Chapter F) Internal migration and regional disparities

F1) Introduction
The first aim of this chapter is to look more specifically at the issue of geographical labour mobility. For this task, it is extremely important to distinguish sharply between the causes for labour migration between regions, and the consequences of labour mobility for the sending and the receiving locations.

The former issue, the causes, will only briefly be discussed in section F2, since it is relatively uncontroversial. The consequences and economic effects of labour mobility on the other hand are heavily disputed in the literature. Is labour mobility an adjustment force that gradually leads to an erosion of existing spatial disparities? Or do differences even get larger and more pronounced when workers are mobile across regions?

Controversies about this issue have a long tradition in the history of economic thought. And this old discussion is also alive in this book. The theoretical models that have been discussed in the preceding chapters have drawn utterly different conclusions about the consequences of labour migration. In section F3 we illustrate the conventional result from neoclassical economics that labour mobility leads to regional factor price convergence. But it will become apparent that this result hinges on quite restrictive assumptions, and empirical evidence does not seem to support the neoclassical convergence hypothesis all too strongly.

In the following two sections we will therefore put forward two distinct arguments why labour mobility might lead to spatial divergence instead of convergence. The first reason, presented in section F4.1.), is the presence of scale and agglomeration economies. This issue has already shown up in the last chapters, but it will be restated here with a special emphasis on the effects of migration on the regional labour markets. The second reason for “non-convergence” is self-selectivity within the group of migrants. We show in section F4.2.) that there is a great deal of evidence in support of the view that skilled workers tend to be geographically more mobile than unskilled workers. This fact can be explained by straightforward economic considerations.

The consequences of selective labour migration are then derived in detail in section F5 where we explicitly distinguish between two types of workers, skilled and unskilled ones. In order to exclusively focus on the selectivity argument for “non-convergence”, we will work with a very simple product market structure with constant returns to scale in this section. We will consider both a full employment version of the model, and a version where immobile unskilled labour is subject to unemployment due to a union minimum wage. The latter model variant is interesting also in another respect, as it specifically addresses the role of national union wage setting for regional unemployment outcomes.
In section F6 we come back to the issue of regional agglomeration by constructing a model that features both localised increasing returns to scale and the selectivity of internal migrants. Within this set-up, we will likewise demonstrate the “non-convergence”-character of labour migration, and analyse the impact of undifferentiated national collective bargaining on regional unemployment rates and effective earnings. In section F7 we compare the theoretical approaches from this chapter with that of chapter E, and we assess again the empirical applicability and relevance of the models from this chapter.

F2) The causes of internal migration
The location decision of any individual is an inherently economic decision. Any individual calculates a gross utility for remaining in the region of residence, and one for moving to the best alternative region of his or her choice. A move takes place if the cost of migration does not exceed the gross gains from moving (Pissarides/McMaster, 1990). There is of course a large array of individual-specific characteristics and political impact factors that influence the mobility choice of a single person. Individual characteristics that influence the mobility behaviour are e.g. the age, family backgrounds, language capabilities, regional preferences etc.\(^1\) Migration might also be driven by jurisdictional policy differences. Certain regions might attract immigrants because of an attractive tax and welfare state regime (Razin/Sadka, 2001; Wildasin, 2000). Supposedly these political factors are less relevant in the context of internal, i.e. intra-national, than in the context of international migration.

But we abstract from these influence factors and focus on two particular important forces that seem most relevant for explaining regional migration patterns. Economic considerations suggest that regions should all in all loose working population the lower is the regional wage level relative to other locations, and the higher is the relative regional unemployment rate.\(^2\) In all models in the preceding chapters, interregional migration was always driven by regional wage and unemployment disparities.

A look back at figure A8 casually supports this theoretical construction also from an empirical point of view: European NUTS2-regions that were subject to population decline between 1993 and 1998 shared on average some common characteristics. Among these were an above-average unemployment rate, and a below-average per capita income (which should be highly correlated with wages). Further pieces of evidence can be obtained from section A5, where we have presented

\(^1\) In a later section of this chapter we will take into account the individual skill level as an important impact factor for the personal mobility decision. For an older, but still very instructive survey about the micro-motives for migration, see Greenwood (1975).

\(^2\) One might also label a high regional unemployment rate a “push factor” that spurs emigration, and high regional wages a “pull factor” that attracts immigrants from other areas.
stylised facts on regional disparities within West Germany. It was shown that the southern part of West Germany reveals higher effective earnings and lower unemployment rates than the northern part. Internal migration flows consequently went from the north to the south.

The plausible hypotheses that internal migration is driven by regional wage and unemployment differences has been verified by numerous econometric studies. Pissarides/McMaster (1990) e.g. estimate a net internal migration equation for nine regions in the UK for the time period from 1961-1982. They find that regional unemployment rates and regional wages are significant explanatory variables (with the expected sign) for regional net migration rates. The result is verified for the case of West Germany by Suedekum (2003), who estimates a similar net migration equation for ten west Bundesländer (without Berlin) covering the time period from 1988-1999. Consistent empirical findings for the case of West Germany come e.g. from Alecke/Untied (2000), Buettner (1999) or Decressin (1994).

Hence, there seems to be considerable consistency in the empirical literature about the causes of internal migration. But does the well established empirical fact that workers move out of regions with high unemployment rates and low wages imply anything about the consequences of migration?

**F3) The consequences of internal migration: The neoclassical view**

Some authors, e.g. Pissarides/McMaster (1990) or Buettner (1999) claim that the empirically observed direction of internal migration flows indeed implies the existence of a long-run equilibrium where regional wage and unemployment differences have vanished, or persist only in form of compensating differentials à la Harris/Todaro (1970) and Hall (1972). It is admitted that migration might be a quite weak equilibrating force. Hence, a very long time period might be needed in order for regional disparities to disappear. But eventually, regional convergence will occur as the result of emigration out of the areas with poor economic conditions. These inferences about the consequences of migration are straightforward derivations from the neoclassical theory of migration. This can easily be demonstrated by a simple theoretical framework that we will introduce in this section.

Suppose there is a static economy consisting only of two regions \( r=\{1,2\} \). Production in either region is described by a Cobb-Douglas production function, where output \( Y_r \) is produced by combining the input factors physical capital \( K_r \) and (homogenous) labour \( L_r \). The parameter \( 0 < \nu < 1 \) is the usual capital share. The capital stock is assumed to be fixed for either region. There is furthermore a term \( A_r \) capturing the total factor productivity of region \( r \).

\[
Y_r = A_r (K_r)^\nu (L_r)^{1-\nu}
\]  
(F.1)
Product and factor markets work perfectly, i.e. we consider at first instance a full-employment model. The good $Y_r$ is freely tradable, and the price $p^Y$ is normalized to unity for both regions. We know that under these conditions the equilibrium wage in region $r$, $w_r^*$, is given by

$$w_r^* = (1 - \nu) A_r \left( \frac{K_r}{L_r} \right)$$

The equilibrium wage rate in region $r$ depends on the capital/labour-ration in the respective location. Ceteris paribus, it decreases when more labour enters the market. Now consider the following thought experiment: suppose that the labour force and the total factor productivity in region 1 and 2 are initially identical ($L_1 = L_2$, $A_1 = A_2$), but the fixed capital stock of region 1 is larger than in region 2 ($K_1 > K_2$). From (F.2) it immediately follows that the equilibrium wage of region 1 is above the level in region 2, because labour is the relatively scarcer factor. With mobile labour, this will cause migration from region 2 to region 1 and put equilibrium wages in $r=1$ under strain, as employment increases and the marginal product of labour declines. The migration process will continue up to the point of wage equalization, which in our example is equivalent with the point where the capital/labour-ratios are equalized across regions. Migration thus leads to wage convergence, since it changes the relative factor intensity in either region. This logic can be illustrated in figure F1, which shows the labour markets of both regions jointly in one diagram.

**Figure F1: Migration in a neoclassical two-region model**
In the initial situation, labour supply (which we assume to be perfectly inelastic) is identical in both regions. But due to the larger fixed capital stock, labour demand is higher for any given wage rate in region 1. The equilibrium wage \( w_1 \) is therefore larger than \( w_2 \). The subsequent labour migration shifts the labour supply schedule of the receiving region 1 to the right, labour supply in the sending region 2 is shifted to the left. The labour demand schedules are not affected by the mobility of input factors. Final equilibrium is obtained when there is interregional wage equalization at \( w^* \).

Introducing unemployment complicates the analysis to some extend, also depending on what is modelled as the source of unemployment. But intuitively, migration will also lead to convergence of regional unemployment rates if neoclassical technological conditions prevail. This point can be made also in a very simple way. Suppose that the initial low wage region 2 is subject to unemployment, whereas there is full flexibility and thus full employment in region 1. If the unemployed of region 2 can freely emigrate, the unemployment rate of region 2 will decrease and converge to the level of zero that prevails in region 1. Similarly, the equilibrium wage in region 1 decreases and converges to the level of region 2, because additional labour is added to production, which causes diminishing returns.

All in all, one can see even from this rudimentary static model what is essentially behind the hypothesis of factor price convergence: it is the prevalence of neoclassical assumptions on the production function, i.e. constant returns to scale and diminishing marginal returns in case of partial factor variations. This leads to the well known proposition that factor prices are determined by factor proportions alone, as the relative scarcity implies the marginal factor productivity that in perfectly competitive factor markets is equal to the real factor price. Of course all factors in the neoclassical production function are assumed to be homogenous, which specifically means that all workers in either region are perfectly substitutable input factors in the production process (F.1).

**F4) Internal migration and regional divergence: alternative views**

But does the empirically observed direction of migration flows from section F2 necessarily imply the long-run equalization of regional disparities like the small neoclassical model from section F3 is telling? In this section we will propose two distinct arguments why this might not be the case.

But before doing so, it is worth noting that there is considerable empirical dissatisfaction with the neoclassical implication that labour mobility will lead to regional convergence. The available empirical evidence on this matter is mixed at best (Walz, 2001). It is highly questionable from an empirical point of view if migration actually spurs \( \beta \)-convergence in the context of international growth regressions (Braun, 1993). The persistence of regional disparities in the EU (section A2-A4), or in West Germany (section A5) alone suggest that migration seems to be a quite poor adjustment force. A very pronounced conclusion can be found in Fass-
mann/Mausburger (1997:190), who point out that "The basic question of regional economics, whether migration contributes to the adjustment of regional [...] disparities, has to be answered with 'no' from a short- and a medium-run perspective".

The purpose of this section is thus to provide theoretical rationale for the possible "non-convergence" character of internal labour migration. We want to point out how sensitive the neoclassical convergence hypothesis is with respect to the underlying technological assumptions. Even slight departures from the straight neoclassical world will lead to fundamentally different conclusions about the spatial consequences of internal labour mobility. The first argument for "non-convergence", presented in section F4.1), concerns the presence of increasing returns to scale in production that prevent the marginal productivity of labour from falling

F4.1) Increasing returns to scale

Can the real world actually be described reasonably well by a neoclassical production function with its restrictive underlying properties? This issue is ultimately a matter of empirical research. But, as pointed out in chapter D, at least on a regional level one can have serious doubts whether a production function with a scale elasticity of one is an appropriate approximation of reality. In this section we show what happens to the impact of labour migration if production on a regional level does not exhibit constant returns to scale, but rather is subject to scale economies.

In the neoclassical model in section F3, the equilibrium factor prices in either region were determined by relative factor intensities alone. During the process of migration only the labour supply functions shifted in both regions, whereas the labour demand schedules remained constant. But labour migration might also have labour demand effects, which can moderate or even offset the supply effects. Recall in this respect the evidence that was presented in section A4.1). There it was shown that regions with declining unemployment rates in the EU faced immigration of workers (i.e. an increase in labour supply). But at the same time, these regions experienced an even stronger increase in employment growth (i.e. in labour demand). It might be, that this increase in employment growth is causally due to the increase in labour supply.

Effects on labour demand can result from various channels. Let us first consider the case where labour demand is affected because of a technological externality. Suppose the production function (F.1) is rewritten in the following way

---

3 For a brief assessment whether neoclassical production functions are an appropriate approximation of reality see Hamermesh (1993).
with $\partial A_r/\partial L_r > 0$. The term that captures the total factor productivity in region $r$ is no longer given exogenously, but now depends endogenously on the size of the regional labour force $L_r$. An intuitive reason why a larger labour force might lead to a higher total factor productivity is e.g. the presence of intra-regional knowledge spillovers. The concentration of more workers in one region leads to a more vital exchange of work experience etc.. Other plausible stories have been presented also sections D2 and D3b), where we have pointed out that this modelling strategy was very popular in the first wave of the endogenous growth literature. One major technical advantage of a formulation like (F.3) is that the parametric introduction of the aggregate externality is consistent with perfect competition. Each single firm maximizes profits while taking the aggregate total factor productivity as given. The equilibrium wage in either region is now given by

$$w^*_r = (1-\nu)A(L_r) \left( \frac{K_r}{L_r} \right)^\nu + \frac{\partial A_r}{\partial L_r} \left[ K_r^\nu L_r^{1-\nu} \right]$$

(F.4)

Compared to (F.2), an additional term enters the expression of the equilibrium wage. This term captures the aggregate externality and is increasing in the labour force size $L_r$. Equation (F.4) implies that during the process of internal migration not only labour supply, but also labour demand curves are affected. This is illustrated in figure F2. Additional to the labour supply shifts, there are now also secondary shifts in the labour demand schedules of both regions.

As migration proceeds from region 2 to region 1, the labour demand schedule of region 1 is shifted to the right due to the external scale economies, the opposite happens in region 2. Depending on the strength of the externality, i.e. on the exact shape of the function $A(L_r)$, the migration of labour might not lead to a closing interregional wage differential, but rather to a widening. Labour mobility in this example thus leads to further regional divergence rather than to regional convergence.4

An argument in this spirit has e.g. been made by Burda/Wyplosz (1992). They also model a two-region economy, where regionally mobile workers embody human capital that enters the production function externally in a way that is loosely

4 If nothing else is added, and labour is perfectly mobile, all workers will move out of region 2 and concentrate in region 1, in order to fully exploit the external economies of scale. Note also that if the externality is not too strong, the shift in labour supply might be stronger than the shift in labour demand. In this case, the regional equilibrium wages would convergence as internal migration proceeds, but the convergence speed is reduced through the aggregate externality.
comparable to (F.3). They also conclude that labour migration might not be an equilibrating force, but might rather lead to a “Mezzogiorno” syndrome for the sending region.

**Figure F2: Migration in a two-region model with an aggregate externality**

Labour migration of course also leads to regional divergence if the scale economies do not enter parametrically through a technological externality, but if the scale economies are internal and transmitted through a pecuniary externality. The difference between the two concepts has been explained in depth in section D2. A good example is our approach from section E3. We have derived in great length that labour migration in our approach does not lead to regional convergence, but that regional disparities get sharper if workers move out of the disadvantaged areas. In the light of the discussion from this section, we can illustrate the divergence result from section E3 also in term of labour supply and labour demand. Consider the product market equilibrium condition for the regions r and s, equation (E.19). Let us abstract for the moment from unemployment and consider the easier full employment case. Manipulating the simplified version of expression (E.19) yields the following upward-sloping labour demand curve for region r

$$L = \left( \frac{\ell}{T} \right) (w)^{\frac{\theta}{1-\theta}}.$$  

(F.5)
An analogous expression applies to region s. The single variables are defined as in chapter E. The intuition for the positive slope of the labour demand curve (F.5) has been extensively discussed in the last chapter. In short, the underlying technology is subject to an internal scale effect. The expansion of the labour force leads to decreasing production costs, and via a zero profit condition to increasing equilibrium wages.

With an upward-sloping labour demand curve, immigration to area r will not lead to falling equilibrium wages. It will rather lead to increasing equilibrium wages, because of a better exploitation of the scale economies. An upward-sloping labour demand curve is an essential feature of NEG- and NTT-models. It implies that labour migration perpetuates rather than cures regional wage disparities. From our approach in section E3 it can be concluded that the non-convergence character of labour mobility also applies to regional unemployment rates.

All in all, the message of this section is that the neoclassical hypothesis, according to which labour migration leads to regional convergence of factor prices (and unemployment rates), does not hold when the technology exhibits sufficiently strong increasing returns to scale. These can be externally or internally transmitted. Of course people migrate from low-wage to high-wage areas. But, as shown in this section, this does not imply that the interregional wage differences must vanish.

F4.2.) Selective labour migration

We now want to turn to a different argument why migration might lead to regional divergence: the issue of self-selectivity of migrants. So far we have assumed that all labour is homogenous. This means, that all workers are assumed to be identical, with no differences in skill, education etc. This of course bares very little real world relevance. It has therefore become common in economic theory to at least distinguish between two types of labour, skilled and unskilled, where members of the former group are e.g. thought of as having completed upper secondary or tertiary education. Increasingly, individual skill is even modelled as a continuous function. This modelling strategy will be adopted in the model of section F6. In the model of section F5 we will at first only distinguish between “skilled” and “unskilled” labour.

a) Evidence on selective migration

Distinguishing between different types of labour is not only important for its own sake. Additionally, there are also drawbacks between the skill level and the propensity of any single individual to be geographically mobile. In the economics literature, it has been acknowledged since decades (and is repeated over and over) that young workers tend to be more mobile than old workers, and that skilled la-
bour is more mobile across regions than unskilled labour. If there is internal migration within an economy, the migrants on average tend to be educated and young individuals.

It is quite obvious that the emigration of such workers from any given jurisdiction does not relieve potential local problems by relaxing competition on the labour supply side. Almost surely there are secondary, (labour) demand side oriented effects that are more pervasive. Otherwise it would be very difficult to understand e.g. the continuous moaning about the large waves of emigration from East Germany (e.g. Wenz, 2002). From a straightforward neoclassical point of view, policymakers who are interested in a convergence process of East German wages and incomes to western levels should be happy about every emigrant, because the marginal productivity and thus the factor price of labour should increase for the remaining Easterners. Yet, arguments in this spirit are hardly ever heard in the current discussion. One major reason seems to be that emigrants from East Germany in particular tend to be young and well educated individuals.

Hunt (2000) has undertaken a large scale micro-econometric study about migration from East to West Germany. She finds strong evidence for selective migration. She concludes: "Emigrants are much younger than stayers, and conditional on age are more skilled, [...] This youth and brain-drain suggests that emigration from the east could be a legitimate concern for policy-makers anxious about the economic viability of the Eastern region." (p. 28). But the skill bias is by no means restricted to the case of East Germany. Mauro/Spilimbergo (1998) present a comparable empirical results for Spain. They analyse regional mobility across 50 Spanish provinces over the time period from 1964-1992. They distinguish between 5 skill groups (from illiterate to college-educated), and find strong evidence for differences in the adjustment behaviour of skill groups to sudden labour market shocks: "The high-skilled are found to migrate very promptly in response to a decline in local labour demand, whereas the low-skilled drop out of the labour force or stay unemployed for a long time." This finding is consistent with the view that an individual’s willingness to be geographically mobile is a strictly increasing function of his of her skill level. Complementary evidence for Spain comes from Antolin/Bover (1997). Evidence for the case of Italy is provided by Gianetti (2001). She looks at migration patterns across Italian regions for the time period from 1980-1992, and similarly concludes that there are "huge differences in the mobility rate of educated workers and less educated ones" (p.24).

b) Reasons for selectivity

The skill bias in the group of internal migrants can be viewed at as a well established empirical fact. In this sub-section, we will present some theoretical reasons

---

why skilled workers tend to be more mobile than unskilled workers. The intuition why younger people have a higher propensity to migrate is quite straightforward: they have a longer income stream of discounted earnings to pay-off the mobility costs.

General theoretical rationale for self-selection of migrants with respect to the skill level has been developed by Borjas (1987) and Borjas et.al. (1992). Within the framework of the Roy-Model, this literature shows that the relative income inequality in source and destination area is an important determinant for the composition of migrants from the source region. Skilled workers seek to migrate to regions with higher income inequality, whereas low skilled workers prefer to migrate to low-inequality jurisdictions.

Theoretical explanations that explicitly address the higher mobility rates of high-skilled workers usually rely on the fact that mobility gains are larger for skilled than for unskilled workers, whereas mobility costs are more or less constant across skill groups. The empirical result of Mauro/Spilimbergo (1998) indicates that skilled and unskilled workers react very differently to sudden and unexpected unemployment spells. The skilled people move to other areas, whereas the unskilled often drop out of the labour force or remain unemployed. A plausible reason is that the value of employment is relatively higher than the value of unemployment for skilled workers than for unskilled labour. The opportunity costs of being unemployed increase with the individual skill level. High-skilled workers have thus a higher incentive to change regions and take up jobs in other areas. The incentives for geographical mobility are relatively lower for unskilled workers, since the material gains from employment are lower and might not even compensate the mobility costs.

This argument has been extended and applied to a context with regional agglomeration. Moeller/Haas (2003), Moeller (2002) and Gianetti (2001) have shown that there is a skill bonus in the agglomeration wage premium. In other words, the earnings gap for observationally equivalent individuals between core regions and rural, peripheral regions is larger for skilled than for unskilled workers. Let mobility costs again be approximately skill invariant. High-skilled workers will then reveal a higher spatial mobility, since they have a higher incentive to concentrate in the industrial core regions in order to realize the agglomeration wage premium.

For most continental European countries, with their heavily regulated and unionised labour markets, also institutional factors are relevant for explaining low and selective mobility rates (Bertola, 2000). Take Germany as a typical example. As argued above in chapter C, there is virtually nil regional differentiation in contracted wages. Collective bargaining mostly applies to workers with low and medium formal skill level. Therefore, these skill groups have little gains from geographical mobility, since spatial wage differences are artificially compressed. Regional effective earnings differentials might not be substantial enough to compensate for mobility costs. Once unemployed, workers in Germany enjoy relatively...
generous benefits. This also will have negative drawbacks on the individual willingness to be geographically mobile, since the pecuniary difference between employment and unemployment is institutionally compressed. This argument applies to a lesser extend to skilled workers. Not surprisingly, unemployment is mainly centred around people from the low and medium skill categories, since the rigid wage setting schemes and the generous welfare state arrangements not only rule out labour market clearing, but also distort geographical mobility.

c) The spatial effects of selective labour migration

What happens to the convergence hypothesis from above if there is selectivity among the group of internal migrants?

We will address this issue in depth in the following sections, where we derive illustrative theoretical models. But in this section, we want to point briefly to some parts of the older literature that is also concerned with the economic effects of selective migration. It is often seen at the root of a cumulative causation logic where migration amounts to a “vicious cycle”. This idea actually dates back to old theorists like Myrdal (1957), Hirshman (1958) or Kaldor (1970). Myrdal (1957:27) writes: “The localities and regions where economic activity is expanding will attract net immigration from other parts of the country. As migration is always selective, at least with respect to the migrant’s age, this movement by itself tends to favor the rapidly growing communities and disfavor the others.” Economic geographers are inclined by this viewpoint until today. Fassmann/Mausburger (1997: 190) e.g. point out that: “Internal migration leads to the social erosion in the regions of origin and not to the automatic adjustment of region’s endowment with factors of production. The origin areas lose human capital to the destination areas. This very uneven process of internal migration in terms of qualification between centre and periphery reinforces the regional economic disparities”.

However, this reasoning, which is fundamentally different from the simple neoclassical model presented in section F3, still has a hard time entering the spheres of mainstream economics. But also neoclassical models increasingly began to acknowledge the importance of heterogeneous labour. This mostly applies to the neoclassical growth theories. In the seminal growth model of Solow (1956), labour is still modelled as a homogenous input factor, and growth of per capita income is seen as an accumulation process of physical capital. In the Solow-model, an increase in the growth rate of the native labour force and immigration of workers from other countries is analytically equivalent. It has negative effects on the steady-state level of income per capita and increases the speed of convergence to the long-run equilibrium. Since modelling labour as a homogenous input factor

---

6 For an extensive discussion of this well known standard model see e.g. Barro/Sala-i-Martin (1999).
was increasingly seen as too simplistic, the Solow-model was extended such that it explicitly accounted for the role of human capital as an important source of growth (e.g. Mankiw/Romer/Weill, 1992). In a model with human capital, there is a fundamental difference between immigration and “normal” labour force growth: immigrants can already carry human capital, whereas young natives still need to accumulate skills. This import of human capital opposes the “Solowian” effects of immigration. The more human capital is brought along by the immigrants, the less adverse is the effect on the steady state income in the destination area. If the immigrants’ level of human capital is large enough, the impact can even become positive and the convergence speed is reduced (see also Walz, 2001; Barro/Sala-i-Martin 1999:285ff.). Hence, this short appraisal of growth theory shows that neoclassical models also draw more differentiated conclusions when a distinction is made between different types of labour.

F5) Selective migration in a two-region model

But let us now come back to the usual static two-region framework from this chapter. In this section we want to illustrate – under the use of a theoretical model – that the convergence character of labour mobility does not hold if only high skilled labour is moving across regions. In order to focus exclusively on the selectivity argument, we abstract for the moment from increasing returns or agglomeration forces as a source for regional divergence. Regional agglomeration, which after all is one of the main themes of this book, will re-enter the theoretical analysis in the next section, where we present a unified framework that jointly analyses agglomeration economies, selective migration and national union wage setting.

In this section, we rather consider an economic area consisting of two regions, both subject to a Cobb-Douglas technology. The only essential difference compared to section F3 is that we now consider two types of labour, skilled and unskilled workers. We first look at a model with full employment. Afterwards, we introduce unskilled unemployment that results because of regionally undifferentiated union wage setting.

F5.1.) The full employment case

Consider a nation consisting of two regions \( r = 1, 2 \) and two factors of production: Skilled labour \( H_r \) and unskilled labour \( L_r \) are both supplied perfectly inelastic. Each region again produces some perfectly tradable good \( Y_r \), for which the price

---

7 A summary about the major developments in growth theory, and a discussion for the reasons to depart from the Solow-model is provided by Romer (1994).
is normalized to unity. For simplicity we abstract from physical capital.\(^8\) The production function of region \(r = \{1,2\}\) now reads as

\[
Y_r = A_r (H_r)^\nu (L_r)^{1-\nu}.
\]

Unskilled labour \(L_r\) is regionally immobile and is distributed equally across regions. Unskilled labour supply in either region is given by \(L_1 = L_2 = L\). Skilled labour on the other hand is perfectly mobile, and the total number of skilled workers is equal to \(H_1 + H_2 = H\). In the first variant of the model we assume that all labour markets work perfectly. Both types of workers are fully employed and paid according to their marginal products. The regional equilibrium wages for skilled and unskilled labour, \(w_r^H\) and \(w_r^L\), can be written as

\[
w_r^H = \nu A_r (H_r)^{\nu-1} (L_r)^{1-\nu}
\]

\[
w_r^L = (1-\nu) A_r (H_r)^{\nu} (L_r)^{-\nu}
\]

The factor prices are fully determined by relative factor proportions and the exogenous parameters \(\nu\) and \(A_r\). Suppose that both regions initially have the same TFP-level \((A_1 = A_2)\), and human capital is distributed equally across the two regions \((H_1 = H_2)\). From (F.7) and (F.8) it follows that there is full equalization of both factor prices across the two regions. There is no incentive for high skilled workers to migrate.

Now suppose that the economy is hit by a permanent exogenous shock that only has negative impacts on the TFP-level in region 2, i.e. \(A_1 > A_2\). As an immediate effect of this asymmetric shock, factor prices for both the skilled and the unskilled decrease in region 2, since they are both directly affected by TFP. As a result, the mobile skilled labour will flow out of region 2 and towards region 1. This has drawbacks on unskilled wages in both regions: Additional to the negative effect from the TFP-shock, the wage \(w_2^L\) decreases further as skilled labour moves out of the region. Vice versa, the wage for unskilled labour in region 1 increases as skilled workers immigrate. The reason is that unskilled labour becomes relatively more abundant in region 2, and relatively more scarce in region 1. One can also think about this impact on \(w_r^L\) in the following way: the Cobb-Douglas function, albeit being a neoclassical production function with substitutability of input factors, also features a complementary relation of input factors. An increase in the

\(^8\) As long as the capital stock is assumed to be fixed, it only enters the production function as an exogenously given number anyway.
availability of one factor (skilled labour) also shifts up the productivity of every other factor.\(^9\) Hence, an increase of skilled labour in region 1 produces a positive pecuniary externality for the local unskilled labour force. The opposite happens in region 2, where the number of skilled workers decreases.

For given values of \(v\), \(A_1\) and \(A_2\), the regional difference in unskilled wages is thus entirely determined by the spatial distribution of skilled labour. This spatial distribution of \(H_r\) can be derived explicitly. After the exogenous TFP-shock has occurred, high skilled workers who live in region 2 face the following trade-off: due to the higher TFP-level, it is attractive to migrate to region 1. But on the other hand, a concentration of human capital in region 1 will induce diminishing marginal productivity, which puts \(w^H_1\) under strain.\(^10\) In equilibrium, these two forces just compensate for each other. Perfect mobility of skilled labour ensures that wages must be equal in both regions.

\[
v A_1 (H_1)^{v-1} (\bar{L})^{1-v} = v A_2 (\bar{H} - H_1)^{v-1} (\bar{L})^{1-v} \quad \text{(F.9)}
\]

Rearranging terms yields the equilibrium fraction of high skilled labour that concentrates in region 1 for any given magnitude of the interregional TFP-difference.

\[
\frac{H_1}{\bar{H}} = \frac{1}{1 + (A_1/A_2)^{1/(v-1)}} < 1 \quad \text{(F.10)}
\]

The larger is the technological shock, the stronger is the concentration of skilled labour is region 1. The stronger is the concentration of human capital, the higher is the regional wage disparity for the low skilled workers. The relative low skilled wage of region 1 is given by

\[
\frac{w^L_1}{w^L_2} = \frac{A_1}{A_2} \left( \frac{H_1}{H_2} \right)^v \quad \text{(F.11)}
\]

which in equilibrium is equal to \(w^L_1/w^L_2 = (A_1/A_2)^{1-v} > 1\). Equation (F.11) shows that the equilibrium regional disparity of low skilled wages is compounded of a

---

\(^9\) More technically, the cross-partial derivative \(\partial^2 Y_r/\partial L_r \partial H_r\) is positive.

\(^10\) Note that this would not be the case in a model with increasing returns to scale, where concentration increases marginal productivity and equilibrium wages. But under neoclassical conditions, such agglomeration or pooling advantages do not exist even for high skilled labour.
direct effect from the asymmetric technological shock, and a secondary effect that stems from the relocation of high skilled workers. If all labour were immobile, the low skilled wage disparity would only be given by the technological component, i.e. \( \frac{w^1_t}{w^2_t} = \left( \frac{A_1}{A_2} \right) \). The mobility of the high skilled workforce thus leads to a stronger regional disparity of low skilled wages. 

What does the simple model from this section tell about the spatial effects of internal migration? Above all it implies that it is much too simple to claim that labour mobility leads to factor price convergence, even with free trade of goods. We have shown (within a straightforward neoclassical model framework) that regional mobility of the mobile parts of the workforce, the high-skilled, leads to a regional divergence of earnings for the immobile parts, the low-skilled. The convergence hypothesis that we have discussed in section F3 is only valid for the mobile fraction of the labour force. At least in Europe, however, large parts of the labour force (the unskilled) can indeed be seen as immobile factors. For them, more migration of the high skilled from the blurring to the blooming regions does not imply factor price convergence. Those who are tied to the disadvantaged locations instead loose out from the emigration of mobile factors.

**F5.2.) Union wage setting and unskilled unemployment**

We now want to extend the model and introduce unemployment. But we pursue a different path than in chapter C by focussing now on the role of national union wage setting for regional unemployment outcomes. The merits of this alternative theoretical approach are discussed in greater detail in section F7.

Suppose that the economy under consideration is subject to high union power, and assume that there is no regional dispersion in contracted wages. We have argued in section C3.1.) that this reflects fairly well an essential institutional feature of the German labour market that is common also in other European countries. Introducing minimum-wage unemployment does not change the basic implication from the last section that mobility of skilled workers causes a wage divergence for the immobile unskilled workers. But furthermore, with our extended model, we can show that selective migration also leads to regional unemployment divergence.

The production function for either region is still of a Cobb-Douglas-type. The only change of the function (F.12) compared to (F.6) is that the employment of unskilled workers in region \( r \) \( (L_r) \) no longer needs to be identical with labour supply.

\[
Y_r = A_r (H_r)^\nu (L_r)^{1-\nu}. \tag{F.12}
\]

This is so because of a contracted union minimum wage, which equally applies to both regions \( (w^1_t = w^2_t = \bar{w}) \). If the union wage is not binding, unskilled employ-
ment is simply equal to labour supply $\bar{L}$. Otherwise, if it is set on a non-
competitive level, employment is determined by labour demand $L^d_r$.

$$L_r = L^d_r = \left(\frac{(1-\nu)A_r}{\bar{w}}\right)^{\frac{1}{\nu}} H_r, \quad (F.13)$$

The mobile high skilled workers on the other hand are still paid competitively
according to

$$w^H_r = \nu A_r (H_r)^{\nu-1} (L_r)^{1-\nu}, \quad (F.14)$$

and are therefore never unemployed. For every fixed wage $\bar{w}$ above the market
clearing level, the unemployment rate in region $r$ is given by

$$U_r = (\bar{L} - L_r)/(\bar{L} + H_r), \quad (F.15)$$

An increase in $H_r$ lowers the unemployment rate in the respective region for two
reasons: Firstly, as a pure matter of accounting, since the fully employed skilled
workers are counted within the region’s labour force. And secondly, the involve-
ment of more skilled workers in the production of region $r$ adds to the productivity
of the local unskilled workers. This pecuniary external effect has already been
discussed above. For any given $\bar{w}$, firms are more willing to hire unskilled labour
because of the enhanced productivity.

Suppose again that both regions initially have the same TFP-level, and human
capital is equally distributed across the two regions. Furthermore, suppose that
there is unemployment because of a non-competitive level of $\bar{w}$. From (F.13) and
(F.15) it follows that unskilled labour demand, unskilled employment, as well as
unemployment rates will be identical across regions. And from (F.14) it is easy to
see that this initial situation is an equilibrium, since there is no incentive for high
skilled workers to change locations.

Now suppose that the economy is hit by a comparable asymmetric exogenous
shock like in the last section. Only region 2 is affected negatively, i.e. $A_1 > A_2$. For
the case of Germany, one might think of this shock as representing the oil crises
of the 1970s, which surely affected all German areas, but the North (region 2) with
its traditional industries like shipbuilding more adversely. As an immediate
effect of this shock, both unskilled labour demand and high skilled wages in region 2 decrease. But it follows from (F.14) that additional to the direct impact, \( w_2^H \) is also negatively affected by the decrease in the unskilled employment of region 2. As a result, human capital will flow out of the region 2. Unemployment will further increase thereby, whereas unemployment drops in the region 1. Importantly, as long as there is unskilled unemployment in region 1 at the given union wage \( \bar{w} \), skilled workers are not subject to diminishing marginal returns as they migrate to region 1. This can be seen by substituting (F.13) into (F.14). As long as unskilled labour is not fully employed, i.e. as long as employment is given by labour demand (F.13), the human capital remuneration can be written simply as a function of exogenous variables.

\[
  w_r^H = v \left( A_r \right)^{1/v} \left( \frac{1-v}{\bar{w}} \right)^{(1-v)/v} 
\]  

Thus, as long as unskilled employment is given by (F.13), high skilled workers move to region 1 that has the higher TFP-level. As human capital concentrates, the marginal productivity of unskilled labour and thereby labour demand increases due to the pecuniary externality inherent to the Cobb-Douglas technology. Put differently, an increase of \( H_1 \) is not a partial, but a total factor variation as long as there are unemployed unskilled workers available who can enter the newly created jobs.

Note that we have just described a cumulative causation mechanism in this extremely simple model: some skilled workers change regions and thereby “create jobs” for unskilled workers in region 1 and destroy some in region 2. This in turn has positive drawbacks on productivity and wages of skilled workers in \( r = 1 \), negative ones in \( r = 2 \). And effectively, the motivation for human capital to concentrate in the advantaged region is even stronger than before. A small technological shock can have pronounced effects, since it starts off a vicious cycle for the disadvantaged region. To which extremity this type of circular logic is taken depends on the magnitude of the asymmetric shock, and on the prevailing wage level \( \bar{w} \).

**Proposition F1.** If \( \bar{w} \) is set on a level equal or higher than \( \hat{w} = (1-v)A_1 \left( \bar{H}/\bar{L} \right)^v \), high skilled workers will completely concentrate in the South. If the union wage is only \( \bar{w} = \gamma \hat{w} \), with \( \gamma < 1 \), the pooling of human capital might not be complete, but at least sufficient to restore full employment in the South.

In the case where \( \bar{w} \geq \hat{w} \), the consequences of an asymmetric shock – no matter how small it may be – are most extreme. Since enough unskilled workers are unemployed to begin with, the move of human capital to region 1 is never associated
with diminishing returns, and all skilled workers move to region 1 that has the higher TFP-level. If $w = \hat{w}$, there would be full concentration of human capital ($H_1 = \hat{H}$) and full employment for the unskilled in region 1. With $w > \hat{w}$, there remains some unemployment in region 1 despite of full human capital concentration.

If the union wage is fixed on a lower non-competitive level $\bar{w} = \gamma \hat{w}$, human capital will flow into the South at least until full employment is reached.\footnote{The parameter $\gamma$ might reflect the union or insider power or the degree of labour market imperfection in the economy. Note further that the union cannot set $\gamma < (\frac{1}{2})^\nu$, because this would be the market wage with equal distribution of human capital.} This is the case if $L_1 = \left( (1 - \nu) A_1 / \gamma (1 - \nu) A_1 (\hat{H} / \bar{L})^\nu \right)^{1/\nu} H_1 = \bar{L}$, which can be rewritten as

$$H_1 = \gamma^{1/\nu} \bar{H}.$$  \hfill (F.17)

Up to the pooling level (F.17), unskilled labour is not a binding factor. But once $H_1$ is as large as (F.17), skilled workers who are still located in region 2 face the trade-off that is already known from the full-employment model in the last section: Due to the higher TFP-level $A_1$, it is attractive to move to region 1. But since there is now full employment of unskilled workers, i.e. $L_1$ is given by $\bar{L}$ instead of $L_1^d$, human capital now faces diminishing returns.

But one can show that the human capital concentration in region 1 will be stronger in equilibrium than described by (F.17). Consider the equilibrium condition $w_1^H = w_2^H$ for the case where $L_1 = \bar{L}$. Using (F.14), this yields after some manipulations

$$\frac{H_1}{H - H_1} = \left( \frac{A_1}{A_2} \right)^{1/\nu} \frac{L}{L_2^d},$$

or

$$H_1 = \left( \frac{A_1}{A_2} \right)^{1/\nu} \bar{L} \left( (1 - \nu) A_2 / \bar{w} \right)^{1/\nu}.$$

We know that the prevailing wage is $\bar{w} = \gamma \hat{w}$, and we can therefore rewrite this expression as
The higher is the difference in TFP, and the higher is the union power \( \gamma \), the greater is the human capital concentration \( \frac{H_1}{H} \), and the wider are the regional unemployment disparities. Since \( A_1 > A_2 \), the equilibrium pooling level (F.18) is stronger than the full employment threshold for unskilled workers in region 1, equation (F.17). Therefore, there is excess demand for unskilled labour in region 1 at the going wage rate \( \bar{w} = \gamma \hat{w} \). This can be verified by plugging the equilibrium pooling level of human capital, (F.18), in the demand function for unskilled labour, (F.13).

\[
L_1^d = \left( \frac{A_1}{A_2} \right)^{\frac{1}{1-\nu}} \bar{L} > \bar{L}
\]  

(F.19)

The excess demand can push up the effective wages of unskilled labour in region 1. This has no effect on the equilibrium pooling level of high skilled workers, since this depends on the actual level of employment \( \bar{L} \), not on labour demand for unskilled workers. An upward tendency of \( w_t^{l} \) above \( \bar{w} \) is to expect, because collective bargaining wages are wage floors with allowance for an upward wage drift. The wage for unskilled labour \( w_t^{l} \) can lie in the range

\[
\bar{w} \leq w_t^l \leq \bar{w} \left( \frac{A_1}{A_2} \right)^{\frac{1}{1-\nu}},
\]  

(F.20)

without having any effects on unskilled employment or wages for the skilled workers. If \( w_t^{l} \) is equal to the upper bound of (F.20), labour demand will exactly match labour supply \( \bar{L} \). But where exactly in this range the actual wage will be, is a matter of bargaining and can not be answered by this model.

F5.3.) Conclusion of the model with constant returns to scale
Let us summarize the model variant from section F5.2.). We have shown that selective internal migration results in regional divergence with respect to both wages and unemployment rates. Starting from a symmetrical initial situation, an exogenous asymmetric shock induces human capital to migrate from region 2 to region 1. As a consequence of this factor movement, jobs are created and unemployment
is decreasing for the low skilled workers in the receiving region, the opposite happens in the sending region. This is due to the mutually beneficial relationship, or complementarity of input factors within a substitutional production function. Higher unskilled employment in region 1 increases the incentives for high skilled workers to leave region 2 even more. At the end of this cumulative causation mechanism, the model predicts that human capital is either completely or partially concentrated in one region, the unemployment rate in that location is lower, and unskilled workers have higher effective earnings. The latter results hold despite of undifferentiated union wages, because excess demand for unskilled labour gives rise to a positive wage drift of effective over contracted salaries.

The fundamental insight from the model of section F5 is that labour mobility does not lead to regional convergence if migration is a selective process. Even within a very simple neoclassical model, that did not make use of any agglomeration or scale effect, migration of high-skilled labour leads to divergence of regional unemployment rates and wages for the unskilled immobile workers. There are thus two truly distinct arguments for the "non-convergence"- character of labour mobility. Firstly, convergence might be ruled out either because the underlying technology exhibits (internal or external) scale economies, so that labour does not face diminishing marginal returns. This has been demonstrated in chapter E as well as in section F4 for a situation where labour was assumed to be a homogeneous input factor. And secondly, convergence does not occur – even under a very conventional neoclassical technology – because internal migrants tend to be high-skilled workers whose emigration does not help but rather hurt the sending region.

The two arguments might be substitutes or complements. Labour mobility can result in spatial divergence either because one of the two described forces is at work, or because they are jointly operating. Various authors have put different emphasis on either of the two sources for non-convergence. Krugman (1991a) e.g. relies on increasing returns to account for the divergent trend brought about by labour mobility, but he neglects the heterogeneity of labour. Other authors, like e.g. Myrdal (1957), are more inclined with the selectivity argument. In the remainder of this chapter, we want to construct a theoretical model that finally pulls all pieces together. We will formulate a model that explicitly entails regional agglomeration and worker heterogeneity in a unified framework.

**F6) Agglomeration, worker heterogeneity and national union wage setting**

The model we are going to develop in this section will be suited to explain the emergence of regional disparities *endogenously*. In the model with constant returns to scale, spatial differences were due to exogenous TFP-shocks, the origin of which has essentially remained an open issue. With endogenous agglomeration economies, such exogenous ad-hoc constructions are no longer needed. In this respect, the framework from this section is again similar to the agglomeration model from chapter E, in particular the production side of the economy. Yet, some
extensions and modifications are needed in comparison to the framework of chapter E, since we want to explicitly introduce worker heterogeneity.

We keep on considering an integrated economic area consisting of two regions labelled \( r = \{1,2\} \). Each region is populated now with two generations of heterogeneous agents, i.e. we construct an overlapping generations (OLG)–model. For simplicity we abstract both from output and population growth, and we assume that in each period a new generation \( L_r \) is born. Young and old individuals in both regions are endowed with one unit of non-leisure time and endogenously decide on the time fraction used for education. Human capital investments pay off in the old age period by expanding the available effective time budget that is then solely used for working. Heterogeneity enters insofar, as learning productivity differs across individuals depending on endowed personal ability characteristics denoted by \( \eta^r \). We therefore do not only distinguish between two groups of workers, skilled and unskilled. We rather (realistically) model skill/education as a continuous variable.

Individuals are tied to their original location of birth during young age. They can, however, emigrate after the first period and spend their second lifetime period in the other region. The incentive for emigration comes from the fact that the technology is characterised by localised increasing returns to scale, i.e. there is an economic benefit from spatial concentration of workers in only one location. But labour mobility is assumed to be costly. More specifically, the migration of an individual from one region to the other imposes costs equal to \( m \), which accrue independently of the agent's skill level.\(^{12}\)

Individuals in this model thus make two endogenous choices at the beginning of their lifetime: they decide on their individual education demand, and they decide on their old age residence region. At first (in F6.1.) we will consider a full-employment version of the extended model, where the labour markets in both regions are flexible and will always clear. In section F6.2.), the issue of unemployment will again be approached in a way comparable to section F5. For both versions of the model we are basically establishing two propositions. Firstly, we show that the individual propensity to be regionally mobile increases with the personal level of education. And secondly, we will show that this selective labour migration, in combination with the increasing returns technology, will lead to a divergence process with respect to regional wages and regional unemployment rates. Labour mobility does thus not cure regional disparities, but rather makes them worse.

\(^{12}\) For a discussion of this assumption see the discussion from section F4.2.)
Internal migration and regional disparities

F6.1.) The full employment case

a) Consumer behaviour

The consumer side of this model is similar to the OLG-model of Haque/Kim (1995). Let us consider an individual i who is born in region 1, and who derives utility from consumption. To keep the analysis tractable, we work with a logarithmic utility function with a time discount rate 0<β<1. The regional superscript \( s = \{1,2\} \) denotes the residence region of the individual at old age. If \( s = 1 \), the individual remains in her original area of birth. Otherwise (if \( s = 2 \)) she emigrates after the first period. Lifetime utility \( U_{i,s}^{i,rs} \) is given by

\[
U_{i,s}^{i,rs} = \log c_{i,t}^{i,ls} + \beta \log c_{i,t+1}^{i,ls},
\]

where \( c_{i,t}^{i,ls} \) denotes consumption of some individual i born at time t in region 1 and residing in region \( s = \{1,2\} \) at time \( t+1 \). Analogous equations apply to individuals born in region 2. \( \ell_{i,t}^{i,ls} \) is the time fraction devoted to education during young age. There are no direct costs of education, but only opportunity costs for foregone earnings. There are also no financial markets and hence no savings in the model, i.e. the education choice is the only means for consumption smoothing.

The budget constraints can be written as

\[
c_{i,t}^{i,ls} = w_{1,t}(1 - \ell_{i,t}^{i,ls}) \]

\[
c_{i,t+1}^{i,ls} = w_{s,t+1}(1 + \eta' \ell_{i,t}^{i,ls}) - m_{i,s}
\]

The variables \( w_{1,t} \) and \( w_{s,t+1} \) denote the wage per effective labour unit devoted to work in the respective region and time period. Mobility costs \( m_{i,s} \) arise only for individuals who choose to leave region 1 after the first period of life.

\[
m_{i,s} = \begin{cases} 0 & \text{if } s = 1 \\ m > 0 & \text{if } s = 2 \end{cases}
\]

Utility maximization with respect to the education choice \( \ell_{i,t}^{i,ls} \) yields the following first-order-condition

\[
\frac{c_{i,t}^{i,ls}}{c_{i,t+1}^{i,ls}} = \beta \eta' \frac{w_{s,t+1}}{w_{1,t}}.
\]
The optimal consumption path is thus simply proportional to the individual's income stream. Together with the intertemporal budget constraint from (F.22) and (F.23), the optimal education choice can be computed as

$$\ell^{*,ls} = \frac{\beta}{1 + \beta} - \frac{1 - (m_{ls}/w_{s,t+1})}{\eta(1 + \beta)},$$

(F.26)

**Proposition F2.** $\ell^{*,ls}$ increases with $\eta$ and $m$. It decreases with $w_{s,t+1}$, and it is greater if $s=2$ than if $s=1$.

An evaluation of (F.26) shows that more able people spend more time on education than individuals with a low learning capability $\eta$. Interestingly, individuals who plan to emigrate after period $t$ ($m_{ls}=m$) ceteris paribus demand more education than do people who are going to remain in region 1 also in $t+1$ ($m_{ls}=0$). The anticipation of future emigration already induces stronger educational attainment today, which is an argument close to Stark/Helmenstein/Prskawetz (1997). The effects of $w_{s,t+1}$ on the optimal learning choice of emigrants represents an income effect in the intertemporal smoothing of the lifetime earnings profile. By substituting (F.26) into the budget constraints, we can compute the optimal consumption path for given residence choices. Provided that the individual will remain in her home location $r=1$ during $t+1$, she will reveal the following consumption profile

$$c_{t,t+1}^{i,1} = w_{1,t} \left( \frac{1 + \eta}{\eta(1 + \beta)} \right)$$

$$c_{r,t+1}^{i,1} = \frac{\beta}{1 + \beta} w_{1,t+1} (1 + \eta)$$

If she spends her second lifetime period in region 2, the consumption path is

$$c_{r,t}^{i,2} = w_{1,t} \left( \frac{1 + \eta - m / w_{2,t+1}}{\eta(1 + \beta)} \right)$$

13 The negative impact of $w_{2,t+1}$ on $\ell^{*,ls}$ actually requires that $(m_{ls}/w_{2,t+1}) > 1$. 

Downloaded from PubFactory at 08/10/2019 04:16:27PM via free access
and

\[ c_{i,t+1}^{i,12} = \frac{\beta}{1 + \beta} w_{2,t+1} \left( 1 + \eta' - m / w_{2,t+1} \right), \]

Not only education activity, but also consumption in both periods differs depending on the old age residence choice that is anticipated in the first period of lifetime. By inserting these consumption levels in the utility function (F.21), we can compute individual i’s utility levels for the case that she remains in her original location \( (U_{i,11}) \), and for the case of emigration \( (U_{i,12}) \).

\[ U_{i,11} = \log \left( 1 + \eta^i \right)^{1+\beta} w_{1,t+1}^\beta K^i \]  \hspace{1cm} (F.27)

\[ U_{i,12} = \log \left( 1 + \eta' - m / w_{2,t+1} \right)^{1+\beta} w_{2,t+1}^\beta K^i \]

where \( K^i = w_{1,t} \left( \frac{1}{\eta' (1 + \beta)} \right) \left( \frac{\beta}{(1 + \beta)} \right)^\beta \).

An individual i will reside in that region during old age that offers the higher utility level for given unit wage rates \( w_{1,t+1} \) and \( w_{2,t+1} \). By equating \( U_{i,11} \) and \( U_{i,12} \) we find after some manipulation the level of personal ability \( \tilde{\eta}^i \) at which an individual is indifferent between migrating and remaining in region r for given unit wages

\[ \tilde{\eta}^i = \left( \frac{m / w_{2,t+1}}{1 - \omega_1} - 1 \right) \]  \hspace{1cm} (F.28)

with \( \omega_1 = \left( w_{1,t+1} / w_{2,t+1} \right)^{1+\beta} \)

\( \omega_1 \) measures the unit wage of region 1 relative to region 2. It can be shown that individuals with personal skills below \( \tilde{\eta}' \) derive higher utility from remaining in region 1 \( (U_{i,11} > U_{i,12}) \), whereas individuals with skills larger than \( \tilde{\eta}' \) are better of spending their second lifetime period in region 2. Thus, (F.28) can be understood as a cut-off ability level beyond which migrating to region 2 is more attractive than staying in region 1.

Individuals of course only consider to emigrate if the unit wage in region 2 is higher than in region 1, i.e. (F.28) is only a meaningful condition as long as \( \omega_1 < 1 \). We will assume that this condition is satisfied initially. More specifically, we take
the unit wage of region 2 as exogenously given at first instance, and assume that it is higher than the initial value of \( w_{1,t+1} \). This renders an initial migration incentive. Below we will fully endogenise the unit wage of region 2. The important message for the time being, however, is that if region 2 offers a higher unit wage level, only the high skilled workers from region 1 will migrate.

The intuition for this result is central to this model. Note that the agglomeration wage premium is higher for single individuals the higher is the number of embodies labour units. This is so, because \( w_{s,t+1} \) denotes a unit wage. Skilled workers therefore have a greater incentive to work in the central region than unskilled workers. The skill bonus in the agglomeration wage premium is a well established empirical fact, as it was argued above in section F4.2b). Together with the skill-invariant mobility costs, it is clear that the gains from migrating only overcompensate the costs for individuals with a high number of embodied labour units. Low skilled workers with low levels of \( \eta^i \) would also like to move from region 1 to region 2 in order to increase their income level. But for them, the wage premium for living in the "core" region 2 is not high enough to compensate for the mobility costs.

What fraction of each young generation \( L_1 \) has learning abilities larger than \( \bar{\eta}^i \) is a matter of the distribution of learning skills across the young population. Suppose that \( \eta^i \) is uniformly distributed across the \( L_1 \) individuals in the range \([1;d]\), i.e. the least talented individual (indexed \( i=0 \)) can not expand her effective labour units through education, whereas the average learning efficiency is \( 1+d/2 \). With this distribution of skills, the fraction \( \mu \) of the population \( L_1 \) that is going to remain in region \( r \) is given by

\[
\mu = \frac{1}{d}\left(\bar{\eta}^i - 1\right) \tag{F.29}
\]

From (F.28) and (F.29) it can be seen that emigration is attractive to a smaller fraction of the population (i.e. \( \mu \) is larger), the higher is the regional unit wage \( w_{1,t+1} \) relative to the (given) level in region 2 (\( w_{2,t+1} \)), and the higher are mobility costs \( m \).

**b) The production side**

We now turn to the production side of this economy, which is structurally almost identical to the model from chapter E. There is a single final consumption good \( Y_r \) which is produced in both regions \( r = \{1,2\} \) without direct use of labour by assembling a large number of symmetrical intermediate inputs \( X_r \). We keep the assumption of perfect competition and perfect tradability in the \( y \)-sector, which renders the numeraire price \( p^y = 1 \) that is valid in both regions.
The interregional transport of intermediate inputs $X_r$ on the other hand imposes the usual 'iceberg'-costs. Let the production function for $Y_r$ in each region again be given by a symmetrical CES function that is described in equation (E.13.). Due to symmetry, we can again write the minimum unit cost function in region $r = 1$ as

$$G_1 = \left( N_1(p_1)^{\theta-1} + N_2(p_2)^{\theta-1} \right)^{\frac{1}{\theta}}$$  \hspace{1cm} (F.30)

An analogous equation applies to region $r = 2$. The production of the single intermediate inputs in either region is still done by small, monopolistically competitive firms that use labour only. But here a change is needed compared to the construction introduced in section E.3.1.), since we no longer work with homogeneous workers. We assume that firms now demand effective units of labour $\zeta$ rather than a certain number of working people. Recall from above that single individuals do now embody different numbers of labour units, depending on endowed personal ability characteristics and on the individual educational choice. The requirement of labour units necessary to produce the quantity $X$ of an intermediate good is given by

$$\zeta = a + bX \hspace{1cm} \text{with} \hspace{0.2cm} a>0, \hspace{0.2cm} b>0$$  \hspace{1cm} (F.31)

Every single intermediate will be produced by only one firm and $N_1$ indicates the number of active firms in region $r = 1$. Profit maximizing prices are a constant mark-up over marginal costs, $p_{1,t} = (b/\theta)w_{1,t}$, and firm's profits are driven down to zero by the entry of potential competitors, i.e. $\pi_1 = p_{1,t} X - w_{1,t}(a + bX) = 0$. Solely for notational convenience we again choose units such that $b=0$. By using (F.31), we can rewrite the zero profit condition in the $X$-sector as

$$X = \zeta = \frac{a}{1-\theta}$$  \hspace{1cm} (F.32)

Equation (F.32) is analogous to (E.5.), i.e. it shows the equilibrium scale of each single firm regardless in which region it is operating. The output scale is identical with the effective demand for labour units per firm. Note, however, that firm sizes can very well differ with respect to the number of employed persons, as a firm does not care if it employs one worker with $\zeta$ embodied labour units or $\zeta$ workers with one labour unit each.
As before, the maximum number of intermediates that can be produced in region \( r \) is restricted by effective regional labour supply. Let \( S_{r,t} \) denote labour supply in region \( r \) at time \( t \). The number of firms and varieties is then simply

\[
N_{r,t} = \frac{S_{r,t}}{X} = \frac{(1-\theta)}{a} S_{r,t}.
\]  

(F.33)

It is now straightforward to compute the equilibrium remuneration per labour unit \( w_{r,t} \). This is done in the same way as in section E3.2.). The equilibrium condition of this model is that the unit costs \( G_r \) must be equal to one in both regions. Using (F.30) this yields the familiar result that regional unit wages are an increasing function of regional labour supply alone.

\[
w_{r,t} = \left( T N_{r,t} \right)^{\frac{1-\theta}{\theta}} = \left( \frac{T}{X} S_{r,t} \right)^{\frac{1-\theta}{\theta}}
\]  

(F.34)

where the economic interpretation of the parameter \( T = (1-\tau^{\frac{2\theta}{\theta-1}})/(1-\tau^{\frac{\theta}{\theta-1}}) \) is as in the preceding chapter E.

In this benchmark version of the model, labour supply is equal to employment. This implies that the unit wage consistent with product market equilibrium is an upward sloping function of employment. This is intuitive, as more labour supply in region 1 implies a higher number of locally produced intermediate inputs. Unit costs \( G_r \) decline, while the price \( p^Y \) remains unchanged. Temporary profits arise in the Y-sector that induce producers to enter the market. Prices for intermediates \( X_r \) are competed up, and these higher prices must completely be absorbed by higher unit labour remunerations via the zero profit condition in the X-sector.

Our assumption from above, that the unit wage in region 2 is initially higher than in region 1, can thus be rationalized by assuming the each new born generation in region 2 is larger than in region 1. In other words, \( r = 2 \) is assumed to be the larger region, which translated via the embedded scale effect in a higher regional unit wage level.

c) Effective labour supply

Due to the increasing returns technology, we have again embedded a centripetal tendency that has been intensively discussed in section E4.1.b). Hence, given the technology alone, all workers regardless of their skill level would want to concentrate in only one region. Due to the initial advantage, this pooling process will
occur in region 2.\footnote{One can show that in our model a “core-periphery” structure would also develop endogenously when starting from a completely symmetrical initial situation. Which region is going to be the core, and which one the periphery would be indeterminate. But the centripetal tendencies, the gains from concentration that re to exploit, are also valid if both regions are ex-ante identical, i.e. the symmetrical equilibrium is unstable.} However, not all workers born in \( r = 1 \) will actually change regions after the first period of lifetime, since we have also introduced an opposing “quasi-centrifugal” force in this approach. This is no longer the prevalence of ad-hoc regional preferences as in chapter E, but rather the presence of mobility costs. Since interregional migration is costly, only such individuals will migrate for whom the agglomeration wage premium at least compensates for the mobility costs. As shown in (F.28) and (F.29), these are the high skilled individuals with abilities greater than the critical level \( \eta^i \). In other words, the workers who remain in region 1 (i.e. those who do not participate in the regional pooling process) are now the low skilled workers for whom migration does not pay off. It are no longer such individuals who feel intrinsically attached to the peripheral location as in section E3.3b).

We can now explicitly derive an expression for the effective labour supply in region 1 at time \( t \). This variable \( S_{r,t} \) does not only depend on the population size in region 1, but also on the education and migration decisions of the individuals. Labour supply at time \( t \) consists of the number of labour units that the two generations offer. For the young generation with size \( L_1 \), this is the amount of time that they do not devote to education. The old generation only has the size \( \mu L_1 \), since the \((1 - \mu)L_1\) most talented workers spend their old age in region 2. Recall that members of the young generation reveal different education demand depending on their old age residence choice. Effective labour supply in region 1 can be written as

\[
S_{1,t} = \int_{i=0}^{\mu L_1} \left( 1 - \phi_{i,11} \right) \, i = \mu L_1 \int_{i=0}^{L_1} \left( 1 - \phi_{i,12} \right) + \int_{i=0}^{\mu L_1} \left( 1 + \eta^i \phi_{i,11} \right)
\]

or

\[
S_{1,t} = (1 + \mu)L_1 + \int_{i=0}^{\mu L_1} \phi_{i,11}(\eta^i - 1) - \int_{i=\mu L_1}^{L_1} \phi_{i,12}
\]

(F.35)

The first term in the second row of (F.35) represents the pure population size that is constant in the steady state when \( \mu \) is at its equilibrium level. The second term are the net returns to education of those who remain in region 1 also in the second period. The third term indicates the costs for region 1 that arise because later emigrants do not use their entire time budget for working. From (F.35) it can be seen that \( S_{1,t} \) increases with \( \mu \) for several reasons. Firstly, because the pure population
mass is larger the fewer people emigrate to region 2. Secondly, because more people realize the returns to education in region 1. And thirdly, because fewer opportunity costs arise in region 1 for educating people whose private and social returns will be realized elsewhere.

It also becomes clear that the linkage effect that runs from labour supply to equilibrium remunerations in (F.34) can represent both a pure scale effect and a human capital externality: $S_{1,t}$ and thereby $w_{1,t}$ can be high either because many people are around ("agglomeration wage premium"), or because they embody a high number of labour units.

d) Equilibrium

With the model that has been outlined in this section, we have established a circular, or cumulative causation mechanism. We have derived the fraction $(1-\mu)$ of each generation $L_1$ that derives a higher lifetime utility by leaving region 1 after the first period of lifetime for given regional unit wages. This fraction is larger, the lower is the unit wage rate in region 1 relative to region 2. On the other hand, we have shown that the equilibrium wage in region 1 decreases, the lower is labour supply. Put differently, people leave if wages are low, and wages are low if people leave.

This circular logic in particular applies to individuals with strong learning capabilities $\eta^1$. Their emigration has a stronger bearing on region 1, firstly because they have demanded a high amount of education during young age. At time $t+1$, when the investment pays off both privately and socially, the high skilled workers leave the small region, which consequently foregoes the positive linkages that are associated with their human capital. Put differently, high skilled workers embody a higher amount of effective labour supply in our model. And since the individual earnings of each individual positively depend on the overall labour supply in the region of residence, the emigration of high skilled workers out of region 1 is particularly harmful for all workers who remain in that location.

The purpose of this section is to derive the equilibrium level of emigration $(1-\mu^*)$ out of region 1 for given exogenous parameter values $m$, $\theta$ and $w_{2,t}$. Note that we will continue to treat the wage in the core region $r = 2$ as an exogenous parameter in this section, and we assume that it does not change with $\mu$. An exact analytical expression of $\mu^*$ can be obtained by plugging (F.34) and (F.35) into (F.28) and (F.29). This yields a rather complicated expression that in principle can be solved

---

15 With the adopted technology, the effective labour supply in region 2, $S_{2,t}$, and thereby the regional unit wage $w_{2,t}$ would endogenously increase as workers immigrate from the small region 1. We come back to this issue in below.
for \( \mu^* \).\footnote{The (highly non-linear) expression that has to be solved for \( \mu \) is given by}

\[
\mu = \frac{1}{d \Phi_1} \left( \frac{m}{\bar{w}_{2,t+1}} / \left[ 1 - \frac{1 - \theta}{a} S_{1,t} (\mu) \right] - \frac{1 - \theta}{\bar{w}_{2,t+1}} - 1 \right) - \frac{1}{d}
\]

For expositional purposes, however, we have chosen to pursue an instructive graphical approach, which is presented in figure F3.

**Figure F3: The determination of \( \mu^* \)**

The locus \( V_0V_1 \) is derived from (F.29), the optimal residence choice based on consumers’ utility maximization. It shows the fraction \( \mu \) of the generation \( L_1 \) that chooses to remain in region 1 as a function of \( w_{1,t} \) and for given parameter values. The positive slope represents the fact that \( \mu \) is increasing in \( \omega \). The locus \( R_0R_1 \) represents the technological relation (F.34) and depicts equilibrium unit wages \( w_{1,t} \) as a function of labour supply \( S_{1,t} \), which is endogenously increasing in \( \mu \).

Within the feasible range \( \mu \in \{0;1\} \), the adjustment mechanisms in this system work as follows: for points above (below) the \( R_0R_1 \) schedule, the wage \( w_{1,t} \) is too high (low) for any given value of \( \mu \). Using the zero profit condition described above, the wage must realign such that it is consistent with the equilibrium locus.
R_0R_1. This determines the phase arrows in the vertical direction. Similarly, for points to the right (left) of V_0V_1, µ is too high (low) for any given wage w_{1,t}. Individuals can still increase lifetime utility through changing locations, and migration will occur until µ is consistent with V_0V_1.

As long as V_0V_1 is steeper than R_0R_1, which will be the only case we consider throughout, there is a unique and stable equilibrium at point A with a spatial equilibrium configuration µ*. This level µ* is consistent both with efficient production and with optimal residence choice.

It is instructive to look at some comparative statics and analyse changes in the exogenous parameter m, w_{2,t} and θ. Changes in θ are easiest to study, since only the R_0R_1 locus is affected. The parameter θ reflects the differentiability of the single intermediate inputs in region 1 and can be understood as an inverse measure of the degree of increasing returns. The higher is θ, the lower is the equilibrium wage w_{1,t} for any given value of µ. The curve R_0R_1 shifts down as θ increases, which implies that µ* is a decreasing function of θ. There is thus more population drain from region 1 the less important are the localized increasing returns.

A change in the (exogenous) wage w_{2,t} affects both curves in figure F3. The impact on V_0V_1 is obvious: if the attainable wage in region 2 increases, the incentive to leave home after the first period is larger for given values of w_{1,t} and m. The curve V_0V_1 is shifted to the left. The curve R_0R_1 is also affected, however, because education demand and thereby labour supply change. This can be seen best by considering the following: the point R_1 shows the equilibrium wage w_{1,t} if nobody of the young generation L_1 will emigrate after the first lifetime period. This wage can be computed as

\[ w_{1,t} = \frac{1}{X} \left[ \frac{2L_1}{\theta} + \sum_{i=0}^{L_1} \left( \mu^* \eta_i - 1 \right) \right]^{(1-\theta)/\theta}, \]

which is independent of w_{2,t}. Yet, at all other points along the R_0R_1 schedule, any given fraction of later emigrants (1-µ)L_1 will spend less time on education as w_{2,t} increases. This consequently increases labour supply of later emigrants during their young age in region 1 and thus has positive impacts on wages w_{1,t} for any given value of µ. Graphically, an increase in w_{2,t} implies a clockwise rotation of R_0R_1 around the point R_1. The net effect of an increase in w_{2,t} on µ* is thus theoretically ambiguous.

A similar point applies to changes in the parameter m, the level of mobility costs. V_0V_1 shifts to the left as migration barriers are removed, because emigration is

\[ \text{In the other case with } R_0R_1 \text{ steeper than } V_0V_1 \text{ the system is characterised by dynamic instability of } \mu^*, \text{ and will in general be driven towards a corner solution.} \]
more attractive for given values of $w_{1,t}$ and $w_{2,t}$. But again, a decrease in $m$ implies a reduction in the education demand of later emigrants, thereby an increase in labour supply and thus a clockwise rotation of $R_0R_1$ around $R_1$. Supposedly (given some numerical simulations of parameter changes in both $w_{2,t}$ and $m$) the "direct" effects on $V_0V_1$ will dominate over the effects on $R_0R_1$ that originates in the individuals' intertemporal substitution, but theoretically the other possibility can not be excluded.

e) The effects of labour mobility

After having described the complete model structure, we can now draw conclusions about the impact of labour mobility. Recall that the model in this section features both worker heterogeneity and thus the element of selective labour migration, as well as endogenous agglomeration economies through the presence of localised increasing returns to scale.

Not surprisingly, it is easy to see that labour mobility is a device for regional divergence. The higher is the equilibrium fraction of emigrants from region 1, $(1-\mu^*)$, the lower is the regional equilibrium wage per effective labour unit, i.e. the lower are the per capita incomes in region 1. Contrarily to chapter E, it are specifically the high skilled workers who emigrate from the periphery to the core in this chapter. This is so, because the agglomeration wage premium is greater than the mobility costs only for such workers whose ability level is greater than some critical level. But in principle, due to the localised increasing returns to scale, everyone has an incentive to move to the core region 2.

The "unfortunate" consequences of labour mobility out of the already poorer area 1 become even more pronounced when we generalize our approach and explicitly model the wage formation in the core region 2. Consumer behaviour and goods production in region 2 is structurally identical to region 1 as described above. This specifically means that the final output $Y_2$ is manufactured under the use of $N_2$ symmetrical local intermediates $X_2$, and that the number of firms $N_2$, as well as the equilibrium producer wage for each effective labour unit $w_{2,t}$ are then functions of regional labour supply in region 2. By applying (F.34) to region 2 we find that

$$w_{2,t} = \left( \frac{T}{X} S_{2,t} \right)^{1-\theta}$$

We have assumed that labour supply in region 2 is initially larger, e.g. because each new born generation in region 2 is larger than in region 1. This translates to a higher unit wage in region 2, and thus implies that there is a migration incentive only from region 1 to region 2. If this is warranted, we can apply the consumer
problem also to individuals from region 2 and compute the optimal education choice $\ell^{*,22}$ as

$$\ell^{*,22} = \frac{\beta}{1+\beta} - \frac{1}{\eta' (1+\beta)}.$$

The overall labour supply in region 2, $S_{2,t}$, is given by

$$S_{2,t} = 2L_2 + \int_{i=0}^{L_2} \ell^{*,22}(\eta' - 1) + (1-\mu)L_1 + \int_{i=\mu L_1}^{L_2} \eta' \ell^{*,12}, \quad (F.36)$$

which is an increasing function of the immigrant population $(1-\mu)L_1$. Hence, $w_{2,t}$ is not independent of $\mu$, but it rather endogenously increases the more people emigrate from region 1. This has drawbacks on the optimal location decision of individuals from region 1 as described by (F.29). Since $w_{2,t}$ is larger the lower is $\mu$, the actual cut-off ability level beyond which emigration starts off is actually lower than implied by figure F3, where the endogenous impact on the $w_{2,t}$ has been neglected. This effect can again be graphically illustrated in figure F4.

The graphical relation that describes utility equalization across the two regions is not given by the $V_0V_1$-curve. It is rather given by some curve $V_2V_3$ that runs strictly to the left of the $V_0V_1$-schedule. For any given value of $w_{1,t}$, the corresponding value of $\mu$ consistent with interregional utility equalization is strictly lower, if individuals take into account the endogenous effects of migration on the wage in region 2. Equilibrium is thus reached at $A'$ rather than at $A$. The fraction $\mu$ of the generation $L_1$ that remains in region 1 during $t+1$ is only $\mu_2^*$ rather than $\mu_1^*$. If one acknowledges that the $R_0R_1$-schedule will rotate around $R_1$ in the clockwise direction (see the discussion above), the additional push for emigration out of region 1 is moderated to some extend. But the actual equilibrium value of $\mu^*$ will be lower than $\mu_1^*$, and the "true" equilibrium will lie somewhere between $A$ and $A'$.

The cumulative causation spiral described above is thus accentuated if we endogenise the wage formation of the core region. Any emigration out of region 1 puts the relative wage $\omega$ under strain from two sides. The true amount of brain drain has been understated by $\mu_1^*$. But other than that, the central insights from our theoretical approach remain qualitatively unchanged. Labour mobility leads to regional divergence of wages and income levels, firstly because of the underlying increasing returns technology. And secondly, this divergence is magnified by the fact that the emigrants are the high skilled, most talented workers.
F6.2.) The case with unemployment

The final thing to do is to let unemployment re-enter the theoretical analysis. This will be done in a very simple (and admittedly artificial) way. We will also discuss this model version only briefly and mostly verbally, since it is supposedly already clear that labour migration in this approach will likewise induce regional unemployment divergence (in analogy again to section F5).

For analytical simplicity, we assume that there is a national union, which does not set a minimum unit wage, but rather a minimum income level per employed person. This minimum income standard $\bar{W}$ is regionally undifferentiated and applies to individuals from both regions. Firms have to pay at least this minimum income to each individual that they wish to employ. In our view it is not unrealistic to assume this type of union policy, since in the real world unions often negotiate about the growth rate of monthly incomes rather than e.g. about hourly wages (which can be seen as a proxy for unit wages in our model). The reasons why we do not study a minimum unit wage, however, are mainly technical.\(^1\)

\(^1\) A minimum unit wage would imply that firms in the X-sector in both regions will charge product prices $p$, equal to the union unit wage. However, the zero-profit conditions and thus the product market equilibrium condition might be violated at the predetermined union level of $p$. But producers can not restore product market equilibrium by lowering the level of employment, since this
The consequences of this type of minimum income policy are straightforward. Firms in region \( r = \{1,2\} \) will only employ such workers who, for a given regional unit wage \( w_{r,t} \), embody enough effective labour units to render a profitable employment. The other workers, who do not possess enough human capital, will have to remain unemployed because of the union’s minimum standards. Unemployment will thus be centred around the low skilled individuals. More specifically, only those young individuals will be employed for which the following condition holds

\[
w_{r,t} (1 - \ell^{rs}) \geq \bar{W} \quad (F.37)
\]

Old individuals are also not fully employed, but only if

\[
w_{r,t} (1 + \eta \ell^{rs}) \geq \bar{W} \quad (F.38)
\]

The left-hand sides of the two inequalities represent the market income level of individual \( i \), which is given by the regional unit wage times the number of effective labour units that individual \( i \) can offer. Individuals who do not cross this threshold have to stay (unwillingly) unemployed.

We could assume that the young unemployed individuals can use their “free time” to invest in education in order to increase the employment opportunities for the old age period, but we abstract from this mechanism again for analytical simplicity.

Recall that the unit wage in the core region 2 is higher than in the peripheral region 1. It is then clear that the periphery suffers more from this regionally undifferentiated union policy, because it is more difficult for individuals from region 1 to match the conditions (F.37) or (F.38) respectively. In other words, the peripheral location 1 will face a higher regional unemployment rate. Furthermore, the minimum income reduces employment and thus the effective labour supply in both regions (and particularly in region 1). This has, via (F.34), negative drawbacks on the equilibrium unit wage and thus on the income level of all workers, also the employed ones. In other words, the union policy not only invokes unemployment, it also depresses the earnings of employed workers because the agglomeration economies are only sub-optimally exploited.
If now migration from region 1 to the relatively advantaged region 2 starts off, labour supply is reduced further in the peripheral location. The opposite happens in the central region 2. Thereby unemployment will increase further in $r = 1$, because the unit wage $w_{1,1}$ decreases and even fewer individuals will cross the employment threshold (F.37) or (F.38). In region 2 on the other hand, immigration will lead to an increase of $w_{2,1}$, and more individuals will become employed.

To sum up, labour migration does – not surprisingly – also lead to a divergence trend with respect to regional unemployment rates. Unemployment is entered into the model of this section in an admittedly very simplistic way. But still it is sufficient to see the consequences of regionally undifferentiated union policy for spatial unemployment disparities, and to examine the impacts of geographical labour mobility.

**F7) Discussion of the theoretical models and their empirical relevance**

In this final section, we want to discuss how the models from section F5 and F6 relate to the approaches that have been presented in the preceding chapters, in particular to the wage curve/agglomeration-model from chapter E. Afterwards we want to assess the empirical relevance of the models from this chapter.

**F7.1.) Discussion from a theoretical point of view**

The first contribution of the models from chapter F is to provide a systematic analysis about the impact of labour mobility on relative regional income levels and unemployment rates. We have shown that the neoclassical convergence hypothesis only holds under very special circumstances. Under an increasing returns technology, or with selective labour migration (or a combination of the two), internal labour migration will lead to regional divergence rather than convergence with respect to both income levels and unemployment rates. This point has been discussed at length above and shall not be repeated here.

The second objective of this chapter has been to provide an alternative theoretical analysis about regional unemployment disparities that explicitly addresses the role of national labour market institutions for unemployment outcomes on the regional level. The approach is different from chapter E insofar, as the models from this chapter are no longer wage curve models. What are the reasons for the departure from the wage curve as an analytical tool? Recall that a wage curve shows how the local unemployment rate has an impact on the local wage level. In the vein of the efficiency wage model that we have used as the micro-foundation, the wage curve relation was understood as a regional non-shirking condition: The lower is unemployment in region $r$, the more efficiency wages had to be paid in order to restore labour market equilibrium in that location.

In section F5, there is no downward sloping labour market equilibrium curve that would constitute an inherent link between local labour market variables. The bottom line message is rather that a regionally undifferentiated union wage, which
does not take into account regional productivity differences, can lead to differentiated unemployment outcomes on an intra-national level. Hence, union wage setting matters for spatial unemployment disparities. But the channel through which it matters is different from the wage curve reasoning. We have argued in section C3.1.) that in principle a wage curve relation could also be rationalized on the basis of a collective bargaining model. The intuition of such a model would be that the bargaining power of a regional union is a decreasing function of the regional unemployment rate. As a consequence, regional wage claims are lower the higher is the regional rate of joblessness. An approach like this, however, that inherently relates local bargaining power with local labour market conditions is inappropriate given the institutional structures of most European labour markets and completely misses the point. Union wage setting affects regional unemployment precisely because there is so little intra-national differentiation in collective bargaining wages.

This point has been made e.g. by Faini (1999). He also builds a two-region model with mobile skilled and immobile unskilled labour. Unemployment is not a feature of his model. But union wage setting plays a role, because he distinguishes two sectors, agriculture and manufacturing. Unions set wages for the unskilled in the latter sector. Faini (1999) shows that unions, which fix national wages irrespective of local productivity conditions, depress labour demand particularly in the backward region. Due to the national union wage, unskilled workers particularly from the backward location are driven to the less well paid agricultural sector. A greater acknowledgement of regional circumstances through a regionalization of the bargaining process would moderate this process. Essentially the same point can be made for our regional unemployment models from this chapter. Unemployment disparities could be smaller if the unions would comply more to regional productivity levels.

F7.2.) Discussion from an empirical point of view
Even though the models from sections F5 and F6 are no wage curve-models from the theoretical point of view, their final results and implications are somehow consistent with the predictions of the wage curve literature.

In the models from this section, the low skilled workers who live in the economically backward region are disadvantaged along two dimensions. Firstly, the unemployment rate is higher than in the blooming location. And secondly, unskilled workers in the advantaged region have higher effective earnings. This was so in the model from section F5, because there is a positive wage drift of effective earnings over contracted salaries. And similarly in section F6, the wages in the advantaged region are higher because of a better exploitation of the agglomeration economies. All in all, high regional wages and low regional unemployment rates coincide in the theoretical models from this chapter. Such a negative correlation between wage levels and unemployment rates on a regional level is the basic pre-
diction of the wage curve literature. The subsequent migration flows worsen the regional disparities with respect to both wages and unemployment rates. The "wage curve" is thus again a seemingly stable socio-economic relation that is not eroded through labour mobility.

These theoretical results are broadly consistent with the stylised facts on the spatial structure of unemployment rates and effective earnings in West Germany as described in section A5. We have shown that there are large and persistent regional unemployment disparities between the North and the South. Workers in the low-unemployment region, the South, also have higher effective earnings, even though union wages do not differ. The migration flows from the North to the South (which have been going on since decades) have apparently not led to an erosion of the existing disparities. Hence, the theoretical models from this chapter seem to be applicable to explain the regional disparities in West Germany.

These theoretical predictions are broadly consistent with the stylised facts on the relevance of agglomeration and the wage curve that have been discussed in section E4.2.), and the models are thus relevant also in a wider sense. Yet, one important qualification is needed with respect to the empirical applicability. Unemployment results because a national union fixes minimum wages on a national level without any regional differentiation. Therefore the models from this chapter are really only applicable to explain intra-national regional disparities. This is so because collective bargaining is often coordinated within single countries, but not yet on an international scale. It seems fair to say that contracted wages are the same in the northern and in the southern part of West Germany. But this is not true e.g. for the case of Germany vs. France, or Spain vs. Portugal. Up to now, trade unions have not spent a truly notable effort to coordinate wage claims across the single EU-countries. In principle, however, a development in this direction is very well conceivable in the future when European integration proceeds. The slogan "equal pay for equal work" that is behind the low regional differentiation of contracted wages e.g. in Germany could in principle also be extended to an European scale. The theoretical models from this section suggest that severe economic problems would be associated with such a policy that pays low attention to spatial productivity differences. However, a "Europeanization" of collective bargaining has so far not become reality.

Hence, the models from chapter F are not suited to address the trans-national unemployment clusters in the EU-15 as a whole that we have talked about in the chapters A and E. But the models are applicable to the wide and persistent intra-national unemployment and wage disparities that exist in virtually all EU countries. The fact that the main theoretical predictions from this chapter (the negative correlation of wages and unemployment rates, the influence of agglomeration, the divergence character of labour mobility) are compatible with the results from chapter E indicates, that the two broad theoretical approaches which have been introduced in this book are complementing and reinforcing each other.