + A_s^i) + 0.5F(c_s + B_s^i)) = 0.5F(A) + 0.5F(B) = F(A) = F(B) \), where the inequality follows by concavity of the \( F \). If for a certain \( t \) we had \( F_A^t \neq F_B^t \), it would follow that \( A_t + \sum A_s^i \neq B_t + \sum B_s^i \) and thus the inequality would be strict by strict concavity of the \( F \). This contradicts the maximality of \( F(A) \).

Proof of Lemma 2

Note that since we are given \( n \) firms, \( K_i \) and \( \theta \), we are implicitly given \( (M_s) \) and \( (A_s).W.l.o.g. \) let \( f'_1 \geq f'_2 \geq \cdots \geq f'_n > 0 \). We will show that there exist Lagrange multipliers \( (\lambda) \) and \( (\mu) \) and an allocation of the workforce \( (A) = (A_s^i) \) such that conditions (2)–(5) hold for any \( S = \{s_1,\ldots,s_t\} \) with \( |S| \geq 2 \) and \( t \in S \).

Let \( S = \{s_1,\ldots,s_t\} \), with \( |S| \geq 2 \) with \( s_1 < \cdots < s_t \) (implying \( f'_{i_1} \geq \cdots \geq \)}
Set $A^*_j := f'_j$, and for all $r = 1, \ldots, k$ set $\mu^*_j := \lambda^*_j - f'_j \geq 0$ (The so defined $\mu^*_j$ will fulfill condition (4) with $F'_j$ replaced by $f'_j$ and one part of condition (5)). If $\mu^*_j > 0$ set $A^*_j = 0$ (to guarantee the other part of condition (5)). Note that always $\mu^*_j = 0$, so that $A^*_j$ has not been fixed yet and can still be chosen, allowing us to satisfy condition (2). If several of the $\mu^*_j = 0$, choose the corresponding $A^*_j$ such that $A^*_j \geq 0$ and condition (2) is fulfilled, which is obviously possible.

Doing this for all $S \subseteq \{1, \ldots, n\}$, we define vectors $(A^*_k)$, $(\lambda)$ and $(\mu)$ that satisfy conditions (2)–(5) with $F'_j$ replaced by $f'_j$. We have in particular pinned down how many workers work for the respective firms. We still have one degree of freedom left: the choice of technology $F_t$ for each firm. By choosing an appropriate technology $F_t$, for each firm, we can guarantee that firm $t$ really has marginal productivity $F'_t = f'_t$ when employing the share of the workforce $W_t$ fixed by the allocation $(A^*_k)$, namely $W_t = A_t + \sum_{s, t\not= s} A^*_s$.

Proof of Lemma 3

By the symmetry of the statement it is enough to establish the $\Rightarrow$ direction of the lemma. In case (a) of the definition of binding mobility restrictions, the implication is clear by Lemma 1. Thus, let $W$ be as in part (b) of the definition. Again, by Lemma 1, it follows that the mass of workers employed by firms in $CW$ (by $CW$ we denote the complement of $W$) must be equal in $(A)$ and $(B)$. In $(A)$, all workers, who can work for firms in $CW$, do so. Therefore, the same must be true for $(B)$. Thus $B^*_t = 0$, for all firms $t$ in $W$ for the same $S$ that is used for the definition of binding mobility restrictions for the allocation $(A)$. □

Proof of Proposition 1

1. Suppose $(A) = (A^*_k)$ is a decentralized equilibrium which is inefficient. We show that this yields a contradiction. Let $(B) = (B^*_k)$ be an efficient allocation. It follows immediately that in $(B)$ there must be at least one firm which employs a strictly bigger share of the workforce than in the allocation $(A)$. Of all the firms employing a strictly bigger share of the workforce in $(B)$ than in $(A)$ let firm $s$ be one for which $F^s_B - F^s_A > 0$ is maximum (by $F^s_A$ resp. $F^s_B$ we denote the marginal productivity of firm $s$ in the allocation $(A)$ resp. $(B)$). Of all firms employing strictly more workers in $(B)$ than they did in $(A)$, $s$’s loss of productivity when changing from allocation $(A)$ to $(B)$ is biggest. Let $K := \{t \in \{1, \ldots, n\} \mid F^s_B = F^s_B \text{ and } F^s_A = F^s_A \}$ and $k := \#K$. $K$ is the set of firms having the same marginal productivity in $(A)$ and $(B)$ as firm $s$. It is clear that all firms $K$ employ a bigger share of the workforce in $(B)$ than in $(A)$. Trivially, $k < n$ as not all firms can employ a bigger share of the workforce in $(B)$ than in $(A)$. Therefore there must be firms $r \in K$ and $t \in K$ with the following properties: $\exists S \subseteq \{1, \ldots, n\}, r, t \in S, 0 \leq A^*_S < B^*_S$ and $A^*_S > B^*_S \geq 0$. (At least one firm in $K$ must employ workers in the allocation $(B)$ which in the allocation $(A)$ worked for a firm which is not in the set $K$.)
Because (A) is a decentralized equilibrium, we have \( F_t^{A'} \geq F_t^{A'} \). (This is true because in the allocation (A) a share of the workforce in \( M_s \) worked for firm \( t \) since \( A_s' > 0 \). These workers could have worked for firm \( r \) too. Thus the wage of firm \( t \) must have been bigger or equal than the wage of firm \( r \) in (A).)

Now consider the following three cases:

1. \( F_t^{B'} < F_t^{B'} \): we have \( 0 < B_s' \) and therefore the output in (B) could be increased by letting workers of firm \( r \) work for firm \( t \) which contradicts the efficiency of (B).

2. \( F_t^{B'} > F_t^{B'} \): because of \( F_t^{A'} \geq F_t^{A'} \), \( F_t^{B'} > F_t^{B'} \) implies that \( F_t^{A'} - F_t^{B'} > F_t^{A'} - F_t^{B'} \) contradicting the choice of firm \( s \).

3. \( F_t^{B'} = F_t^{B'} \): because of \( F_t^{A'} \geq F_t^{A'} \), \( F_t^{B'} = F_t^{B'} \) implies that \( F_t^{A'} - F_t^{B'} = F_t^{A'} - F_t^{B'} \) and thus \( F_t^{A'} = F_t^{A'} \). This implies that \( F_t^{A'} = F_t^{A'} \) because \( F_t^{B'} = F_t^{B'} \). Thus, \( t \in K \), a contradiction.

Therefore, we end up with a contradiction in all possible cases. Thus, the decentralized equilibrium \( A \) must have been efficient.

2. Let \( (A) = (A_s') \) be an efficient allocation of labor. (A) determines the marginal productivities \( F_t' = F_t' (A_t + \sum_{s \in S} A_s') \) and thus the wages of the firms. Let \( \emptyset \neq S \subseteq \{1, \ldots, n\} \) and let \( \max := \max \{F_i'| t \in S\} \). By efficiency of (A) and the remark in Section 3, all individuals in \( M_s \) work for firms \( t \in S \) such that \( F_t' = \max \).

This proves that the allocation (A) is a decentralized equilibrium. \( \Box \)

**Proof of Proposition 2**

1.\( \Rightarrow \)2. We will show that (not 2.) implies (not 1.). Let (A) be the original efficient allocation and assume the economy is not a region. Then (i) there are two firms that cannot be linked by a chain of directly connected firms or (ii) there exist binding mobility restrictions. We want to show that in both cases (i), (ii) there is a possibility to raise taxes that are not equal for all firms and that do not affect the efficiency of the allocation. For case (i) pick two such firms \( s \) and \( t \). Let \( M \) be the set of all firms that can be reached from \( s \) by a chain of directly connected firms. Then there is no \( S \subseteq \{1, \ldots, n\} \) with \( u \in S \cap M, v \in S \cap CM \) and \( A_s' > 0 \) (by CM we denote the complement of M). This is saying that there is no individual who could work for a firm in \( M \) and for a firm in \( CM \). It is then clear, that imposing any tax rate \( T \) on all the firms in \( M \) and no tax on the firms in \( CM \), will not affect the efficient allocation. For (ii) define \( F' := \max_{s \in \{1, \ldots, n\}} F_s^{A'} \). If there is a firm \( s \) such that \( F_{s}^{A'} < F' \) then let \( M \) be the set of all those firms whose marginal productivity in (A) is \( F' \). Let \( \tilde{F}' \) be the maximal marginal productivity of all firms not contained in \( M \). Then imposing the positive tax \( T_{M} := F' - \tilde{F}' \) on all the firms in \( M \) and no tax on the other firms will not affect the allocation of labor by the assumption before Proposition 2. In the remaining case, all the firms have marginal
productivity $F'$ in (A). By part (b) of the definition of binding mobility, there exists a nonempty set $W \neq \{1, \ldots, n\}$ such that for each $S \subset \{1, \ldots, n\}$ containing both elements of $W$ and $CW$ such that $A_S > 0$, we have $A'_S = 0$ for all firms $t$ in $W$. Thus imposing any tax $T > 0$ on all the firms in $W$ and no tax on the other firms will not have an impact on the resulting allocation.

2. $\Rightarrow$ 1. Let (A) be the original efficient allocation and (B) the decentralized equilibrium after the imposition of the taxes $T_t$ ($t = 1, \ldots, n$). By assumption, not all $T_t$ are equal. By Proposition 1, (B) is an efficient equilibrium for production functions of the firms given by $G_i(x) := F_i(x) - T_i \cdot x$. We show by contradiction that (B) is not efficient for the original production functions $F_i$.

Claim: There is a firm $v$ such that $F'_{v} \neq F'_{v'}$.

If not, let $W := \{t \in \{1, \ldots, n\} \mid T_t = \max_{u \in \{1, \ldots, n\}} T_u\}$ and let $s \notin W$. Assume $S \subset \{1, \ldots, n\}$ with $s, t \in S$ for a $t \in W$ and $s \notin W$, such that $A_S > 0$. Applying the FOC for an efficient equilibrium (4) in the case of production functions $G_i$, we have $F'_{u} - T_u - \lambda_i + \mu_{s} = 0$ for $u = s, t$ and thus $\mu_{s} - \mu_{s} = T_t - T_i > 0$. This implies $\mu_{s} > 0$ and thus by (5) that $B_{s}^{t} = 0$. Thus, we have showed that in the allocation (B) no individual who can work for a firm in $W$ and for a firm in $CW$ will work for the firm in $W$.

Now, there exists $s \notin W$ that is directly connected to a firm $t \in W$. Therefore, $S \subset \{1, \ldots, n\}$ with $s, t \in S$ such that $A_S > 0$.

Since no mobility restrictions are binding $S' \subset \{1, \ldots, n\}$ with $s', t' \in S'$, $s' \in CW, t' \in W$ such that $A_{S'} > 0$ and $A'_{S'} > 0$. But we have already showed that in the allocation (B) $B_{s'}^{t} = 0$. Thus $W$ faces a net loss of workers when changing from (A) to (B). Thus at least one firm in $W$ faces a net loss of workers which implies that its marginal productivity is going to increase, a contradiction. This proves the claim.

Let $W := \{t \in \{1, \ldots, n\} \mid F'_{v} - F'_{v'} = \max_{u \in \{1, \ldots, n\}} F'_{u} - F'_{u}\}$. It is $W \neq \{1, \ldots, n\}$. Let $s \notin W$ be any firm that is directly connected to a $t \in W$. It is $F'_{v'} > F'_{v}$ and thus $B_{s}^{t} = 0$ for any set $S$ containing both $s$ and $t$. Using the definition of binding mobility restrictions and repeating once more the above argument it follows that at least one firm $t'$ in $W$ faces a net loss of workers while one firm $s'$ outside $W$ faces a net gain. This implies $F'_{v'} > F'_{v}$, a contradiction. $\square$

Proof of Corollary 1

Let (A) be an efficient allocation and let $F'_{i_{1}} \leq F'_{i_{2}} \leq \cdots \leq F'_{i_{n}} \leq F'_{i_{n+1}}$ be the ordering of the firms according to their marginal productivities in that allocation. For each firm $s$, if $s = i_{s}$, impose the firm-specific tax $T_s := F'_{i_{s}} - F'_{i_{s}}$ on firm $s$. It is clear that the tax does not affect the allocation and that wages will be equalized after the tax is imposed. $\square$
Proof of Proposition 3

(a) We start with the ⇐ direction of the proof. We first show that condition 1. in the proposition implies efficiency.

Let (A) be the allocation of the workforce before the imposition of the tax. The following implications hold:

[There is firms s ∈ R and t ∈ Q such that after the tax is imposed an individual changes her job from s to t] ⇒ [There are firms s ∈ R and t ∈ Q and ∃S ⊂ {1, ..., n} with s, t ∈ S, A_s^t > 0 and F_s^t - T < F_s^t] ⇒ [T > T']

Negating these statements yields:

[T ≤ T'] ⇒ [For all firms s ∈ R and t ∈ Q and for all S ⊂ {1, ..., n} with s, t ∈ S, (F_s^t - T < F_s^t) implies (A_s^t = 0)] ⇒ [No individual changes her job from a firm s ∈ R to a firm t ∈ Q after the tax is imposed] ⇒ [The equilibrium after the imposition of the tax is efficient].

The very last implication follows from the fact that the decentralized equilibrium before imposition of the tax is efficient (Proposition 1). Therefore if no individual changes her job after imposition of the tax (the only possible changes of jobs after imposing a tax in R, are from firms in R to firms in Q) the resulting equilibrium with taxes is still efficient. We have shown so far that condition 1. of the proposition implies efficiency.

If condition 2. of the proposition holds, no individual who could have worked for firms in R and Q before imposition of the tax, actually worked in R. It is clear that after imposition of the tax in R, there is no reason why such an individual should now start working for a firm in R. Therefore, the same reasoning as before implies efficiency.

(b) ⇒ direction of the proof:

Let (A) be the allocation before imposition of the tax and (B) the decentralized equilibrium after the tax. Assume that both conditions 1. and 2. of the proposition do not hold. We want to show that (B) is inefficient. We will reason by contradiction. The proof consists of first showing that efficiency of (B) implies the following two statements (i) and (ii).

Assume that (B) is efficient. Then:

(i) Let s be a firm with strictly bigger marginal productivity in (B) than in (A) and let t be a firm that employs a nonzero mass of workers in (B) that in the allocation (A) worked for s. Then t’s marginal productivity in (B) must also be bigger than in (A). Formally:

Let F_s^t < F_s^t and assume ∃S ⊂ {1, ..., n} with s,t ∈ S and A_s^t > B_s^t and A_s^t < B_s^t, then F_s^t < F_s^t.

(ii) It is true that (∃s ∈ Q such that F_s^t > F_s^t) or (∃s ∈ R such that F_s^t < F_s^t).

Once (i) and (ii) are established we can finish the proof by the following reasoning: Define M := {s ∈ {1, ..., n} such that F_s^t < F_s^t}. Changing from (A) to (B) each firm in M loses a nonzero mass of workers. By (ii) M ≠ ∅ (this is obviously true if in (ii) ∃s ∈ R such that F_s^t < F_s^t). If ∃t ∈ Q F_s^t > F_s^t, t employs strictly more workers in (B), thus there must exist another firm t’ that
faces a net loss of workers and thus \( F^A_i < F^B_i \) and thus the firms in \( M \) face a net loss of workers. Thus \( \exists s \in M \) that employs workers in \( B \) coming from firms in \( M \). By (i) \( s \in M \), a contradiction. Thus (B) cannot be efficient.

Ad (i): By assumption, \( 0 \leq B^t_s < A^t_s \) (implying that \( F^A_i \leq F^B_i \) by efficiency of (A)) and \( 0 \leq A^t_s < B^t_s \) (implying that \( F^B_i \leq F^A_i \) by efficiency of (B)). We thus have that \( F^A_i \leq F^A_i < F^B_i \leq F^B_i \).

Ad (ii): If (ii) does not hold then

\[
\forall t \in Q, F^A_i \leq F^B_i \quad \text{and} \quad \forall s \in R, F^A_i \geq F^B_i \quad (*)
\]

Since condition 1. and 2. of the proposition are both violated by assumption, it follows that

\[
\exists s \in R, t \in Q, S \subset \{1, \ldots, n\}, s, t \in S, A^t_s > 0 \text{ and } F^A_i - T < F^A_i \quad (**)
\]

(B) is a decentralized equilibrium when workers at firm \( t \) gain wages \( F^A_i(x) - 1(t \in R) \cdot T \cdot x \). By Proposition 1, (B) is then also an efficient equilibrium, if the production functions of the firms are given by \( g_i(x) = F^A_i(x) - 1(t \in R) \cdot T \cdot x \). By Eq. (4) the FOC \( F^A_i - T - \lambda A_t^s + \mu^s_A = 0 \) and \( F^B_i - \lambda A_t^s + \mu^s_A = 0 \) hold. Thus \( F^A_i = F^B_i - T + \mu^s_A - \mu^s_B \). By (**), \( \mu^s_A - \mu^s_B > 0 \) implying \( \mu^s_A > 0 \) and thus by (5) \( B^t_s = 0 \).

Since \( A^t_s > 0 \) and \( B^t_s = 0 \), firm \( s \in R \) lost some workers (we do not claim that it is a net loss in the total workforce of firm \( s \)). In (B) these workers must be employed at a firm in \( Q \) (if not, they must be employed at a firm \( s' \in R \). Thus, \( \exists s' \in R, S \subset \{1, \ldots, n\}, s', s, t \in S, F^A_i - T < F^A_i, A^t_s > 0 \) and \( B^t_s > A^t_s \). This implies \( F^B_i - T \geq F^B_i \forall t \in S \cap Q \). But then \( F^A_i - T < F^B_i \leq F^B_i - T \) and \( F^A_i < F^B_i \leq F^B_i \) by (**), contradicting \( A^t_s > 0 \). On the other hand no worker employed for a firm \( t \in Q \) in (A) changes to a firm \( s \in R \) in (B). (If not, \( \exists s \in R, t \in Q, S \subset \{1, \ldots, n\}, s, t \in S, 0 = B^t_s < A^t_s \) (implying \( F^A_i \geq F^B_i \)) and \( 0 \leq A^t_s < B^t_s \) (implying \( F^B_i - T \geq F^B_i \)). By (*) we get \( F^A_i \leq F^A_i < F^B_i \leq F^B_i - T \) and thus \( F^B_i > F^A_i \geq F^B_i \), a contradiction). The last statements combined yield that, net, a strictly positive mass of workers changes from firms in \( R \) to firms in \( Q \). Thus, for some \( s \in R \) we have \( F^A_i < F^B_i \), which contradicts (**). □

Proof of Proposition 4

We know from Proposition 2 that both the sets \( H = \{ s \in \{1, \ldots, n\}, F^A_i < F^B_i \} \) and \( L = \{ s \in \{1, \ldots, n\}, F^A_i > F^B_i \} \) are nonempty. We have to show that \( E = \{ s \in \{1, \ldots, n\}, F^A_i = F^B_i \} \) is empty. Assume the contrary. Then there is three possibilities. Switching from the allocation (A) before the tax to the allocation (B) after the tax, there is (1) a positive net stream of workers (pns) from the firms in \( E \) to the ones in \( H \) and a negative net stream (nns) from \( E \) to \( L \) (2) nns from \( E \) to \( H \) and nns from \( E \) to \( L \) (3) there is a zero net stream of workers between \( E \) and \( H \) and between \( E \) and \( L \). Case (1) and (3) both imply that there is a pns from \( H \) to \( L \). This implies that there exist \( s \in H \) and \( t \in L \) and \( \exists s \in \{1, \ldots, n\} s, t \in S, B^t_s > 0. \)
This contradicts $F_t^{B_t} < F_s^{B_s}$. Similarly, case (2) implies that there exist $s \in H$ and $t \in E$ and $\exists S \subset \{1, \ldots, n\}$ s, $t \in S$, $B_t^S > 0$. This contradicts $F_t^{B_t} < F_s^{B_s}$. □

References