**Institut für
Volkswirtschaftslehre
und Statistik**

No. 593-00

Ethnic Dualism and Communication Costs – Explaining Segmentation and Wage Inertia

Ulf von Kalckreuth

**Beiträge zur
angewandten
Wirtschaftsforschung**

**Universität Mannheim
A5, 6
D-68131 Mannheim**

Ethnic Dualism and Communication Costs – Explaining Segmentation and Wage Inertia

Ulf von Kalckreuth

Economic Research Centre of the Deutsche Bundesbank

First Version: February 2000

This version: November 2000

Abstract:

What is the impact of international migration on the wage structure? Empirically, it is difficult to find any effect at all. This essay gives a new theoretical explanation for this conspicuous absence, emphasising non-convexities in the technology of individual firms due to communication costs. With high costs of coordination between workers, the labour force will segregate on the workshop level. In this case, the aggregate production technology has linear segments, and within certain bounds, additional labour input is absorbed without changing the relative factor prices. A local non-substitution theorem is derived.

Key Words:

Labour Markets, Ethnic Dualism, International Migration, Communication Costs, Segregation, Wage Structure

JEL Classification: D24; D33; J31; F22

Correspondence:

Deutsche Bundesbank, Economic Research Centre, P.O. Box 10 06 02, 60006 Frankfurt am Main, Germany. Tel: +49 69 9566 2217; e-mail: Ulf.von-Kalckreuth@bundesbank.de; Homepage: http://www.von-kalckreuth.de. Any opinions expressed are those of the author alone and should not be interpreted as the opinion of the Deutsche Bundesbank.

Ethnic Dualism and Communication Costs – Explaining Segmentation and Wage Inertia *

1. A Model of Ethnic Dualism: Introduction and Summary

In December 1999, the Council of the European Union decided to enlarge the group of candidates to membership. In addition to ongoing talks with Poland, Hungary, the Czech Republic, Estonia, Slovenia and Cyprus, the EU will open negotiations with Bulgaria, Latvia, Lithuania, Romania, Slovakia and Malta. Furthermore, the status of a candidate for membership is also conferred on Turkey, without immediate negotiations, though.

The enlargement process is being followed with particular attention in Germany. Membership of the EU means freedom to settle and to trade. Germany and Austria will naturally have to bear the brunt of any large immigration wave that follows. A large majority of all immigrants from South Eastern Europe already residing in the EU are living in one of these tangential countries. In addition, more than 2 million Turkish nationals living in Germany might act as a bridgehead for massive immigration from that region, too. Judging from the past, it is unlikely that these immigrants would integrate quickly. What would be the consequences for the German labour market? Are there any special features of "ethnic dualism"?

This essay puts forward a new way of thinking about dualistic structures, centring on ethnic differences and the resulting communication costs between agents. If two types of labour are not perfect substitutes, there are comparative advantages that may be exploited by division of labour. If there are no further interactions between members of the labour force, it will always be efficient to employ both types jointly in the same production unit. Dividing labour, however, means coordinating work effort and in order to do this, agents must communicate. This will not be costless, and communication between people from different cultural backgrounds is more difficult, takes more time and may involve a higher risk of misunderstanding and conflict than communication between people sharing the same culture.

High communication costs will lead to efficient sorting. There will be firms that specialise in employing mainly one type of labour. A dualistic industrial structure will result from the ethnic cleavage. Segregation, however, will probably not be complete: the marginal productivity of a factor might increase when it gets scarcer. Hence, it will usually be inefficient for firms to dispense with one type of labour completely.

---

* Any opinions expressed are personal and ought not to be interpreted as the opinions of the Deutsche Bundesbank. This paper partly draws on von Kalckreuth (1999a), Chap. 3. I owe a great debt to Jürgen Schröder, Martin Hellwig and Norbert Schulz for illuminating discussions and many valuable comments. A seminar audience in Mannheim and colleagues at the Bundesbank gave important feedback, thanks especially to Johannes Hoffmann. The responsibility for all remaining errors and omissions rests with me.

On the firm level, high communication costs render certain factor proportions inefficient and firms will never use them. This, however, makes *aggregate technology* linear within these bounds. If aggregate supply of labour changes, substitution will not take place by firms smoothly adapting their factor intensities to changing prices, but by two efficiently producing sectors adapting their output quantities. Ideally, no wage reaction at all is necessary to restore equilibrium.

Fig. 1 demonstrates the basic argument. The bold line shows the unit isoquant of a representative firm, factors A and B being the labour input from workers of different ethnic origin. Because of communication costs, technology on the firm level will be non-convex, and firms will not use the input vectors along the line $\mathbf{z}_{\mathbf{I}}$ to $\mathbf{z}_{\mathbf{II}}$. The *aggregate* input requirement set, however, is given by the convex hull of the representative firm's set. If aggregate endowment happens to lie in the cone defined by $\mathbf{z}_{\mathbf{I}}$ and $\mathbf{z}_{\mathbf{II}}$, these two unit activities will be used simultaneously in equilibrium. Changing aggregate factor supply will be accommodated by a changing intensity of their usage, and factor prices will not change, nor will the set of efficient activities. In other words: as a result of efficient sorting, the aggregate isoquant will have a linear portion.

2. An Empirical Puzzle: Why Do Wages Seem Invariant to Migration?

In order to measure the impact of migration on the wage structure of the recipient country, one ideally would like to observe random streams of migrants into otherwise closed labour markets. The work of Grossman (1982) was the first of a whole series of studies taking this approach. Typically, a city or metropolitan area is treated as a closed labour market and the researcher performs regressions between different wage rates on one hand and the number of migrants on the other. Quite regularly, the outcome indicates that migration has *almost no effect on native wages.* Borjas (1994) and Greenwood and McDowel (1986) survey the literature on the US labour market, see, for example, Borjas (1990) or Altonij and Card (1991). The point estimates of the elasticity of wage rates to immigration cluster around -0.01 to -0.02. If a city has 10 per cent more immigrants than another, the native wage in that city is only about 0.2 per cent lower. This numerically weak relationship is "observed across all types of native workers, white or black, skilled or unskilled, male or female".[1] Zimmermann and Bauer (1999) give an extended review of the relatively recent literature on West European labour markets. Overwhelmingly, these studies conclude that the wage effects of immigration are negligible or non-existent and – in some cases – even positive. Examples are the contributions by Pischke and Velling (1994), Winter-Ebmer and Zweimüller (1996), Velling (1995), Bauer (1998) and the comprehensive study by Boeri and Brückner (2001). Taking an entirely

---

1 Borjas (1994), p. 1697.

different approach, Gang and Rivera-Batiz (1994) reach the same conclusions. De New and Zimmermann (1994) find substantial negative effects, but seem to remain the exception.

It has been argued by Borjas and others that migration flows are not exogenous and the observed labour markets are not closed. Migrants have strong incentives to choose that local labour market which offers the best economic perspectives. Natives, for their part, may move to neighbouring labour markets and thereby mask the aggregate significance of migration. If natives and migrants eliminate the differences between local labour markets, the observed correlations have no structural interpretation. This problem cannot be solved completely by using instrumental variable techniques.

Card (1990) reports a convincing natural experiment that calls into question the validity of this argument. Probably in order to put political pressure on the US, Fidel Castro unexpectedly declared on April 20, 1980 that anyone who wished to do so could leave Cuba via the port of Mariel. Between the months of May and September 1980, about 125,000 Cuban immigrants reached the port of Miami in a fleet of small boats. About half of them stayed permanently in Miami. Almost overnight, the city's labour force had grown by 7%, and the number of Cuban migrants by 20%.

These new migrants were mostly unskilled and barely spoke any English.[2] In his study, Card traces the development of the wage structure in Miami and four neighbouring cities. Table 2 gives a first impression. For all ethnic groups, wages in Miami were lower than in comparable cities. The incomes of Blacks, of Whites and of Hispanics (without Cubans) stayed constant in Miami, whereas they tended to fall in the control group.

The wages of Cuban immigrants did decline, by 7% between 1979 and 1981. However, Card shows that this is almost entirely the consequence of a particularly low skill level on the part of the Mariel immigrants, whose wages were about 34% less than wages of other Cuban immigrants. In effect, the newcomers diluted the skill level of the entire group. Card undertakes a careful analysis of the wage dynamics, controlling for the changing qualification structure and finally states that a comparison of Cuban wages inside and outside the Miami labour market does not provide any evidence of a widening wage differential during the years following the influx of the Mariel immigrants. There was no worsening of the earnings of native low skilled workers either.

A violent expansion of the labour force by 7% due to the influx of almost exclusively low skilled immigrants from one single source country had no effect on wages or unemployment rates! Of course, it may be the case that earnings of less qualified workers and Hispanics in general and Cubans in particular *would have* increased, had the Mariel incident not taken place. Card's evidence does not point in this direction. His explanation centres on special fea-

2 It even seems that Castro used this opportunity to empty the country's prisons and mental hospitals.

tures of the labour market of Miami. As a result of earlier waves of immigrants, an industry structure had evolved that not only was *tailored to the skills* of Hispanic immigrants but made *communication costs irrelevant:*

> Two factors that may have been especially important in facilitating the absorption of the Mariel immigrants are related to the distinctive character of the Miami labour market. First, Miami's industry structure was well suited to make use of an influx of unskilled labour. This structure, and particularly the high concentration of textile and apparel industries, evolved over the previous two decades in response to earlier waves of immigrants and may have allowed the Mariel immigrants to take up unskilled jobs as earlier Cuban immigrants moved to better ones. Second, because of the high concentration of Hispanics in Miami, the lack of English-speaking ability among the Mariels may have had smaller effects than could be expected for other immigrants in other cities. [3]

The following sections will work out in a stylised, but formal way what the *interplay of comparative advantages and communication costs* might mean for the industry structure and its capacity to absorb migrants.

3. Communication Costs and Convexity

A convincing analytical foundation for the convexity assumption in labour economics is presented in an essay by Rosen (1978) on the division of labour, refining a model of international trade by Dornbusch, Fisher and Samuelson (1977). If there are a variety of tasks that must be assigned to two different types of workers, these tasks can be ordered by decreasing comparative advantages of one group with respect to the other. An efficient work assignment involves allocating tasks to workers in the order indicated by their comparative advantage. Consider a movement down the isoquant, where successively workers of type A are substituted against workers of type B. In order to replace one unit of A labour, ever more units of B are needed as the comparative advantage of type B for the last task becomes smaller and smaller. Thus, the rate of technical substitution falls. If there is a continuum of tasks to be performed, and comparative advantages are described by a twice-differentiable function, the familiar neoclassical, strictly convex and differentiable isoquant obtains.

Rosen's theoretical account abstracts from the need to coordinate and other interactions between workers. Becker and Murphy (1992) point out that the various costs resulting from the co-ordination of specialised activities in effect impose a limit to the realisation of gains from labour division. This repeats an argument made earlier by McManus (1985) in his fundamental study on the costs of linguistic heterogeneity: if the languages of two groups of agents differ, they are unable to reap all the efficiency gains that could be realised by labour division, given their different skill endowment.

Linguistic deficits reduce the productivity of workers. McManus (1985) investigates the earning losses for Hispanics in the USA and estimates their present value to be $36,000 for the most disadvantaged group. Kossoudij (1988) and Chiswick (1991) also detect high losses

[3] Card (1990), p. 257.

accruing to immigrants due to their lack of proficiency in English. Lazear (1999) points out that the losses from not knowing the majority language might be inversely related to the relative size of a minority group.

In order to analyse the consequences of ethnic dualism on the labour market, it is crucial to see that the convexity of technology on the firm level will be lost if the costs of communication between workers of two different types are sufficiently high. If workers are separated at different production sites – without any labour division whatsoever – there will be no costs of communication resulting from cultural interferences either. These will occur as soon as workers of different ethnic type are employed jointly in the same production unit. If the costs of co-ordination are sufficiently high, the maximum output resulting from labour division is not higher, but lower than the output attainable at separate productions sites. The efficiency gains resulting from the division of labour are more than outweighed.

In order to model ethnic dualism and explain the low elasticity of wages, we employ a simple but powerful basic hypothesis: Between workers of different ethnicity there are co-ordination costs that do not exist between members of the same group.

4. A Model of Ethnic Division

4.1 Production Plans and Production Sets

Suppose a single consumption good is produced by firms $\mathrm{j}=1,\dots,\mathrm{J}$, with $\mathrm{J}\geq 2$. The *production plan* of firm j is characterised by:

$$\mathbf{y}_{\mathrm{j}} \in \mathrm{R}^{3} \text{ , with } \mathbf{y}_{\mathrm{j}} = \left(Q_{\mathrm{j}}, -A_{\mathrm{j}}, -B_{\mathrm{j}}\right) .$$

Here, Q_{j} is the output. The entries A_{j} and B_{j} are quantities of labour input by workers of nationality A and B, respectively. Following the convention, they enter with a negative sign. Using another vector, the *input*:

$$\mathbf{z}_{\mathrm{j}} \in \mathrm{R}^{2} \text{ , with } \mathbf{z}_{\mathrm{j}} = \left(A_{\mathrm{j}}, B_{\mathrm{j}}\right),$$

we can write the production plan as $\left(Q_{\mathrm{j}}, -\mathbf{z}_{\mathrm{j}}\right)$.

All producers use the same technology. The set of production plans available to producers is their common production set Y^{N}. A production plan in Y^{N} is called *technologically feasible.* The production set can be described by a *production function,* $\mathrm{f}^{\mathrm{N}} : \boldsymbol{\Omega}^{2} \rightarrow \mathrm{R}$. The set $\boldsymbol{\Omega}^{2}$ is the nonnegative orthant, i.e., the production function is defined for all $\mathbf{z}_{\mathrm{j}} \geq (0,0)$. A production plan $\mathbf{y}_{\mathrm{j}} = \left(Q_{\mathrm{j}}, -\mathbf{z}_{\mathrm{j}}\right)$ is contained in Y^{N} if and only if $Q_{\mathrm{j}} \leq \mathrm{f}^{\mathrm{N}}\left(\mathbf{z}_{\mathrm{j}}\right)$, in short:

$$\mathrm{Y}^{\mathrm{N}} = \{\mathbf{y}_{\mathrm{j}} \in \mathrm{R}^{3} \mid Q_{\mathrm{j}} \leq \mathrm{f}^{\mathrm{N}}\left(\mathbf{z}_{\mathrm{j}}\right)\} \text{ .} \tag{1}$$

For every output $Q_{\mathrm{j}} \in \mathrm{R}$, there is an *input requirement set* $\mathrm{V}^{\mathrm{N}}\left(Q_{\mathrm{j}}\right)$. This is the set of all inputs $\mathbf{z}_{\mathrm{j}}$ sufficient to produce Q_{j}, in other words:

$$\mathrm{V}^{\mathrm{N}}\left(Q_{\mathrm{j}}\right) = \{\mathbf{z}_{\mathrm{j}} \in \mathrm{R}^{2} \mid Q_{\mathrm{j}} \leq \mathrm{f}^{\mathrm{N}}\left(\mathbf{z}_{\mathrm{j}}\right)\} \text{ .}$$

The boundary of the input requirement set is the *isoquant* for the output Q_j.

4.2 Communication Costs

We assume that the individual production function can be written more specifically in the following way:

$$f^N(\mathbf{z}_j) = f^G(\mathbf{z}_j) - \kappa c(\mathbf{z}_j) \ .$$

The term $f^G(\mathbf{z}_j)$ is interpreted as the gross output that would be attainable in the absence of communication costs. The function $f^G : \Omega^2 \to R$ is continuous, linear homogeneous and concave. Moreover, both types of workers are assumed to be non-essential, so that $f^G(\mathbf{z}_j) > 0$ for $\mathbf{z}_j = (1,0)$, as well as for $\mathbf{z}_j = (0,1)$. The output losses caused by communication costs are denoted by $\kappa c(\mathbf{z}_j)$. The function $c : \Omega^2 \to R$ indicates the efficiency losses resulting for both types of labour from the necessity to cooperate with workers of the other type.[4] The scalar $\kappa > 0$ serves to parameterise the level of communication costs.

The function $c(\cdot)$ is continuous and linear homogeneous. Apart from this, our assumptions concerning communication costs are very general.[5] If only one type of worker is employed, there will not be any output loss due to cultural interferences. We therefore assume $c(\mathbf{z}_j) = 0$ for all $\mathbf{z}_j = (A_j, 0)$, as well as for all $\mathbf{z}_j = (0, B_j)$. On the other hand, if both types of labour are used in the same production process, there will be output losses caused by the necessity to coordinate. This is captured by assuming $c(\mathbf{z}_j) > 0$ for all $\mathbf{z}_j >> \mathbf{0}$.

If we consider two inputs $\mathbf{z}_j' = (A_j, 0)$ and $\mathbf{z}_j'' = (0, B_j)$ with $A_j, B_j > 0$, and any $\mathbf{z}_j = \lambda \mathbf{z}_j' + (1-\lambda)\mathbf{z}_j''$, we have

$$f^N(\lambda \mathbf{z}_j' + (1-\lambda)\mathbf{z}_j'') < \lambda f^N(\mathbf{z}_j') + (1-\lambda) f^N(\mathbf{z}_j'') \ , \tag{2}$$

whenever

$$\kappa > \frac{f^N(\mathbf{z}_j) - \left(\lambda f^N(\mathbf{z}_j') + (1-\lambda) f^N(\mathbf{z}_j'')\right)}{c(\mathbf{z}_j)} > 0 \quad . \tag{3}$$

For sufficiently high values of κ, therefore, the net production function will no longer be concave. Below we will always assume that κ is high enough for this to be the case.

For lack of knowledge, we have put only few restrictions on the nature of communication costs, and this carries over to the net production function. It is true that f^N is continuous and

---

4 It may be presumed that the efficiency of a given type of worker will be smaller if the share of the other type becomes higher. The investigation of McManus (1990) supports this hypothesis: the income loss for a member of the Hispanic minority resulting from not speaking English is a decreasing function of the percentage of his own ethnic group in the district.

5 Statements on monotonicity are not easily made. If the labour input of A-labour rises, the efficiency of B-labour decreases. The efficiency of M-labour itself, on the other hand, might increase, and in addition the relative weights change. Statements on concavity are equally difficult: strong assumptions on the nature of social interactions are needed, like perfect mixing, for example.

linearly homogeneous. But the function will be neither necessarily monotonous, nor differentiable, nor even positive throughout. Fig. 2 shows on the left a standard, linear homogeneous concave gross production function of the CES-type and on the right a non-concave net production function, resulting from sufficiently high communication costs. Fig. 3 shows some systems of isoquants that are consistent with our restrictions on technology.

The properties of the production set are easily derived from the characteristics of the production function f^{N}. The set Y^{N} is closed and represents a single output technology: For all $\left(Q_{\mathrm{j}},-\mathbf{z}_{\mathrm{j}}\right)\in \mathrm{Y}^{\mathrm{N}}$ we have $\mathbf{z}_{\mathrm{j}} \geq \mathbf{0}$. By (1) we have assumed free disposal of output: if a plan $\left(Q_{\mathrm{j}},-\mathbf{z}_{\mathrm{j}}\right)$ is contained in Y^{N}, the same will be true for any plan $\left(Q_{\mathrm{j}}',-\mathbf{z}_{\mathrm{j}}\right)$ with $Q_{\mathrm{j}}'< Q_{\mathrm{j}}$. As f^{N} is linearly homogeneous, Y^{N} exhibits constant returns to scale: from $\mathbf{y}_{\mathrm{j}} \in \mathrm{Y}^{\mathrm{N}}$ follows $k\mathbf{y}_{\mathrm{j}} \in \mathrm{Y}^{\mathrm{N}}$ for all $k \geq 0$. This also includes the possibility of inaction: the plan $\mathbf{y}_{\mathrm{j}} = \mathbf{0}$, the so-called zero-activity, is feasible. Geometrically, Y^{N} is a cone with vertex at the origin. Factors are productive and non-essential. In the set Y^{N}, production plans yielding a positive output exist for $\mathbf{z}_{\mathrm{j}} = (1,0)$ as well as $\mathbf{z}_{\mathrm{j}} = (0,1)$. Ultimately, the technology is non-convex. There are plans $\mathbf{y}_{\mathrm{j}}'$ and $\mathbf{y}_{\mathrm{j}}''$ such that $\lambda\mathbf{y}_{\mathrm{j}}'+(1-\lambda)\mathbf{y}_{\mathrm{j}}''\notin \mathrm{Y}^{\mathrm{N}}$ for some scalar $\lambda\in \left]0,1\right[$. This property is crucial for our explanation of ethnically segregated economic sectors.

4.3 Price System and Profits

If p_{Q} represents the price of output and $p_{\mathrm{A}}, p_{\mathrm{B}}$ are the wages for A-labour and B-labour, a price system is defined by the vector:

$$\mathbf{p}\in \mathrm{R}^3 \qquad \text{with} \qquad \mathbf{p} = \left(p_{\mathrm{Q}}, p_{\mathrm{A}}, p_{\mathrm{B}}\right) .$$

The value of plan $\mathbf{y}_{\mathrm{j}}$ with respect to the price system $\mathbf{p}$ is the inner product of $\mathbf{p}$ and $\mathbf{y}_{\mathrm{j}}$:

$$\mathbf{p}\mathbf{y}_{\mathrm{j}} = p_{\mathrm{Q}}Q_{\mathrm{j}} - p_{\mathrm{A}}A_{\mathrm{j}} - p_{\mathrm{B}}B_{\mathrm{j}} \quad .$$

The producers consider the price system as given. A profit-maximising plan solves the problem

$$\underset{\mathbf{y}_{\mathrm{j}}\in \mathrm{Y}^{\mathrm{N}}}{\mathrm{Max}}\, \mathbf{p}\mathbf{y}_{\mathrm{j}} \ .$$

Such a plan will not exist for every price system. For a given $\mathbf{p}$, the value of an arbitrary production plan will be positive, equal to zero or negative. If $\mathbf{p}\mathbf{y}_{\mathrm{j}} > 0$, the same will be true for all production plans $k\mathbf{y}_{\mathrm{j}}, k \geq 0$ on the same ray through the origin, and potential profit is unlimited. With constant returns, a solution to the problem of profit maximisation will exist if and only if

$$\mathbf{p}\mathbf{y}_{\mathrm{j}} \leq 0 \text{ for all } \mathbf{y}_{\mathrm{j}} \in \mathrm{Y}^{\mathrm{N}} \quad . \tag{4}$$

The zero activity yields a profit of zero for any price system, so maximal profits cannot be negative. Thus, if $\mathbf{y}_{\mathrm{j}}^*$ is a solution to the problem of profit maximisation, the standard *zero profit condition* must hold:

$$\mathbf{p}\mathbf{y}_j{}^* = 0 \ . \tag{5}$$

In fact, conditions (4) and (5) are necessary and sufficient for a plan to be optimal. The first condition makes sure that a maximum exists. If it holds, any plan is optimal that satisfies the second condition.

4.4 Factor Supply and Factor Market Equilibrium

The model is closed by the assumption that the economy's factor endowment, the stock $\bar{\mathbf{z}} = (\bar{z}_A, \bar{z}_B)$, belongs to consumers who do not draw direct utility from their labour endowment, whereas their preferences for the consumer good are insatiable. The profits of the producers ultimately also accrue to the consumers.

An ordered set of feasible plans for the producers, $(\mathbf{y}_1,\ldots,\mathbf{y}_J)$ is a factor allocation. A factor market equilibrium is defined as follows:

<u>Definition 1:</u> *A price system* $\mathbf{p}^* >> 0$ *and a factor allocation* $(\mathbf{y}_1{}^*,\ldots,\mathbf{y}_J{}^*)$ *with* $\mathbf{y}_j{}^* = (Q_j{}^*, -\mathbf{z}_j{}^*) \in Y^N$ *is a factor market equilibrium, if*

(1) *for every producer* j, *the plan* $\mathbf{y}_j{}^*$ *solves* $\max\limits_{\mathbf{y}_j \in Y^N} \mathbf{p}\mathbf{y}_j$ *, and* (6)

(2) *the factor market clears, i.e.:* $\sum\limits_{j=1}^{J} \mathbf{z}_j{}^* = \bar{\mathbf{z}}$. (7)

As only one good is produced, the role of the consumers is limited to ensuring the clearing of the factor market and guaranteeing a strictly positive price system. A price $p_Q \le 0$ is ruled out by the consumers being insatiable. If p_Q is positive, $p_A \le 0$ and $p_B \le 0$ are not possible. Both factors are productive and nonessential, and if one of their prices were negative, producers could make arbitrary large profits, and their demand would be unbounded. Thus, *if* an equilibrium exists, $\mathbf{p}^* >> 0$ must necessarily hold. The entire factor endowment will be supplied and Walras' law makes the market for output goods clear as well.

The investigation ultimately centres on the equilibrium structure of production, i.e. the properties of the production processes used in equilibrium. To this end, it is useful first to prove the *existence* of equilibrium and then its *efficiency*. Having done this, it is surprisingly easy to infer the properties of the production structure.

5. Existence and Efficiency

5.1 The Aggregate Production Set

If we consider a list of feasible production plans $(\mathbf{y}_1,\ldots,\mathbf{y}_J)$ and sum up, we obtain the aggregate production vector $\mathbf{y}$, consisting of an aggregate output Q and an aggregate input $\mathbf{z}$:

$$\mathbf{y} = \sum_{j=1}^{J} \mathbf{y}_j = (Q, -\mathbf{z}) \ .$$

The set of all feasible aggregate production vectors is the aggregate production set Y. It is the sum of the production sets Y^{N} for the individual producers:

$$\mathrm{Y} = \mathrm{Y}^{\mathrm{N}} + \ldots + \mathrm{Y}^{\mathrm{N}} = \left\{ \mathbf{y} \in \mathrm{R}^{3} \middle| \mathbf{y} = \sum_{\mathrm{j}=1}^{\mathrm{J}} \mathbf{y}_{\mathrm{j}}, \text{with } \mathbf{y}_{\mathrm{j}} \in \mathrm{Y}^{\mathrm{N}} \text{ for all j} \right\} .$$

As a counterpart to the production function f^{N} for the individual firm, we define the aggregate production function $\mathrm{f} : \Omega^{2} \to \mathrm{R}$. For any given aggregate input, it indicates the maximum aggregate output that can supplied by the producers as a whole:

$$\mathrm{f}(\mathbf{z}) = \max_{\mathbf{z}_1,\ldots,\mathbf{z}_{\mathrm{J}}} \sum_{\mathrm{j}=1}^{\mathrm{J}} \mathrm{f}^{\mathrm{N}}(\mathbf{z}_{\mathrm{j}}) \quad \text{s.t. } \mathbf{z}_{\mathrm{j}} \geq 0 \forall \mathrm{j} \quad \text{and} \quad \sum_{\mathrm{j}=1}^{\mathrm{J}} \mathbf{z}_{\mathrm{j}} \leq \mathbf{z} \quad .$$

The constraints define a bounded and closed set in R^{J}. Furthermore, f^{N} is continuous and the Weierstrass theorem guarantees the existence of a maximum for any given $\mathbf{z} \geq \mathbf{0}$. Hence the function f is everywhere defined.

We will see that with constant returns, the aggregate production set is convex, even if the technology of the individual producers is not. To this end, we first show that any aggregate production vector can be written as the sum of not more than two individual production plans.

Lemma 1: *Consider any number of production plans $\mathbf{y}_1,\ldots,\mathbf{y}_{\mathrm{n}}$, $n \geq 2$ in Y^{N}. Under the given assumptions, there are two technologically feasible plans $\mathbf{y}_{\mathrm{I}}$ and $\mathbf{y}_{\mathrm{II}}$, such that*

$$\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathbf{y}_{\mathrm{j}} = \mathbf{y}_{\mathrm{I}} + \mathbf{y}_{\mathrm{II}} .$$

Proof: The sum $\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathbf{y}_{\mathrm{j}} = \sum_{\mathrm{j}=1}^{\mathrm{n}} (Q_{\mathrm{j}}, -\mathbf{z}_{\mathrm{j}}) \equiv (Q, -\mathbf{z})$ is a convex combination of feasible plans: $(Q, -\mathbf{z}) \equiv \sum_{\mathrm{j}=1}^{\mathrm{n}} \frac{Q_{\mathrm{j}}}{Q} (Q, -\hat{\mathbf{z}}_{\mathrm{j}})$, with $\hat{\mathbf{z}}_{\mathrm{j}} = \frac{Q}{Q_{\mathrm{j}}} \mathbf{z}_{\mathrm{j}}$. By construction, every $\hat{\mathbf{z}}_{\mathrm{j}}$ belongs to the input requirement set $\mathrm{V}^{\mathrm{N}}(Q)$. The aggregate input is a convex combination of the $\hat{\mathbf{z}}_{\mathrm{j}}$. It lies in the polyhedron PO spanned by the inputs $\hat{\mathbf{z}}_{\mathrm{j}} \in \mathrm{V}^{\mathrm{N}}(Q)$. The construction is illustrated by Fig. 4. We consider the line that connects $\mathbf{z}$ and the origin, $\mathrm{S} = \{ s\mathbf{z} | 0 \leq s \leq 1 \}$. Line S necessarily contains points on the border of PO. Any such point can be written as a convex combination of at most *two* of the points that span the polyhedron. There will be scalars $s \in [0,1]$, $\lambda \in [0,1]$ and inputs $\hat{\mathbf{z}}_{\mathrm{i}}, \hat{\mathbf{z}}_{\mathrm{k}}$ in $\mathrm{V}^{\mathrm{N}}(Q)$, such that $s\mathbf{z} = \lambda \hat{\mathbf{z}}_{\mathrm{i}} + (1-\lambda) \hat{\mathbf{z}}_{\mathrm{k}}$. This means: $\lambda \frac{1}{s} \hat{\mathbf{z}}_{\mathrm{i}} + (1-\lambda) \frac{1}{s} \hat{\mathbf{z}}_{\mathrm{k}} = \mathbf{z}$. The inputs $\lambda \frac{1}{s} \hat{\mathbf{z}}_{\mathrm{i}}$ and $(1-\lambda) \frac{1}{s} \hat{\mathbf{z}}_{\mathrm{k}}$ are sufficient to produce an amount $\frac{1}{s} Q \geq Q$. Because of free disposal, Q can also be produced. Thus we can write the aggregate production as the sum of two production plans both contained in Y^{N}: $(Q, -\mathbf{z}) = \lambda \left(Q, -\frac{1}{s} \hat{\mathbf{z}}_{\mathbf{i}} \right) + (1-\lambda) \left(Q, -\frac{1}{s} \hat{\mathbf{z}}_{\mathbf{k}} \right)$. ■

The Lemma directly leads to

Proposition 1: *Under the assumptions on Y^{N} and for $J \geq 2$, the aggregate production set is convex.*

Proof: Let $\mathbf{y}'$ and $\mathbf{y}''$ be any aggregate production vectors in Y. The convex combination $\mathbf{y} = \lambda\mathbf{y}' + (1-\lambda)\mathbf{y}''$ is the sum of $n = 2J$ production plans in Y^{N}. Lemma 1 states that there are two plans $\mathbf{y}_{\mathbf{I}}$ and $\mathbf{y}_{\mathbf{II}}$ with $\mathbf{y} = \mathbf{y}_{\mathbf{I}} + \mathbf{y}_{\mathbf{II}}$. Therefore $\mathbf{y} \in \mathrm{Y}$. ■

There is a close connection between sets of profit-maximising individual production plans, and profit-maximising aggregate production vectors:

Lemma 2:[6] Let $\mathbf{y}_1, \ldots, \mathbf{y}_J$ *be production plans in* Y^{N}. *For any given* $\mathbf{p}$, *the aggregate production vector* $\mathbf{y} = \sum_{j=1}^{J} \mathbf{y}_j$ *maximizes aggregate profit* $\mathbf{py}$ *in* Y *if and only if* $\mathbf{y}_j$ *is profit maximising in* Y^{N} *for every producer.* ■

If a factor allocation maximises aggregate profits, every production plan involved will maximise the individual profit, and vice versa. As we search for an equilibrium, we can proceed in two steps. First, we can interpret the whole sector as a single, profit-maximising firm that nonetheless acts as a price taker. Then, having found an equilibrium production vector and an appropriate price system, we can decentralise the allocation.

5.2 The Efficiency of Equilibrium

An aggregate production vector is efficient if there is no $\mathbf{y}' \in \mathrm{Y}$ such that $\mathbf{y}' > \mathbf{y}$, i.e. if it is not possible to have a higher output using the same input and if the same output cannot be produced with less input. It is readily shown that if an aggregate production vector $\mathbf{y}$ maximises profits in Y for any price system $\mathbf{p} >> \mathbf{0}$, this $\mathbf{y}$ must be efficient.[7] Together with Lemma 1, this leads to a stripped down version of the first fundamental theorem of welfare economics:

Lemma 3: *If* $(\mathbf{y}_1^*, \ldots, \mathbf{y}_J^*)$ *and* $\mathbf{p}^* >> 0$ *is a factor market equilibrium, then the aggregate production vector* $\sum_{j=1}^{J} \mathbf{y}_j^*$ *is efficient.*

5.3 The Existence of Equilibrium

The search for equilibrium can thus be limited to allocations that lead to efficient aggregate productions. First it has to be checked whether there is an aggregate production vector that is both efficient and uses up the entire endowment. Then we have to ask for the existence of a price system $\mathbf{p}^*$ that supports the efficient aggregate production vector. If the answers to both questions are positive, we can conclude that an equilibrium exists.[8] The first question is addressed by

---

6 See e.g. Debreu (1959).

7 See e.g. Mas-Collel et alt. (1995), S. 150.

8 Existence can also be shown by invoking Debreu's (1959) theorem on the existence of equilibria in competitive economies. The route taken here is analytically less demanding and yields valuable qualitative information on the nature of equilibrium at the same time.

Lemma 4: *Under the given assumptions on* Y^{N}*, the aggregate production set contains an efficient production vector* $\overline{\mathbf{y}} = \left(\overline{Q}, -\overline{\mathbf{z}}\right)$*, with* $\overline{Q} = \mathrm{f}\left(\overline{\mathbf{z}}\right)$ *and* $\overline{\mathbf{z}} >> \mathbf{0}$ *being the aggregate factor endowment.*

Proof: For every $\mathbf{z}$ – and for $\mathbf{z} = \overline{\mathbf{z}}$ in particular – the maximum attainable output exists. Consider the aggregate production vector $\overline{\mathbf{y}} = \left(\overline{Q}, -\overline{\mathbf{z}}\right)$ with $\overline{Q} = \mathrm{f}\left(\overline{\mathbf{z}}\right)$. It is efficient if the same output cannot be produced using an aggregate input $\mathbf{z} < \overline{\mathbf{z}}$. Now if less input could be used, a quantity $\Delta\mathbf{z} = \overline{\mathbf{z}} - \mathbf{z} > 0$ would be redundant. Because all factors are productive and nonessential, further producers $\mathrm{J}+1,\ \mathrm{J}+2, \ldots$ could produce additional output using $\Delta\mathbf{z}$. Lemma 1, however, guarantees that an aggregate production vector based on an arbitrary number of feasible production plans can be realised by just 2 producers. So if $\overline{Q}$ is the maximum output given $\overline{\mathbf{z}}$, there can be no redundant factor quantities. ■

As every factor market equilibrium is efficient and, furthermore, one and only one efficient aggregate output vector exists, we have to look for a price system $\mathbf{p}^*$ that makes $\overline{\mathbf{y}}$ a profit-maximising vector. Part a) of the following Lemma is a version of the Second Fundamental Theorem of Welfare Economics. The proof is well known and rests on the Separation Theorem for convex sets.[9] Part b) uses the properties of the technology assumed here:

Lemma 5: *Let* Y *be a convex production set and* $\overline{\mathbf{y}}$ *be an efficient aggregate production vector in* Y.

a) Then a price system $\mathbf{p} > \mathbf{0}$ *exists, for which* $\overline{\mathbf{y}}$ *maximises aggregate profits.*

b) If factors are productive and nonessential, returns to scale are constant, and $\overline{\mathbf{y}} >> 0$*, this price system satisfies* $\mathbf{p} >> \mathbf{0}$.

Proof for part b): Let $\mathbf{p} = \left(p_{\mathrm{Q}}, p_{\mathrm{A}}, p_{\mathrm{B}}\right) > 0$ be a price system that makes $\overline{\mathbf{y}}$ a profit-maximising production vector. Then p_{Q} must be positive, as for $\mathbf{p} = \left(0, p_{\mathrm{A}}, p_{\mathrm{B}}\right) > 0$, the use of $\mathbf{z} >> \mathbf{0}$ would not be profit-maximising. But then *both* factor prices must be positive. None of the factors is essential: if one of them were free, producers could always increase profits by using more of it. ■

This leads immediately to

Proposition 2: *Under the given assumptions on* Y^{N}*, for any aggregate endowment* $\overline{\mathbf{z}}$ *a factor market equilibrium, consisting of an allocation* $\left(\mathbf{y}_1^*, \ldots, \mathbf{y}_{\mathrm{J}}^*\right)$ *and a price system* $\mathbf{p}^* >> 0$ *exists. The aggregate production vector* $\sum \mathbf{y}_{\mathrm{j}}^*$ *is efficient.*

Proof: Lemma 2 limits the search for equilibrium to those allocations that lead to efficient aggregate production vectors. Lemma 3 shows that there is exactly one efficient aggregate production vector $\overline{\mathbf{y}}$ that uses $\overline{\mathbf{z}}$. Lemma 4 guarantees, together with Proposition 1, that there is a price system $\mathbf{p}^* >> 0$ for which $\overline{\mathbf{y}}$ maximises aggregate profit. If $\mathbf{A}^* = \left(\mathbf{y}_1^*, \ldots, \mathbf{y}_{\mathrm{J}}^*\right)$ is an allocation with $\sum_{\mathrm{j}=1}^{\mathrm{J}} \mathbf{y}_{\mathrm{j}}^* = \overline{\mathbf{y}}$, then for $\mathbf{p}^*$ every single production plan must be profit-maximising. Every such pair $\left(\mathbf{p}^*, \mathbf{A}^*\right)$ is thus a factor market equilibrium. ■

---

9 See e.g. Mas-Collel et al. (1995), p. 151, Takayama (1985), p. 56, or Koopmans (1957), p. 88.

6. Uniform and Multiple Productive Structures

The aggregate production set is convex, but not so the production sets of the individual producers. What are the consequences? The set Y contains the individual technology Y^{N} entirely. But not vice versa. In the aggregate, production vectors are possible that are not feasible individually. If any such aggregate production vector is realised, then the individual plans *must* differ from each other by more than just a scaling factor.

We call *unit activity* any feasible production plan that leads to one unit of output: $\mathbf{y}_{\mathrm{j}}^{\mathrm{u}} = (1, -\mathbf{z}_{\mathrm{j}}^{\mathrm{u}})$, with $\mathbf{z}_{\mathrm{j}}^{\mathrm{u}} > 0$. Any production plan can be written as a scalar product $\mathbf{y}_{\mathrm{j}} = q_{\mathrm{j}}\mathbf{y}_{\mathrm{j}}^{\mathrm{u}}$ of the quantity of output and a unit activity. In describing an allocation $\mathbf{A} = (q_1\mathbf{y}_1^{\mathrm{u}}, \ldots, q_{\mathrm{J}}\mathbf{y}_{\mathrm{J}}^{\mathrm{u}})$, we speak of a *uniform productive structure* if all producers use the same unit activity, i.e. if $\mathbf{y}_1^{\mathrm{u}} = \ldots = \mathbf{y}_{\mathrm{J}}^{\mathrm{u}}$. A *multiple productive structure,* on the other hand, is realised if the allocation makes use of at least two different unit activities. Let $\overline{\mathbf{y}} = (\overline{Q}, -\overline{\mathbf{z}})$ be the efficient aggregate production vector, with $\overline{Q} = \mathrm{f}(\overline{\mathbf{z}})$. Then we can make the following case distinction:

<u>Case 1:</u> $\mathrm{f}^{\mathrm{N}}(\overline{\mathbf{z}}) = \mathrm{f}(\overline{\mathbf{z}}) = \overline{Q}$.

That means, $\overline{\mathbf{y}} \in \mathrm{Y}^{\mathrm{N}}$. Factors can be efficiently utilised by individual producers in the proportion given by $\overline{\mathbf{z}}$. Then every allocation

$$\mathbf{A}^* = \left(Q_1 \frac{1}{\overline{Q}} \overline{\mathbf{y}}, \ldots, Q_{\mathrm{J}} \frac{1}{\overline{Q}} \overline{\mathbf{y}} \right) \quad \text{with } Q_{\mathrm{j}} \geq 0 \text{ for all j and } \sum_{\mathrm{j}=1}^{\mathrm{J}} Q_{\mathrm{j}} = \overline{Q}$$

constitutes an equilibrium, together with a suitable price system $\mathbf{p}^*$. If technology were convex on the firm level, this would always be the case.

<u>Case 2:</u> $\mathrm{f}^{\mathrm{N}}(\overline{\mathbf{z}}) < \mathrm{f}(\overline{\mathbf{z}}) = \overline{Q}$.

Then, obviously, $\overline{\mathbf{y}} \notin \mathrm{Y}^{\mathrm{N}}$. Individual producers are not able to use the factors efficiently in the proportion given by $\overline{\mathbf{z}}$. Because technology is non-convex on the firm level, there certainly are such factor proportions. Moreover, inequalities (2) and (3) show that for *any* factor proportion $\overline{\mathbf{z}} >> 0$, there is a level of communication costs high enough to make this the relevant case.

The equilibrium allocation is efficient. Therefore, in case 2 at least two different unit activities must be in active use. They satisfy the zero-profit condition: $\mathbf{p}^* \mathbf{y}_{\mathrm{j}}^{\mathrm{u}} = 0 \; \forall \, \mathrm{j}$. All profit-maximising unit activities lie on a hyperplane that supports Y and goes through the origin. The aggregate production vector lies in the interior of the cone spanned by the $\mathbf{y}_{\mathrm{j}}^{\mathrm{u}}$.

Let us now return to Fig. 1, which depicts the input space in equilibrium. By the equation $\mathbf{p}^* \mathbf{y}_{\mathrm{j}}^{\mathrm{u}} = 0$, the price system determines a unit cost line consisting of all inputs worth one unit of output. This line supports the input requirement set $\mathrm{V}_{\mathrm{N}}(1)$, and it contains all profit-maximising unit inputs. It also contains the "average" input of the economy, $1/\overline{Q} \cdot \overline{\mathbf{z}}$. The latter is a convex combination of the unit activities used in equilibrium:

$$\frac{1}{\overline{Q}}\cdot\overline{\mathbf{z}}=\sum_{j=1}^{J}\frac{Q_j}{\overline{Q}}\cdot\mathbf{z}_j^u \quad .$$

For the isoquant drawn with a bold line, there are two profit-maximising unit inputs, $\mathbf{z}_I$ and $\mathbf{z}_{II}$. In this case, the quantities produced by means of these activities are uniquely determined.

The dashed isoquant shows a case in which there are not just two, but three profit-maximising activities. The equilibrium quantities are not fully determined. A dualistic productive structure, however, must be regarded as the typical case: The efficient aggregate production can always be realised by using just *two* unit activities. An additional profit-maximising unit activity, however, is restricted to lie on the same supporting hyperplane as $\mathbf{y}_I^u$ and $\mathbf{y}_{II}^u$.

Under what conditions will a multiple productive structure be realised? To begin with, the condition $f^N(\overline{\mathbf{z}})<f(\overline{\mathbf{z}})$ shows that, with a given technology, the *composition of the aggregate factor endowment* is decisive. If the allocation of labour in the proportion given by aggregate endowment leads to high communication costs at the firm level, then multiple unit activities are needed, normally two. The labour force is segregated at the workshop level. Furthermore, a uniform productive structure will be inefficient for any composition of factor endowment, if communication costs are sufficiently high. Therefore, for a given composition of the labour force, it depends on the level of communication costs whether a multiple productive structure will result or not.

7. A Simple Theorem of Nonsubstitution

Now we have to work out a special feature of equilibria characterised by a multiple productive structure. *Within certain bounds, the real wages, as well as the efficient unit activities, are invariant with respect to changes in the aggregate factor endowment.* Let the composition of $\overline{\mathbf{z}}$ be such that a multiple productive structure is needed, and let $\mathbf{p}^* >> 0$ be an equilibrium price system. Then $\mathbf{p}^*\mathbf{y}_j \le 0$ for all $\mathbf{y}_j \in Y^N$. The set M collects all profit-maximising unit activities given $\mathbf{p}^*$:

$$M=\{\mathbf{y}^u \in Y^N \,|\, \mathbf{y}^u=(1,-\mathbf{z}^u), \text{with } \mathbf{p}^*\mathbf{y}^u=0\}\,.$$

By assumption, M contains at least two elements. The activities in M span a convex cone C of aggregate production vectors:

$$C=\left\{\mathbf{y}\in R^3 \,\middle|\, \mathbf{y}=\sum_{j=1}^{J} Q_j\mathbf{y}_j^u, \text{with } Q_j \ge 0, \mathbf{y}_j^u \in M\right\}\ .$$

This cone is depicted in Fig. 5. Every aggregate production vector contained in C is profit-maximising given the price system $\mathbf{p}^*$, and therefore efficient. We have:

<u>Proposition 3:</u> *Let* $\mathbf{y}$ *be any aggregate production vector in the interior of C. Then*
a) all unit activities used in equilibrium will be contained in M, and

b) every unit activity in M *may be used in equilibrium.*

Proof: a) $\mathbf{y} \in \mathrm{C}$ maximises aggregate profits for $\mathbf{p}^*$. Every unit activity $\mathbf{y}_j^u$ used to produce $\mathbf{y}$ must be profit-maximising itself, i.e. $\mathbf{y}_j^u \in \mathrm{M}$. b) The convex cone C is spanned by two[10] linearly independent unit activities in M that may be labelled $\mathbf{y}_{\mathrm{I}}^u$ und $\mathbf{y}_{\mathrm{II}}^u$. If an aggregate production vector lies in the interior of the cone, there will be scalars $Q_{\mathrm{I}}, Q_{\mathrm{II}} > 0$ with

$$\mathbf{y} = Q_{\mathrm{I}}\mathbf{y}_{\mathrm{I}}^u + Q_{\mathrm{II}}\mathbf{y}_{\mathrm{II}}^u \qquad \text{with } Q_{\mathrm{I}}, Q_{\mathrm{II}} > 0. \tag{8}$$

If M does not contain any further unit activity besides $\mathbf{y}_{\mathrm{I}}^u$ and $\mathbf{y}_{\mathrm{II}}^u$, the assertion is proved. Now let $\hat{\mathbf{y}}^u$ be some unit activity in M not identical with either $\mathbf{y}_{\mathrm{I}}^u$ or $\mathbf{y}_{\mathrm{II}}^u$. Then $\hat{\mathbf{y}}^u$ lies in the interior of C. The subsets defined below are a partition of the interior of C:

$$\mathrm{C}_{\mathrm{A}} = \{Q_{\mathrm{I}}\mathbf{y}_{\mathrm{I}}^u + \hat{Q}\hat{\mathbf{y}}^u \mid Q_{\mathrm{I}} > 0 \wedge \hat{Q} > 0\}; \qquad \mathrm{C}_{\mathrm{B}} = \{\hat{Q}\hat{\mathbf{y}}^u \mid \hat{Q} > 0\}; \qquad \mathrm{C}_{\mathrm{B}} = \{\hat{Q}\hat{\mathbf{y}}^u + Q_{\mathrm{II}}\mathbf{y}_{\mathrm{II}}^u \mid \hat{Q} > 0 \wedge Q_{\mathrm{II}} > 0\}.$$

Every $\mathbf{y}$ in the interior of C is contained in one of these subsets and can be written as a linear combination of $\hat{\mathbf{y}}^u$ and other unit activities with $\hat{Q} > 0$. Fig 6 illustrates the argument. ■

The aggregate production vectors in the interior of C are closely related: Each one is optimal for the same (normalised) price system - and only for this price system. If one thinks of the production side of the economy as a representative firm, the price system makes the firm indifferent between all the production vectors represented as points in C.

Proposition 4: Let $\mathbf{y}$ *be any aggregate production vector in the interior of* C. *The price system* $\mathbf{p}^*$ *for which* $\mathbf{y}$ *maximises aggregate profits is uniquely determined up to a scalar, and it is the same price system for all* $\mathbf{y}$.

Proof: The existence of a price system $\mathbf{p}^*$ already follows from the definitions of C and M. As in (8) above, every $\mathbf{y}$ in the interior of C can be written as a linear combination of two unit activities from M with positive coefficients. For every $\mathbf{p}^*$, therefore, both $\mathbf{y}_{\mathrm{I}}^u$ and $\mathbf{y}_{\mathrm{II}}^u$ must be profit-maximising, i.e., $\mathbf{p}^*\mathbf{y}_{\mathbf{I}}^u = 0 \;\wedge\; \mathbf{p}^*\mathbf{y}_{\mathbf{II}}^u = 0$. These are two linearly dependent equations for the prices p_{Q}^*, p_{A}^* and p_{B}^*. If we normalise by p_{Q}^*, there can be no more than one equilibrium price system. ■

The graphical representation of factor market equilibrium in Fig. 1 makes this obvious. Real wages are uniquely determined. There cannot be a second isocost line that both leads through $1/Q \cdot \bar{\mathbf{z}}$ and supports the input requirement set $\mathrm{V}(1)$. This holds true for all factor endowments in the interior of the cone spanned by $\mathbf{z}_{\mathrm{I}}$ and $\mathbf{z}_{\mathrm{II}}$.

Consider the projection of C onto the input space, C^{Z}. The set C^{Z} is also a cone; it contains all aggregate inputs that belong to the aggregate production vectors in C. If we vary the factor endowment within the bounds of C^{Z}, the production sector behaves similar to a Leontief system with substitution possibilities. The real factor prices remain constant, just as the profit-

---

[10] M contains at least two different unit activities, so $\dim(\mathrm{C}) \geq 2$. On the other hand, $\dim(\mathrm{C}) \leq 2$, because all unit activities in M are contained in a supporting hyperplane given by the equation $\mathbf{p}\mathbf{y}^u = 0$.

maximising activities. Thus we can state that Propositions 3 and 4 constitute a local Nonsubstitution Theorem[11] for the case of a multiple productive structure.

Every competitive equilibrium that uses an aggregate input within the cone C^{Z} is characterised by the same efficient unit activities and by the same real wages. The results remain valid in a model with real wage-dependent factor supply. We also can generalise to a situation with convex technologies and constant returns, in which more than two primary factors are used.[12]

The simple model of industry structure described above makes two assertions on economies characterised by ethnic dualism. First, the propensity of the workforce to segregate on the firm level is high if workers are characterised by major cultural differences. The complementarities given by comparative advantages are then countervailed by interferences due to high co-ordination costs, and a separation of the workforce into two (or more) sectors becomes efficient. The approach thus gives us an explanation and interpretation of dualistic economic structures. Second, if a multiple productive structure prevails, the *aggregate* production set is linear in the neighbourhood of the equilibrium production vector. As the aggregate factor input composition varies, this will be accommodated by sectoral output quantities, not the factor intensities on the firm level. As long as the economy stays in the dualistic regime, the wage structure will be independent of the composition of the labour force.

8. Generalisations and Outlook

We have to add some points that have not been mentioned before. Communication costs are not the only inhomogeneity costs capable of causing nonconvexities on the firm level and a dualistic industry structure. Similar to the Schelling (1978) model of residential segregation, direct externalities between workers could depress productivity or make workers demand higher wages. Furthermore, certain indivisibilites might have to be adjusted in a way that suits the productivity of either one or the other type of labour. Examples might be the company language, the form of organisation or the intensity of IT use. Rosen (1978) has put it this way:

> The often observed fact that a factory in one country is more productive than its identical twin in another country can arise because work assignments embodied in the design of capital are optimal for one labour force and not for another owing to differences in the distribution of worker skills and comparative advantage.

---

[11] See e.g. Mas-Collel et al. (1995), pp. 157-160. The Nonsubstitution Theorem was derived by Samuelson (1951), pp. 142-146, and refers to systems of nested production processes with n output goods, but only one primary input. Samuelson assumes continuous, linearly homogeneous production functions and shows that with efficient production, the activities used in production are independent of the composition of output. The same holds true for equilibrium prices. If the composition of output varies, substitution takes place solely by adapting the level on which the efficient activities are used. An important difference remains: the Nonsubstitution theorem by Samuelson makes an assertion on systems with one primary input and several output goods, whereas the Propositions 3 and 4 assume one final good and several primary inputs.

[12] Let k be the number of factors. We have to suppose $J \geq k$ as to the number of producers. After generalising proposition 1 in a straightforward fashion, all propositions hold *mutatis mutandis*. Yet, without further assumptions, a *dualistic* productive structure can no longer be regarded as typical.

Another issue is that the technological assumptions might seem overly restrictive. However, the argument is easily generalised. First, let us introduce capital in a way that leaves the basic conclusions intact. Let

$$W_{\mathrm{j}} = \mathrm{f}^{\mathrm{N}}\left(A_{\mathrm{j}}, B_{\mathrm{j}}\right)$$

be the labour input of producer j, measured in efficiency units. The function f^{N} is assumed to have the same properties as before. Together with capital K_{j}, the labour services are used for the production of the output good, as described by the following meta production function:

$$y_{\mathrm{j}} = g^{\mathrm{N}}\left(K_{\mathrm{j}}, W_{\mathrm{j}}\right) \quad .$$

The function g^{N} is assumed to be linearly homogeneous and weakly concave. If marginal returns to K_{j} and W_{j} in the meta function g^{N} are decreasing everywhere, it is easy to see that the compound production function will be linearly homogeneous with decreasing marginal returns with respect to every factor. In a factor market equilibrium with a dualistic productive structure there will be two types of producers: those that combine capital mainly with labour of type A and those that specialise in using labour of type B, both employing the two types in fixed proportions. *Relative* wages are still invariant to small changes in the aggregate labour endowment. *Absolute* wages will decrease due to an influx of migrant workers, if the economy in question is closed. In a small open economy with capital mobility, however, the price of capital is given by the world market interest rate, and an influx of migrant workers will be accompanied by a corresponding inflow of capital from abroad.

Second, the limitation to one single output can be dispensed with. If we suppose a number of different goods to be produced by a single output technology of the form $\mathrm{f}^{\mathrm{N}}(\cdot)$, the intervals C^{Z} of inefficient factor proportions will vary from one good to the other, as well as the respective rates of substitution in the regimes of multiple productive structures. In equilibrium, however, there will be but one relative wage. Any sector will be in one of three states: (a) Producers use labour of type A intensively, or (b): the sector is characterised by a multiple productive structure, or (c): it mainly uses labour of type B. Relating relative wages to the proportion $\overline{B}/\overline{A}$ of labour endowment in the economy will yield a curve that decreases monotonously, as one sector after the other switches from state (a) to state (c). But because of convexity of aggregate technology, every such switch involves an interval where the sector is situated in a dualistic regime (b). There, rates of substitution are constant and variations of aggregate labour supply are absorbed by accommodating the level of two or more different efficient activities. As long as one sector of the economy is in a dualistic regime, it acts to sterilise the wage structure from supply shocks.

All this is good news with respect to the impact of migration as a consequence of EU enlargement. There are, however, further issues involved. We have shown that the propensity of the labour force to segregate on the workshop level depends on the level of communication costs. A high level of communication costs resulting form marked cultural differences means that the two types of labour are segregated almost completely in the various sectors of the economy. This, however, is not favourable to a smooth assimilation of the migrants in their new socio-economic environment. Assimilation is a process of social learning, and work is very important for the diffusion of knowledge, ideas and values. In von Kalckreuth (1999a and b) we show that a slow rate of diffusion in a dualistic economy leads to sluggish growth and a highly unequal income distribution.[13] This means that the presence of strong cultural and economic differences in a dualistic economy may be conducive to internal structures that are the basis for their very persistence.

Literature

Altonji, Joseph G. and David Card (1991), The Effects of Immigration on the Labor Market Outcomes of Less Skilled Natives. In: John M. Abowd and Richard B. Freeman (ed.), Immigration, Trade, and the Labor Market. A National Bureau of Economic Research Project Report. Chicago, London: University of Chicago Press, pp. 201-234.

Bauer, Thomas (1998), Arbeitsmarkteffekte der Migration und Einwanderungspolitik: Eine Analyse für die Bundesrepublik Deutschland. (In German, *Labour Market Effects of Migration and Immigration Policy: An Analysis for the Federal Republic of Germany*), Heidelberg: Physica.

Bauer, Thomas and Klaus F. Zimmermann (1999), Assessment of Possible Migration Pressure and its Labour Market Impact Following EU Enlargement to Central and Eastern Europe. A Study for the UK Government Department for Education and Employment, IZA, Bonn.

Becker, Gary S. und Kevin M. Murphy (1992), The Division of Labor, Coordination Costs and Knowledge. Quarterly Journal of Economics 107, pp. 1137-1160.

Bénabou, Roland (1996a), Heterogeneity, Stratification and Growth: Macroeconomic Implication of Community Structures and School Finance. American Economic Review 86, pp. 584-609

Bénabou, Roland (1996b), Equity and Efficiency in Human Capital Investment: The Local Connection. Review of Economic Studies 63, pp. 237-264.

Boeri, Tito and Brücker, Herbert (2001), The Impact of Eastern Enlargement on Employment and Labour Markets in the EU Member States. Forthcoming.

Borjas, George J. (1990), Friends or Strangers: The Impact of Immigrants to the US Economy. New York: Basic Books.

Borjas, George J. (1994), The Economics of Immigration. Journal of Economic Literature 32, pp. 1667-1717

Card, David (1990), The Impact of the Mariel Boatlift on the Miami Labor Market. Industrial and Labor Relations Review 43, pp. 245-257.

Chiswick, Barry S. (1991), Speaking, Reading, and Earnings among Low-Skilled Immigrants. Journal of Labor Economics 9, pp. 149-170.

De New, John P. and Klaus F. Zimmermann (1994), Native Wage Impacts of Foreign Labor: A Random Effects Panel Analysis. Journal of Population Economics 7, pp. 177-192.

Debreu, Gérard (1959), Theory of Value; New York: John Wiley.

Dornbusch, Rudiger, Stanley Fischer and Paul A. Samuelson (1977), Comparative Advantage, Trade and Payments in a Ricardian Model with a Continuum of Goods. American Economic Review 67, pp. 823-839.

Durlauf, Steven (1996), A Theory of Persistent Income Inequality. Journal of Economic Growth 1, pp. 75-93

Gang, Ira N. and Francisco L. Rivera-Batiz (1994), Labor Market Effects of Immigration in the United States and Europe. Substitution vs. Complementarity. Journal of Population Economics 7, pp. 157-175.

Greenwood, Michael J. and John M. McDowell (1986), The Factor Market Consequences of U.S. Immigration. Journal of Economic Literature 24, pp. 1738-1772.

Grossman, Jean Baldwin (1982), The Substitutability of Natives and Immigrants in Production. Review of Economics and Statistics 64, pp. 596-603.

von Kalckreuth, Ulf (1999a) Wachstum und soziale Struktur in dualistischen Volkswirtschaften. (In German, *Growth and Social Structure in Dualistic Economies),* Frankfurt am Main, Berlin, Bern etc: Lang.

von Kalckreuth, Ulf (1999b), Social Learning in Dualistic Societies: Segregation, Growth and Distribution. Discussion Paper No. 571-99, Institut für Volkswirtschaftslehre und Statistik, University of Mannheim

---

13 On this, also see Bénabou (1996a and b), as well as Durlauf (1996).

Koopmans, Tjalling C. (1957), Three Essays on the State of Economic Science. New York, Toronto, London: McGraw Hill.

Kossoudij, Sherrie A. (1988), English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men. Journal of Labor Economics pp, S. 205-228.

Lazear, Edward P. (1999), Culture and Language. Journal of Political Economy 107 (Supplement), pp. S95-S126.

Mas-Collel, Andreu; Michael D. Whinston and Jerry R. Green (1995), Microeconomic Theory. New York, Oxford: Oxford University Press.

McManus, Walter S. (1985), Labor Market Costs of Language Disparity: An Interpretation of Hispanic Earnings Differences. American Economic Review 75, pp. 818-827.

McManus, Walter S. (1990), Labor Market Effects of Language Enclaves: Hispanic Men in the United States. Journal of Human Resources 25, pp. 228-252.

Pischke, Jörn Steffen and Johannes Velling (1994), Wage and Employment Effects of Immigration to Germany: An Analysis based on Local Labor Markets. CEPR Discussion Paper No. 935, London.

Rosen, Sherwin (1978), Substitution and Division of Labor. Economica 45, pp. 235-250

Samuelson, Paul A. (1951), Abstract of a Theorem Concerning Substitutability in Open Leontief Models. In: Tjalling C. Koopmans (ed.), Activity Analysis of Production and Allocation. Proceedings of a Conference. New York: Wiley; London: Chapman.

Schelling, Thomas C. (1978), Micromotives and Macrobehavior. Fels Lectures on Public Policy Analysis. New York, London: Norton.

Takayama, Akira (1985), Mathematical Economics, 2. ed.; New York: Cambridge University Press.

Velling, Johannes (1995), Immigration und Arbeitsmarkt. Eine empirische Analyse für die Bundesrepublik Deutschland. (In German, *Immigration and Labour Market. An Empirical Analysis for the Federal Republic of Germany)*, Baden-Baden: Nomos.

Winter-Ebmer, Rudolf and Josef Zweimüller (1996), Immigration and the Earnings of Young Native Workers. Oxford Economic Papers 48, pp. 473-491.

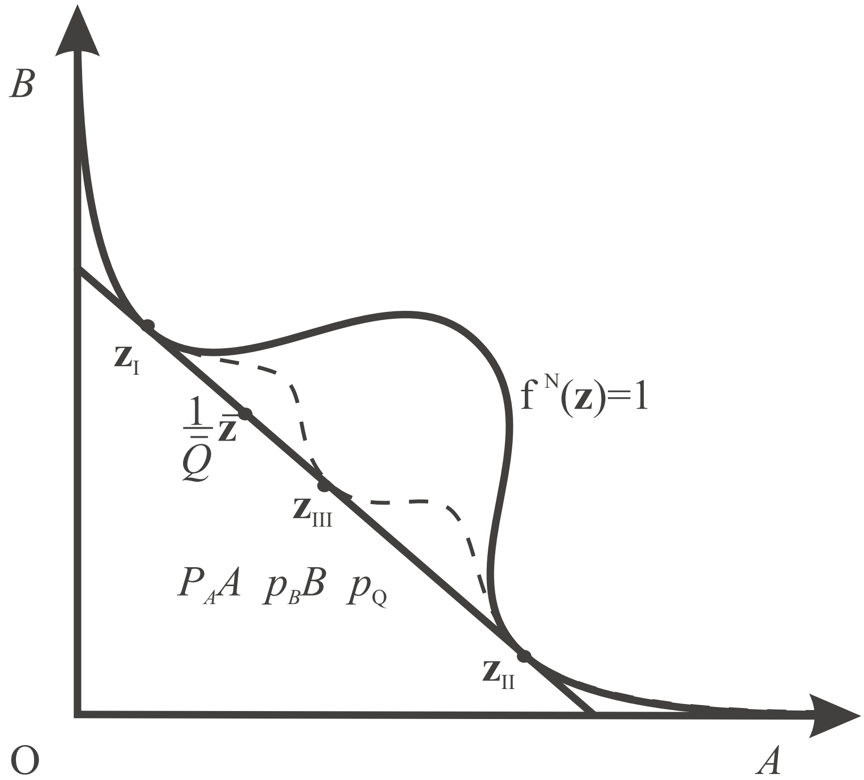

Fig. 1: Unit Isoquant and Factor Market Equilibrium with High Communication Costs

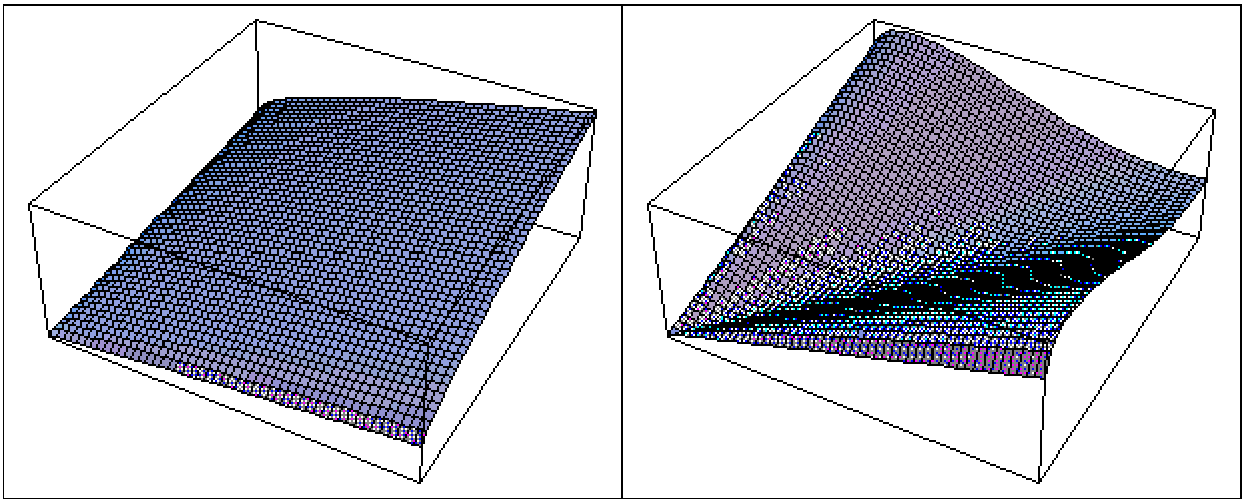

Fig 2: Gross Production Function and Net Production Function

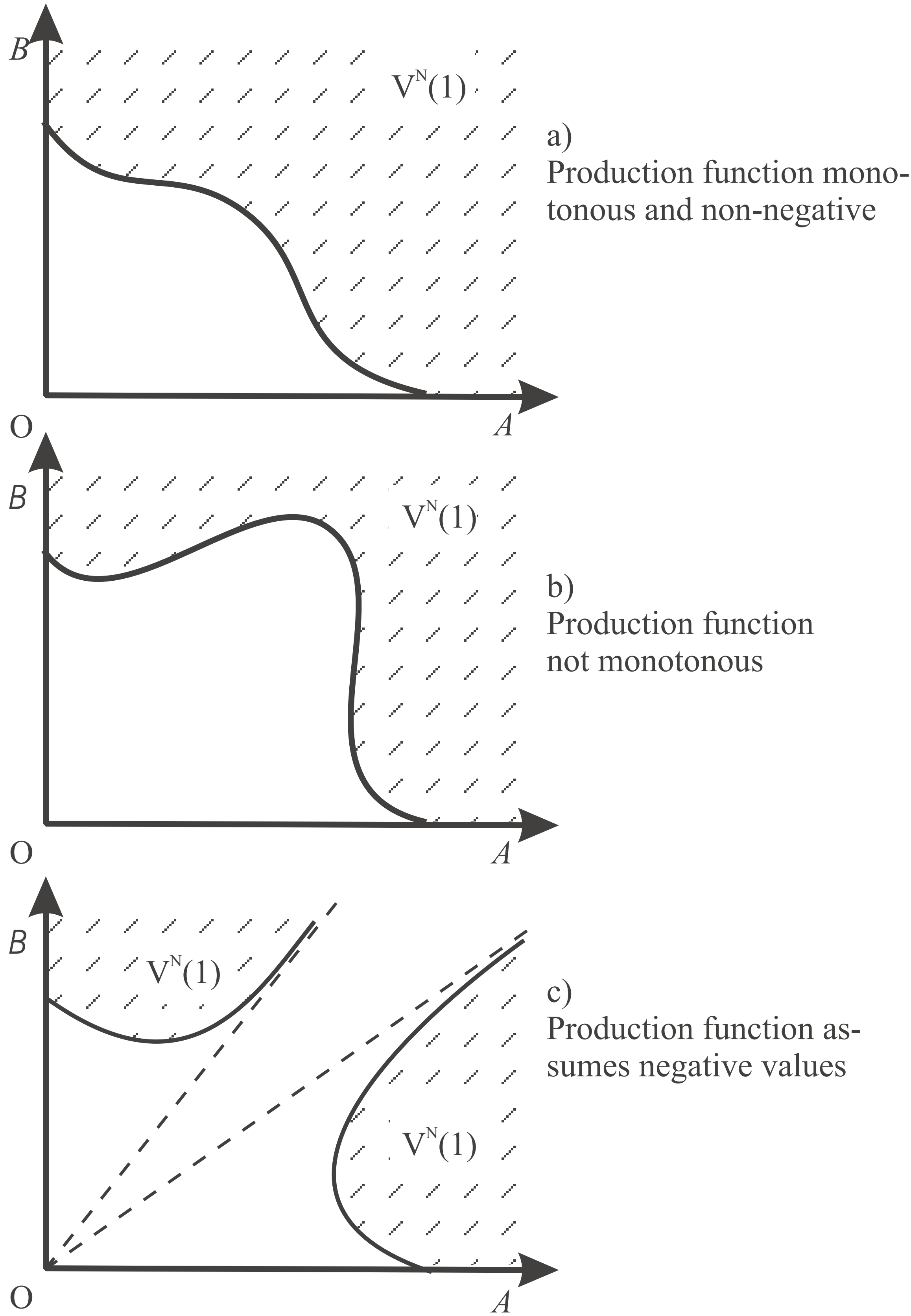

Fig. 3: Input Requirement Sets for Non-Convex Technologies

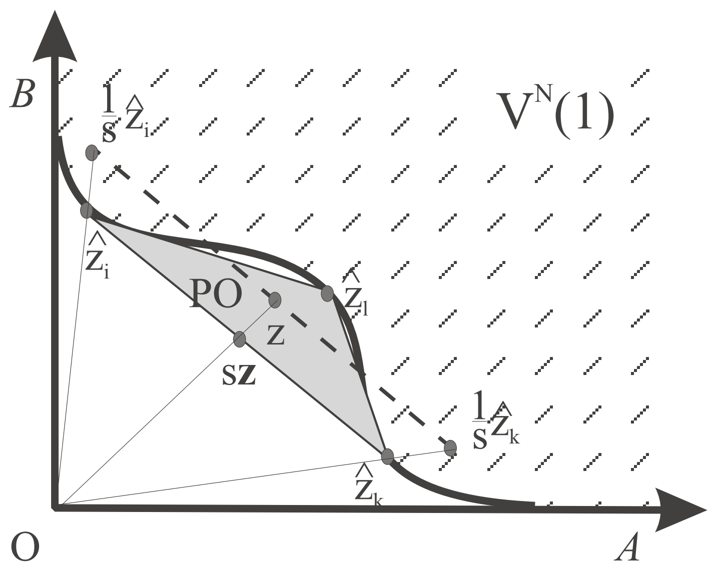

Fig. 4: Every aggregate production vector is the sum of not more than two production plans

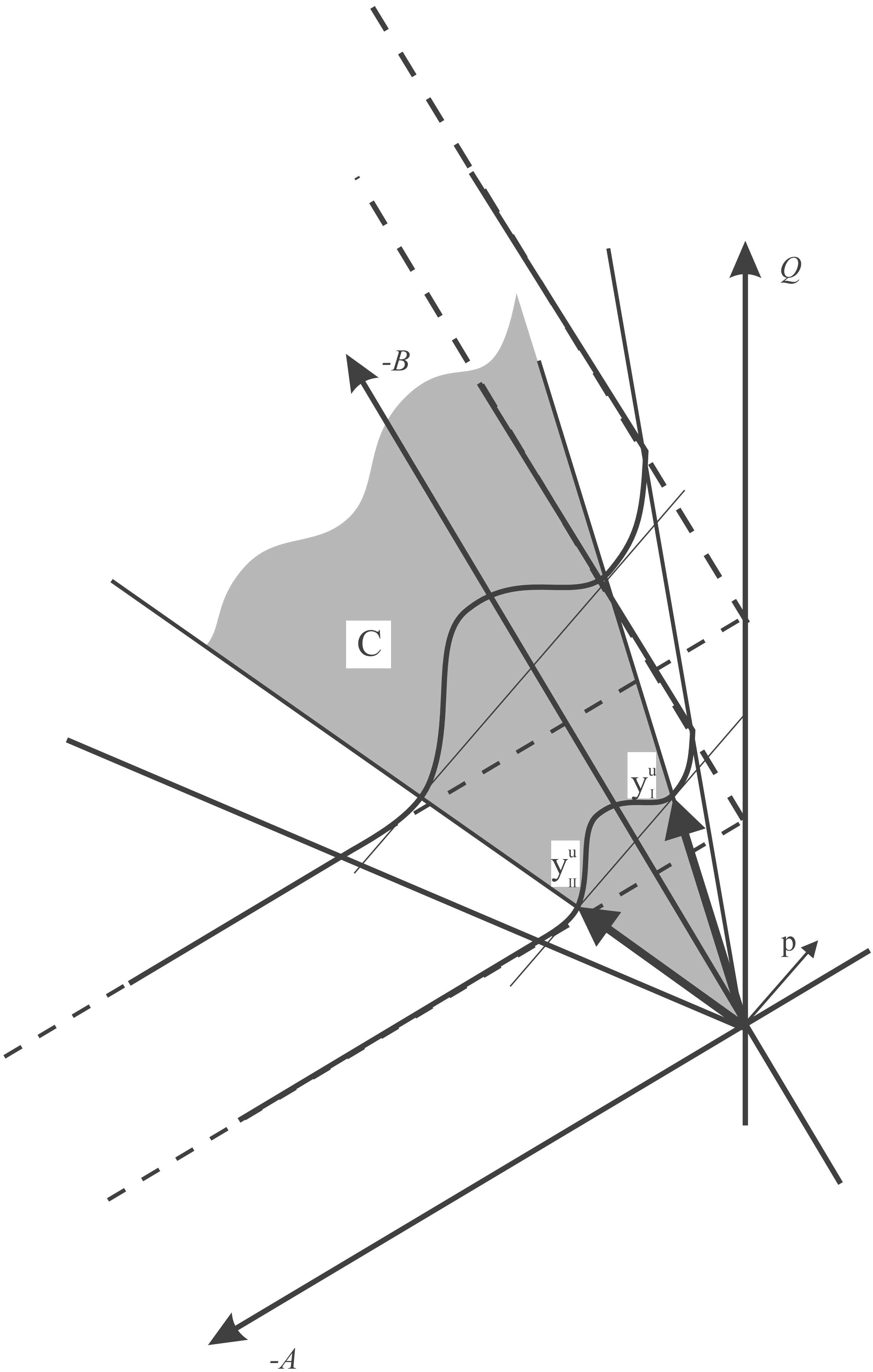

Fig. 5: Production Set and Cone of Profit-Maximising Aggregate Productions

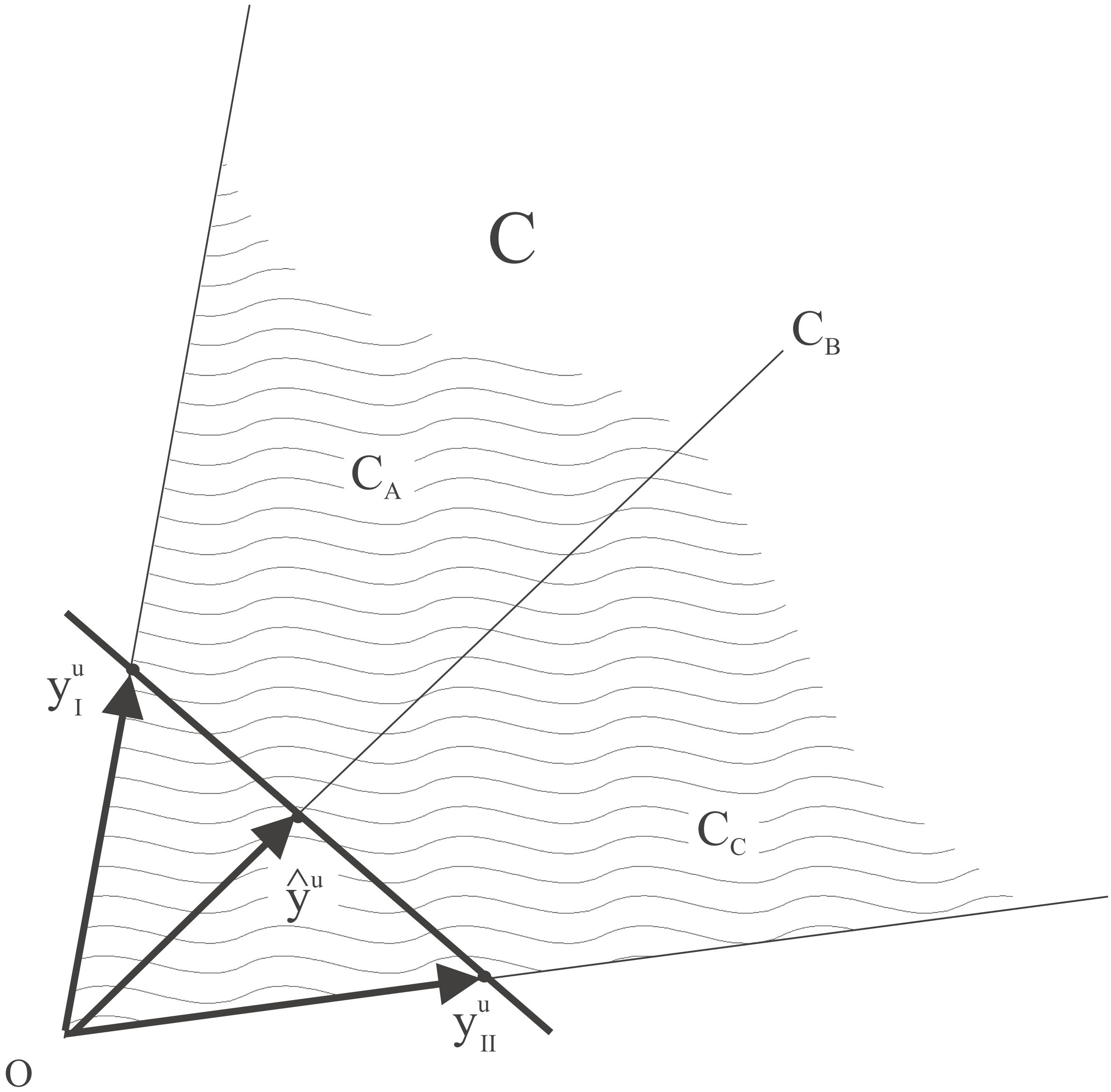

Fig. 6: Partitioning the Cone of Profit-Maximising Aggregate Production Vectors