Chapter E)
Regional agglomeration and regional unemployment

E1) Introduction
During the last two chapters we have essentially laid out the theoretical foundations necessary for the analysis of regional unemployment in relation to regional agglomeration. In chapter C we have introduced the wage curve literature as a regional labour market theory. But we have argued that the product market structure in these models is problematic from a theoretical point of view, because regional disparities can only exist if they are exogenously assumed. Regional asymmetries can not develop endogenously.

In chapter D we have therefore introduced theoretical approaches that aim to explain the endogenous emergence of regional agglomeration and core-periphery patterns of economic activity. This literature descended from the insights of Marshall (1890), and is represented in modern economic theory through the models of NTT and NEG. However, this strand of theory abstracts from unemployment and focuses on the explanation of regional disparities in production and income. It can thus not be used to analyse spatial unemployment differences.

The aim of this chapter is to combine elements from both strings in the literature and propose a theoretical model that allows to analyse regional unemployment and regional agglomeration in a unified framework. There is very little that has been done in this “boarder region” of different fields within economic theory. Still, very little is more than nothing. The existing work on the relation of trade and location theory with unemployment theory will be discussed in section E2. We then turn to the introduction of our model approach in section E3. In section E4 we critically review our approach from a theoretical point of view, and against the available empirical evidence from the EU-15. Section E5 concludes and raises some further issues for the upcoming chapter.

E2) The existing literature
Our theoretical analysis is related closest with the NTT-model of Matusz (1996). Another contributions that is worth mentioning is the paper of Peeters/Garretsen (2000).

E2.1.) The model of Matusz (1996)
Matusz builds a NTT-model in spirit of Ethier (1982), where final goods are assembled under the use of intermediate inputs. The production of intermediates involves internal increasing returns to scale, whereas a greater variety of industrial inputs induces a pecuniary externality on the unit costs in the final goods sector. Matusz works with a two-country model without labour mobility. In autarky, each country can assemble final goods only with local intermediates. The more inter-
mediates can be produced in a nation, the more efficient is the final goods production and (via a zero profit condition) the higher is the real wage consistent with product market equilibrium. Unemployment enters in his model through a shirking approach and yields a labour market equilibrium condition that is analogous to a wage curve.

Matusz moves from an autarky setting to an analysis of international trade, but he does so in a quite extreme way: he assumes that once trade barriers fall, there is costless trade in both the final good and the intermediates sector across countries. The introduction of trade widens the scope of available intermediates and thereby reduces unit costs in the final goods sector for both countries. Since firms make zero profits, this higher efficiency is absorbed by higher wages and lower unemployment rates that necessarily must be equalized across countries. Since trade is costless, no element exists that could discriminate between the factor income in the different spatial units. The final result of Matusz (1996) is thus that international trade increases wages and (via the shape of the wage curve) must reduce unemployment.

The model of Matusz is discussed here, because we will use elements of his paper also for our approach. But it has already become clear that substantial changes are necessary in order to make the model structure applicable for regional issues. This firstly concerns the fact that there are no regional disparities in wages and unemployment rates across nations (regions). This artefact stems from Matusz’s assumption of perfect and costless trade. But since we are precisely interested in the nature of regional disparities, we generalize his model and introduce transportation costs. And secondly, a regional analysis must also derive implications with respect to labour mobility.

With the appropriate model extensions and modifications, our generalized version of the model comes much closer in spirit to the theories of regional agglomeration and NEG, and it provides then a useful framework to study the interrelation of agglomeration and unemployment on a regional basis.

E2.2.) The model of Peeters/Garretsen (2000)
The contribution from Peeters/Garretsen (2000) is noteworthy, since it is to our knowledge the only existing attempt where unemployment is integrated into a standard and full fledged NEG-model.

The research focus of Peeters/Garretsen is the impact of globalisation (falling transportation costs) on the labour market position of unskilled workers in advanced countries. They mean by that the skilled/unskilled-wage gap that might change as globalisation proceeds. But they also analyse the impact on unemployment if unskilled labour is subject to an exogenously given minimum wage. They formulate a NEG model with two countries (labelled as home and foreign), mobile skilled and immobile unskilled labour. They adopt a more complicated production setting compared to the standard Krugman-model. They consider a two-step pro-
duction process of manufacturing goods. Firstly, there are so called “unfinished” goods that are produced by unskilled labour only. To these raw goods, producer services of high skilled workers must be added in order to make the goods marketable. Both unfinished goods and services exhibit scale economies in the production, and both are tradable across regions subject to transportation costs. The standard agricultural sector with constant returns that employs low skilled labour only is also part of their model. In principle, they derive standard NEG-implications with respect to the impact of globalisation: the effect of falling transportation costs on the labour market position of low-skilled workers is ambiguous and depends on the range of transportation costs. They are able to make some further qualifications, as they can distinguish what type of transportation costs declines (for raw goods, for producer services, both). But we want to focus on their implications with respect to unskilled unemployment.

They consider different minimum wage regimes for the low-skilled workers in the manufacturing sector of the home country. All other labour markets, including that for low-skilled manufacturing workers in the foreign country, must clear for the analysis to be internally consistent. The prevalence of a minimum wage induces unemployment for the unskilled in the manufacturing sector of the home country. In view of falling transportation costs, parts of the globalisation impacts on low skilled workers do not occur through changes in the relative skilled/unskilled wage, but rather through changes in the unemployment exposure of the low skilled workers. Of course these conclusions can not be obtained as closed-form analytical solutions, but Peeters/Garretsen (2000) have to rely on numerical solution techniques.

For the purpose to analyse regional unemployment disparities within the EU-15, the approach of Peeters/Garretsen (2000) is hardly applicable. Recall that there will be unemployment only in the home country, whereas all foreign labour markets clear. For an analysis of spatial disparities, there should be unemployment in both regions. Furthermore, Peeters/Garretsen rely on institutional differences, since the minimum wage legislation can exist only in one country, but not in the other. For our purposes, however, we would like to see regions with identical labour market institutions in order to analyse if disparities in unemployment rates can still arise. Furthermore, Peeters/Garretsen do not incorporate a wage curve, or a labour market equilibrium relation between wages and unemployment, but rather impose a specific minimum wage. Hence, the model of Peeters/Garretsen, that should really be seen as pioneering work, can serve for us as a related piece of literature. But it is designed to address a totally different economic problem, not to analyse regional unemployment disparities. There is thus still enough room for us to formulate an own theoretical approach.
E3) Regional agglomeration and the wage curve: The model
The exposition of our model goes in three steps: first we present a closed-economy setting where there is neither trade in intermediate inputs, nor factor mobility. This setting is identical to the autarky model of Matusz (1996). We then allow for trade in intermediate inputs. But we generalize the Matusz-model to account explicitly for geographical factors by assuming the presence of iceberg costs for interregional transportation of intermediate inputs. Yet, we keep at first the assumption of immobile workers. The last step is to relax also this assumption and analyse the impacts of (imperfect) labour mobility.

E3.1.) The closed economy setting
Consider an economy that produces (without using labour) a final consumption good Y under the use of a large variety of N single intermediate inputs $X_i$. The production function of the final product Y is given by the symmetrical CES function

$$Y = \left( \sum_{i=1}^{N} X_i^\theta \right)^{1/\theta} \quad 0 < \theta < 1. \quad (E.1)$$

The parameter $\theta$ is a measure of the differentiability of single intermediate inputs. If $\theta$ is close to one, inputs are nearly perfect substitutes. The elasticity of substitution between the single intermediates is given by $\sigma = 1/(1-\theta)$.

The minimum cost function of producing Y can be obtained by minimizing total consumption expenditure subject to (E.1). This yields

$$C(p_1, ..., p_N, Y) = GY \quad \text{where} \quad G = \left( \sum_{i=1}^{N} p_i^{\theta/(\theta-1)} \right)^{\theta-1} \quad (E.2)$$

The term G in (E.2) can be understood as a minimum unit cost function for the final good Y. We apply the standard assumption of NTT that all intermediate inputs enter symmetrically into the production function. This implies that the production function simplifies to $Y = \left( N(X)^{\theta} \right)^{1/\theta}$, and furthermore implies that the minimum unit cost function becomes

$$G = \left( N(p)^{\theta/(\theta-1)} \right)^{(\theta-1)/\theta} \quad (E.3)$$

where p is simply the price of one of the symmetrical intermediates produced in that economy. As can be seen, the minimum unit costs decrease with N. Accord-
ing to Ethier (1982) and Matusz (1996), this intends to capture the famous “pin factory”-idea of Adam Smith. It is an advancement for an economy to have a deeper division of labour, i.e. more narrowly defined sub-steps in which a specific production task (Y) is partitioned.

An important assumption made by Matusz (1996), that will also be used here, is perfect competition in the Y sector, which ensures that profits must always be equal to zero. We furthermore use the price of the final good \( p_Y \) as the numeraire and normalize it to one. By the condition \( \pi_Y = 1 \), \( Y - G.Y = 0 \), we can easily derive the first product market equilibrium condition of the model. It is given by the requirement that minimum unit costs need to equal the product price

\[
G = \left( N(p)^{\theta/(\theta-1)} \right)^{(\theta-1)/\theta} = 1.
\]

Since Y is the only commodity in this economy that is directly consumed, also the consumer price index is given by the value \( p_Y = 1 \) at any time. This assumption is thus simplifying in another respect, as it allows us to abstract from the distinction between nominal and real wages. In equilibrium, prices of the intermediates will have to adjust such that this condition is satisfied.

Each of the N single intermediates \( X_i \) is produced by using labour only. The production function in the X-sector is virtually identical to that described for the manufacturing sector in section D5.2.). There are N single firms, each producing one distinct (but symmetrical) intermediate under increasing returns to scale and within a monopolistically competitive market. The labour requirement \( \ell_i \) necessary to produce the quantity \( X_i \) is given by

\[
\ell_i = \alpha + \beta X_i \quad \text{with} \quad \alpha > 0, \quad \beta > 0
\]

Each firm sets prices as a constant mark-up over marginal costs\(^1\).

\[
p = \frac{\beta}{\theta} w \tag{E.4}
\]

Despite of the assumption of monopolistic competition, profits for every single intermediate good are driven down to zero by the entry of potential competitors. This again implies that all X-firms operate at a unique scale of output, and employ a well defined number of workers, respectively given by

---

\(^1\) In the formulation with \( \sigma \), which is also the perceived elasticity of demand, the pricing rule takes the form \( p(1-1/\sigma) = \beta w \). This form might look more familiar.
The determination of the equilibrium price and the equilibrium number of intermediate inputs can most easily been done for the case with full employment. We know that the equilibrium number of workers that each firm employs is given by \( (E.5) \). Let the size of the total labour force be given by \( L \). The equilibrium number of firms \( N \) is then simply given by

\[
N = \frac{L}{\ell} = \frac{L(1-\theta)}{\alpha}.
\]

(E.6)

This condition states that the equilibrium number of firms is higher, the larger is the local labour force \( L \). But recall that the number of firms \( N \) will also affect the production costs in the Y-sector. We have already established the equilibrium condition to be \( G = N^{(\theta - 1)/\theta} p = 1 \). Using \((E.4)\) and \((E.6)\) in this equilibrium condition, we can derive the equilibrium wage per worker \( (w) \) as a function of \( N \):

\[
w = \frac{\theta}{\beta} \left( \frac{(1-\theta)}{L} \right)^{\frac{1-\theta}{\theta}}.
\]

(E.7)

The individual wage is higher, the larger is the local labour force and the higher is the equilibrium number of firms \( N \). What is the intuition of this result? We have seen that production costs for the final product decrease in the number of available intermediate inputs. But the price \( p^Y \) is the numeraire and always equal to one. Suppose \( N \) increases. On instance, there will be positive profits in the Y sector, because costs have decreased at constant sales prices. With perfect competition, new entrepreneurs will enter the market for \( Y \) production and compete profits down to zero. This must be done by paying more for intermediate products. By \((E.4)\) and the assumption of zero profits in the X sector, these price increases will be absorbed by higher wages. In other words: If more intermediates can be produced, the increasing returns to scale can better be exploited. This will lead to higher wages.
b) The model with unemployment

Matusz (1996) now adds unemployment to the closed-economy model through a standard efficiency wage approach in the vein of Shapiro/Stiglitz (1984). His set-up is in principal the same as described in section C3.1a. We shall therefore present the essence of this model here only very briefly. We stick to the notation of chapter C and apply our version of a shirking model rather than the set-up of Matusz (1996).

Workers living in the closed economy are risk-neutral and derive utility from wage income $w$, but disutility from work-effort $e$, which is simply a technologically fixed number. The work utility is

$$ V = w - e. \quad (E.8) $$

"Shirking" individuals spend zero effort ($e=0$), but run a risk $(1-\gamma_e)$ of being detected and then fired. By using the unemployment rate $U = 1 - \frac{N \ell}{L}$, we can derive the utility levels of an unemployed individual ($V_u$), a non-shirking employed worker ($V_{en}$) and a shirker ($V_{es}$) respectively.

$$ V_u = \alpha (w - e) $$
$$ V_{en} = w - e $$
$$ V_{es} = \gamma w + (1-\gamma)(\alpha(w - e)), $$

where $\alpha$ is again the outflow probability from unemployment that negatively depends on the unemployment rate $U$ (see section C3.5.). The firms pay efficiency wages such that the "non-shirking condition" $V_{es} = V_{en}$ holds. The regional wage curve, or aggregate non-shirking condition, is given by the following expression

$$ w = e + \frac{\gamma e}{(1-\gamma)(1-\alpha(U))} \quad (E.9) $$

Of course the required efficiency wage is again lower the higher is the regional unemployment rate $U$. Equation (E.9) is the labour market equilibrium curve, and can be viewed as as the "first half" of full general equilibrium.

The product market equilibrium condition has also been derived above. It only needs to be slightly modified if we allow for unemployment. The maximum number of varieties $N$ that can be produced in this economy is no longer given by labour supply. Since labour supply and employment can now differ, it is by definition now given by the latter. The equilibrium condition (E.7) becomes
The two equilibrium relations (E.9) and (E.10) can be illustrated within the same graph, given in figure E1.

**Figure E1: Equilibrium in the closed economy**

The locus VV represents the familiar wage curve locus. It is the graphical illustration of all combinations of wages and unemployment rates where shirking is just prevented for workers in the intermediate good sector. For all points to the right of VV, unemployment is too high for any given wage. Consequently firms can hire new workers and trust that they do not shirk. Hence, equilibrium unemployment must fall. This determines the phase arrows in the horizontal direction. The locus BB represents all combinations of w and U where all goods markets (Y and all X) are in equilibrium: there are no profits for either firm, there is cost minimization in the Y-sector and profit maximization for the X-sector. The curve is downward-sloping and convex, because of the involved increasing returns in the technology.
For all points below BB, wages are too low for a given unemployment rate. Prof-
its arise in the Y-sector than attract new producers. Labour demand rises and con-
sequently wages have to go up.

Full equilibrium in this closed economy is given at an intersection of VV and BB. The phase arrows indicate that only one equilibrium is stable. This one is at point A, with equilibrium unemployment U* and equilibrium wages w*. Note that the equilibrium unemployment is strictly involuntary. Workers would in principle be willing to work at going wages w. But employers know that further recruiting would lead the incumbent workers to shirk at going wages. Therefore they do not hire additional workers. Put differently, the combination with full employment (U=0) and wages equal to \( \bar{w} \) might be desirable, but it is not feasible, since ra-
tional individuals would respond to this constellation with shirking behaviour.

Let us briefly look at some comparative statics. No parameter change affects both curves. The wage curve VV shifts with the disutility of effort e, or the shirking detection rate \( \gamma \). As argued in chapter C, these parameters might reflect structural characteristics of the respective labour market and could (through appropriate model extensions) be explicitly modelled as contingent on public policy, like e.g. the welfare state regime or employment laws.

The goods market curve BB shifts out to the right top as \( \alpha, \beta, \text{ or } \theta \) decrease. Lower values of \( \alpha \) or \( \beta \) reflect lower production costs at given output prices, i.e. an increase in profits. Due to perfect competition, this must lead to higher real wages stemming from the product market equilibrium condition. A decrease in \( \theta \) reflects a higher degree of increasing returns to scale. By a similar reasoning this advantage for the production sector must be absorbed by higher equilibrium wages.

c) The introduction of perfect trade

The final reason why the BB-locus could shift out to the top right is an increase in the exogenously given size of the labour force L. Since the VV-curve is not af-
fected by this parameter change, the resulting new equilibrium would be one with higher equilibrium wages w and a lower equilibrium unemployment rate U. The intuition is straightforward. An increase in L implies that it becomes feasible for the economy to produce a larger number of intermediate goods, which leads to a decline in production costs for one unit of Y. This intuition is also at the core of the analysis of Matusz (1996), who moves from a closed-economy setting to a two-country case with perfect (=costless) trade in final goods and intermediate products.

Suppose that the closed economy described in this section now starts to trade with a structurally identical foreign country. The final good Y as well as all intermedi-
ates X can be freely shipped across space, and all intermediates will be used symmetrically in the Y-production of either country. In such a setting, the home country can draw on imported inputs from the foreign country as if the intermedi-
Regional agglomeration and regional unemployment

ates were produces locally. In other words, there is no substantial difference between the introduction of trade and an increase in the size of the domestic labour force in the Matusz-model. The unit costs for the home country now read as

\[ G = \left( N(p)^{\theta/(\theta-1)} + N^*(p^*)^{\theta/(\theta-1)} \right)^{(\theta-1)/\theta} = 1, \quad (E.11) \]

where an asterisk denotes variables of the foreign country. An analogous equation like (E.11) applies to the foreign country. Since there is free tradability and perfect competition in the Y-sector, the equilibrium condition is still given by the requirement that unit costs \( G \) (and \( G^* \)) need to equal the numeraire price \( p^Y = 1 \). Due to free trade in the intermediates sector, there must also be price equalization for the symmetrical industrial inputs \( p = p^* \). The equilibrium condition for the product market in both countries is thus given by

\[ w = w^* = \frac{\theta}{\beta} (N + N^*)^{(1-\theta)/\theta} \quad (E.12) \]

If both countries face structurally identical labour markets, i.e. the same wage curve locus, (E.12) inevitably implies that there must be equalization of real wages and unemployment rates across the two countries. The real wage is higher and the unemployment rate is lower in either country than it has been the case in autarky. The basic implication of the model is thus that the introduction of free trade per se has positive welfare effects by increasing wages and lowering unemployment.

E3.2.) The two-region case with imperfect trade

As argued above, this basic version of the Matusz-model is ill-equipped to study regional disparities and agglomeration. We will therefore extend and generalize the model of Matusz (1996) in this section. In chapter D we have shown that there are basically three core elements of the regional agglomeration theories in the vein of NEG. The first, increasing returns to scale, is already a feature of the model approach by Matusz. The second, transportation costs, will now be introduced. It will open up the possibility of regional disparities, since now one can explicitly discriminate the economic variables of the different spatial units. The third core element, the presence of an endogenous mechanism that pushes for spatial concentration, will be introduced in the next section through (imperfect) labour mobility.
a) The product market

We now consider a nation consisting of two regions r and s with identical technology and preferences (if preferred, one can use the original terminology of Matusz with a home and a foreign country). Both regions produce the final consumption good Y under the use of symmetrical intermediate inputs which are partly manufactured in the respective region itself (X_{rr}), and partly imported from the other region (X_{sr}). Transportation of intermediates across regions now imposes \( 'iceberg' \) costs \( \tau > 1 \). For each unit \( X_{sr} \) dispatched, only \( 1/\tau \) units arrive. The standard CES production function for region r is now given by

\[
Y_r = \left( N_r X_r^\theta + N_s \left( \frac{X_{sr}}{\tau} \right)^\theta \right)^{1/\theta} \quad (E.13)
\]

The minimum unit cost function \( G_r \) becomes

\[
G_r = \left( N_r (p_r)^{\theta-1} + N_s (\tau p_s)^{\theta-1} \right)^{\theta/(\theta-1)}, \quad (E.14)
\]

and the demand functions for intermediates \( X_{rr} \) and \( X_{sr} \) can be written as

\[
X_{rr} = p_r^{\theta-1} G_r^{\theta/(\theta-1)} Y_r \quad \text{and} \quad X_{sr} = (\tau p_s)^{\theta-1} G_r^{1-\theta} Y_r. \quad (E.15)
\]

where \( p_r \) is the mill price of an symmetrical intermediate from region r. \( G_r \) is decreasing in both \( N_r \) and \( N_s \): A larger array of intermediate inputs reduces costs in the Y-sector in both regions. But due to the transportation costs, the decline is larger in the region where the increase in the number of varieties has occurred. Put differently, an increase in \( N_r \) has a stronger negative effect on \( G_r \) than on \( G_s \).

With respect to the final goods sector, we maintain the assumptions from the last section. The Y-sector is perfectly competitive, and the final consumption good can be freely traded across space. This implies that there is price equalization on the market for the final consumption good. Y-producers in both regions have to take the price \( p^Y \) as given, potentially as determined on world markets outside the nation. Without loss of generality we keep \( p^Y = 1 \) as the numeraire. At first sight the assumption of perfect tradability in the consumption goods sector, but imperfect tradability in the sector of industrial inputs might seem peculiar. But apart from offering analytical convenience, its use can be justified on intuitive grounds. Production in the Y-sector is just the assembly of intermediates without use of labour.
Therefore, the function $Y_r$ simply reflects with what level of technological sophistication the region $r$ can produce a given final output bundle. The advantage of sophisticated (large) regions is the high availability of locally produced intermediates, whereas small peripheral regions have higher costs of producing the same level of output.

Zero profits and efficient production in the $Y$-sector imply that unit costs need to equal one in both regions. Using (E.14), the respective equilibrium condition is

$$1 = N_r(p_r)_{\frac{\theta}{\theta-1}} + N_s(p_s)_{\frac{\theta}{\theta-1}}$$

for region $r$ and

$$1 = N_r(p_r)_{\frac{\theta}{\theta-1}} + N_s(p_s)_{\frac{\theta}{\theta-1}}$$

for region $s$. (E.16)

The production of the single intermediates is unchanged compared to the closed-economy case. Therefore, the number of locally produced intermediates is restricted by regional employment:

$$N_r = (1 - U_r) \frac{L_r}{\ell}$$

and

$$N_s = (1 - U_s) \frac{L_s}{\ell}$$

(E.17)

where $L_r$ and $L_s$ denote the exogenously given sizes of the regional labour forces. The regional wage levels $w_r$ and $w_s$ are determined in the same way as in (E.4), i.e. prices are constant mark-ups $\beta/\theta$ over marginal costs.

Contrary to the case with free tradability of intermediates, there is no longer price and wage equalization in the $X$-sector if the two regions differ in size and $\tau > 1$. If one region is larger than the other one, it has an advantage since it can produce more intermediates locally. We substitute (E.17) into (E.16) and set $\beta=\theta$ for notational simplicity. We obtain

$$w_r = \left[ \frac{(1 - U_r) \frac{L_r}{\ell}}{1 - (1 - U_s) \frac{L_s}{\ell} (\tau w_s)_{\frac{\theta}{\theta-1}}} \right]^{\frac{1-\theta}{\theta}}$$

(E.18)

The nominal (=real) regional wage $w_r$ derived from the condition for product market equilibrium is increasing in employment in both regions, but decreases
with higher wages in region \( s \) and transportation costs \( \tau \). An analogous equation applies to region \( s \). Solving for \( w_r \) and \( w_s \), we can obtain closed-form solutions for the regional equilibrium wages

\[
\begin{align*}
\frac{w_r}{w_s} &= \left( \frac{T}{1-U_r} \right)^\frac{1-\theta}{\theta} \\
\frac{w_s}{w_r} &= \left( \frac{T}{1-U_s} \right)^\frac{1-\theta}{\theta}
\end{align*}
\]

where \( T = \frac{1-\tau^{\theta-1}}{\theta} \geq 1 \)

At these wage levels, there is profit maximization and zero profits in both sectors and both regions. As can be seen, \( w_r \) and \( w_s \) only depend (positively) on employment in the respective region itself. An increase in \( L_s \) e.g. in equilibrium has only positive effects on the wage in region \( s \), but not on the wage of region \( r \). The analytical reason is the symmetrical use of all intermediates in both regions. It can be seen in (E.18): an increase in \( L_s \) lowers the denominator and ceteris paribus has positive spillover effects in region \( r \). But once the endogenous effect on \( w_s \) is taken into account, the impact will cancel out. Economically, equation (E.19) implies that the model incorporates a purely regional scale externality. Despite the openness, there are no interregional effects of an increase e.g. in the regional labour force \( L_r \). But within region \( r \), a larger labour force induces positive pecuniary spillovers on all workers in the form of higher equilibrium wages. The variable \( T \) can be understood as an inverse measure of the resource waste from shipping. If trade is prohibitively costly \( (\tau \rightarrow \infty) \), \( T \) takes on the value \( T=1 \) and (E.19) turns into the autarkic expression (E.10). The closer the economy comes to free trade \( (\tau \rightarrow 1) \), the value of \( T \) approaches \( T=2 \). The regional wages in (E.19) are consistent with efficient production, but we still have to show that the consumption good market and the market for intermediate products clear. Only then will the wages be the true equilibrium wages in this economy.

**Proposition E1:** Regional wages in (E.19) imply not only efficient production, but also market clearing in both sectors and regions. They are the true equilibrium values.

**Proof:** Each firm in the intermediates-sector supplies an regionally invariant quantity given by \( X=\alpha/(1-\theta) \) (Equation (E.5) with \( \beta=\theta \)). Equilibrium requires that \( X \)
equals total sales to both regions $X_{rr} + \tau X_{rs}$. Using (E.15) together with the equilibrium condition $G_r=1$ we can write this as

$$X = (w_r)^{\frac{1}{\theta-1}} \left( Y_r + \frac{\theta}{\tau^{\theta-1}} Y_s \right) \tag{E.20}$$

Using (E.19) and solving for $Y_r$ yields

$$Y_r = \frac{X}{2\theta} \left( (TN_r)^{\frac{1}{\theta}} - \tau^{\frac{\theta}{\theta-1}} (TN_s)^{\frac{1}{\theta}} \right) \tag{E.21}$$

Equation (E.21) determines the regional production level $Y_r$ at which markets for intermediates clear. The total national production of $Y$ is

$$Y_r + Y_s = \frac{X}{T} \left( (TN_r)^{\frac{1}{\theta}} + (TN_s)^{\frac{1}{\theta}} \right) \tag{E.22}$$

Since $Y$ is freely tradable at $p^Y=1$, (E.22) needs to equal total national income and consumption expenditure, which is given by

$$w_r(1-U_r)L_r + w_s(1-U_s)L_s \tag{E.23}$$

By using (E.19) in (E.23) and rewriting (E.22), one can show that both expressions are equivalent to

$$\left( \frac{T}{X} \right)^{\frac{1-\theta}{\theta}} \left[ ((1-U_r)L_r)^{\frac{1}{\theta}} + ((1-U_s)L_s)^{\frac{1}{\theta}} \right], \tag{E.24}$$

which proofs proposition E1. Markets for $Y$ and all intermediates $X$ clear with wages and prices given by (E.19). Equation (E.24) is the gross national product of this two-region economy: it is increasing in employment, and decreasing in transportation costs.

Let us consider the intuition of the expression (E.19). It is telling that for identical unemployment rates $U_r = U_s = U_1$, the region with the larger labour force pays the higher wage. Graphically, this can be seen in figure E2). The two downward sloping curves $B_rB_r$ and $B_sB_s$ illustrate those combinations of unemployment rates and wages consistent with product market equilibrium in both regions. The labour force in region $r$ is assumed to be larger than in region $s$. 

Regional agglomeration and regional unemployment

Figure E2: Product market equilibrium in both regions

The intuition for this scale effect is straightforward. At any given unemployment rate, region r produces more local intermediate goods than region s. Since all intermediates from both regions are used symmetrically in the Y-production, the larger region saves on transportation costs. It must consequently pay higher wages for the zero profit conditions in the Y- and the X-sector to hold. The higher is the difference between \( L_r \) and \( L_s \), the further apart are \( B_r \) and \( B_s \). An increase in \( \tau \) shifts both curves downwards and to the left, because the dead weight loss of resources wasted in transportation increases. The same shift occurs upon an increase in \( \alpha \), \( \beta \), or \( \theta \).

b) The labour market

We have explicitly described the product market in this economy and derived a negatively sloped curve for each region that represents the equilibrium relation of wages and unemployment consistent with efficient production and market clearing. To close the model, the regional unemployment rates \( U_r \) and \( U_s \) need to be determined. This will of course be a matter of the wage curve that constitutes the second equilibrium relation in this model.

The derivation of the wage curve remains unchanged compared to the closed economy setting. The wage curve in both regions r and s is thus given by (E.9). There are no structural differences between the labour market in region r and region s. Both regions face the same downward sloping labour market equilibrium curve in the \((w,U)\)-space that already showed up as the VV-locus in figure E1.
c) Equilibrium in the two-region economy with trade and immobile agents

Full equilibrium is obtained if both product and labour markets are jointly in equilibrium. Both regions face the same wage curve, represented by the curve VV in figure E3). The two BB-loci are the product market equilibrium curves from (E.19) that were graphically represented in fig. E2. The stable regional equilibrium points are at A and B respectively. As can be seen, region r has both the higher equilibrium wage and a lower unemployment rate.

Figure E3: Equilibrium in the two-region economy with immobile agents

Recall that we have assumed that region r is larger than region s, and can therefore better exploit the scale economies entailed in the adopted technology. For any given wage rate, it has a higher labour demand. This drives down unemployment at first instance and simultaneously increases the necessity to pay efficiency wages in order to prevent individuals from shirking. In equilibrium, the larger region ("the core") is advantaged over the smaller one along two dimensions: by having higher wages and a lower unemployment rate.

The fact that larger regions pay higher wages in presence of increasing returns in combination with transportation costs has already been noted in the NTT model of Krugman (1980) that has been discussed in the last chapter. Yet, the existence of unemployment exacerbates this agglomeration wage premium. It is not only because of technological factors and the better exploitation of scale economies. With efficiency wage based unemployment, we can see that the larger region must pay an additional wage premium to deter shirking in view of the low unemployment rate.
E3.3.) The impact of labour mobility

So far he have taken the sizes of the regional labour forces as exogenously given, which is equivalent to assuming that there is no labour mobility. But spatial differences as indicated in fig. E3) will induce workers from region s to migrate to region r. This conclusion seems less clear for Europe than for the US, since geographical mobility in Europe is much lower than in the US e.g. because of cultural and language barriers (Puhani, 1999). But nevertheless, for an analysis on the basis of regions, factor mobility must be taken into account. Furthermore, also in Europe there is some degree of responsiveness of workers to interregional wage and unemployment differentials (Decressin/Fatas, 1995; Pissarides/McMaster, 1990). A more general discussion about the causes and consequences of interregional labour migration is presented in the next chapter.

a) Perfect labour mobility

Labour migration is not an equilibrating force in our model framework, it rather perpetuates and accentuates regional disparities. In fact, if workers were perfectly mobile across regions, there would always be full concentration of all economic activity in one region regardless of the level of transportation costs.

For notational convenience, let us normalize the size of the total national labour force to \( \tilde{L} = L_r + L_s \). A fraction \( \lambda \) lives and works in region r, \( L_r = \lambda \tilde{L} \). The remaining fraction \( (1-\lambda) \tilde{L} \) is located in region s. From (E.19), the relative regional wage \( \hat{w} = w_r/w_s \) is given by

\[
\hat{w} = \frac{w_r}{w_s} = \left( \frac{(1-U_r)\lambda}{(1-U_s)(1-\lambda)} \right)^{1-\theta}
\]

As shown in figure E4), the relative wage \( \hat{w} \) is greater than one if \( \lambda > \frac{1}{2} \). This is so for two reasons. Firstly, the better exploitation of increasing returns warrants a higher production wage. And secondly, unemployment is lower in region r. Thus, higher efficiency wages need to be paid. Suppose that we are initially in a symmetrical situation with \( \lambda=1/2 \). Initially, there will thus be identical wage levels and unemployment rates across regions.

But workers have an unambiguous incentive to concentrate in one region, say region r, since the relative real wage rate is steadily increasing in the degree of spatial concentration. All workers would thus migrate up to the point with \( \lambda=1 \) under perfect mobility. This concentration process is also socially optimal, as can be seen from the expression of total national output (E.22), or gross national income (E.23) in this economy. Both expressions, which have been proofed to be equivalent to (E.24), reveal larger values under an asymmetric distribution of the total labour force across space than under a symmetrical one with \( L_r=L_s \).
The relative wage function is non-linear and strictly upward sloping for all values of $\tau > 1$ and $0 < \theta < 1$. In other words, there is no break on the feasibility of regional agglomeration in the model with perfect mobility; there is no "centrifugal" force. We will come back to the discussion of this issue below, when we critically assess our approach and compare it with a "typical" NEG-model.

b) The degree of agglomeration under imperfect labour mobility

For now, we will introduce an "ad-hoc" centrifugal force by assuming imperfect labour mobility. Two reasons for this construction can be put forward at this point: Firstly, the assumption of perfect mobility is neither appealing per se for most countries, nor is the result of complete agglomeration very reasonable. And secondly, recall from chapter C that the collapse of the wage curve relation in the model of B/O (1996) is only prevented by the presence of "ad-hoc" intrinsic regional preferences. By adopting a comparable set-up in our approach, we will be able to highlight in the clearest possible way in what respects our approach differs from that of B/O.

We add an individual specific discount factor $\kappa_i > 0$ that captures the relative attractiveness of region r perceived by individual i. The value $\kappa_i$ is equal to one if an individual has no intrinsic preference for either region. Values $\kappa_i > 1$ indicate preference for region r. Vice versa, $\kappa_i < 1$ indicates a discount for living in region r. We assume that in the initial situation that half of the total labour force that is located in region r also has an intrinsic preference for this location (i.e. $\kappa_i > 1$). The other half of $\tilde{L}$, which is located in region s, instead dislikes region r.
Consider all \( \tilde{L} \) workers in the nation rowed up in a line from \( i=0 \) up to \( i=\tilde{L} \). We assume that the distribution of the individual preference parameter \( \kappa_i \) over the whole working population will be described by the following function

\[
\kappa_i = \left( \frac{\tilde{L} - i}{\tilde{L}} \right)^d \quad \text{with } d \geq 0. \tag{E.26}
\]

The worker \( i=\tilde{L}/2 \) is the only one who has no intrinsic preferences for either region (\( \kappa_i=1 \)). All workers in the range from \( \lambda=0 \) to \( \lambda=1/2 \), i.e. those who are originally located in region \( r \), also have a preference for that location. Since these workers are indexed in the range \( i \in (0; \tilde{L}/2) \), their value of \( \kappa_i \) is greater than one. On the contrary, all workers indexed in the range \( i \in (\tilde{L}/2; \tilde{L}) \), i.e. those who initially live in region \( s \), discount to live in region \( r \). The individuals at the extremes, \( i=0 \) and \( i=\tilde{L} \), have values \( \kappa_i=\infty \) and \( \kappa_i=0 \) respectively, and are thus inevitably tied to one specific region. The parameter \( d \) is a measure of how strong the overall labour force is biased towards living in the home location. The lower is \( d \), the more pronounced are the preferences for one specific region. In the extreme case with \( d=0 \), all individuals are intrinsically indifferent between locations.\(^2\)

Let us consider the workers living in region \( s \), who face the following trade-off:\(^3\) they would receive an "economic" benefit if they were to move to region \( r \), but they are intrinsically attached to their home location \( s \). Thus, these workers must receive a premium in order to actually move. They move if and only if the relative wage of region \( r \) is greater than the individual dislike of that location. The analytical condition for a move of individual \( i \) to occur is \( \kappa_i \hat{\omega} > 1 \), or

\[
\hat{\omega} > \frac{1}{\kappa_i}.
\]

\(^2\) Note that the modelling of preferences \( \kappa_i \) only captures intrinsic preferences for the region \( r \) that are independent of the actual degree of spatial concentration in that location. The term \( \kappa_i \) does not represent a dislike of congestion etc.. However, given that the construction of regional preferences is "ad-hoc" anyway, one could easily extend the preference formation and introduce a term \( \kappa(\lambda) \) that depends on the overall degree of spatial concentration. Total individual preferences might e.g. look like \( \kappa(i,\lambda) = \kappa_i + \kappa(\lambda) \), where the former term captures purely individual factors and the latter the preference for or against congestion.

\(^3\) Again, the analogous case where region \( s \) is the agglomeration area also applies. But we consider here the migration direction from \( s \) to \( r \).
Regional agglomeration and regional unemployment

To obtain a better intuitive feel for this trade-off, it is graphically illustrated in figure E5. The thick solid line represents the relative regional wage $\hat{w}$ that was introduced in figure E4. The dotted lines represent three different examples of the inverted regional preferences $1/\kappa_i$. The steeper are the $1/\kappa_i$-curves, the more biased are individuals towards their home region, i.e. the lower is the parameter $d$. People move only if the curve $\hat{w}$ runs above curve $1/\kappa_i$. For the case with high regional preferences (the steepest curve in fig. E5), the final equilibrium will be the symmetrical initial situation, because the higher relative wage (the technological advantages from concentration) can never compensate for the individual dislike of region $r$. The opposite happens with low attachment to specific regions, the flattest dotted line in fig. E5). Here, agglomeration will be almost complete, as the location of workers is driven mainly by regional relative wages. With intermediate regional preferences, the equilibrium distribution of the workforce $\lambda^*$ is given by the intersection point of the two curves.

Figure E5: Distribution of the labour force with regional preferences

The regional preferences do thus not change the fundamental result that the larger region is better off than the periphery with respect to both $w$ and $U$. The existence of $\kappa_i$ introduces parameter constellations, in which migration does not occur due to very pronounced attachment to home locations. Importantly, the construction with regional preferences allows for less-than-full agglomeration. There are interior equilibria $\frac{1}{2} < \lambda^* < 1$, where wages and regional preferences just compensate for each other. This enables us to draw implications from the degree of agglomeration ($\lambda$) for the magnitude of regional disparities.
Proposition E2: Regional disparities are more pronounced the higher is $\lambda^*$ in the range $\lambda \in [1/2; 1]$. $\lambda^*$ is larger, i) the stronger are the increasing returns to scale $\theta$; ii) the less important is the overall home bias $d$.

We have shown that regional disparities are highest, the closer $\lambda$ is to the boundary case $\lambda=1$. Regional disparities are nil if $\lambda=1/2$. In between the two extremes the relation between $\lambda^*$ and the magnitude of regional disparities is smooth.

E4) Critical discussion of our model approach
The model presented in section E3 aimed to integrate a wage curve into a general equilibrium framework where the technology allows for agglomeration forces similar to the literature in NTT of NEG. The partial equilibrium foundation of the wage curve remained unchanged compared to existing work. But the product market structure in our model has been changed. We have formulated a generalized and extended version of the NTT model of Matusz (1996). Our framework allows us to explicitly take up the issue of regional unemployment disparities in relation to regional agglomeration. In this section, we want to first summarize the crucial mechanisms and results of our model. We then compare our approach with the core models of wage curve theory and NTT/NEG. This discussion, that will highlight in what respects our approach improves on the current state of art in the literature, will be followed by an assessment of the empirical implications of our model. These shall be contrasted with the available empirical evidence from chapter A. Finally we will highlight some further issues that motivate the contents of the next chapters.

E4.1.) Discussion of the model from a theoretical point of view
The central mechanisms in our model can be summarized as follows: Larger regions are advantaged over smaller regions, because they can better take advantage of the adopted technology with localised increasing returns to scale. This advantageous scale effect leads to higher regional wages in the larger region. This result has been known since Krugman (1980). If unemployment is explicitly introduced through a wage curve, one can show that the large region is advantaged also through a lower unemployment rate, and it needs to pay an additional efficiency wage premium. This insight can be seen as the fundamental contribution of the model approach from section E3.

Stated differently, it implies that unemployment rates will exhibit the same core-periphery structure as wages and output. Regional unemployment disparities are driven by the degree of regional agglomeration, as large ("core") regions exhibit low unemployment rates and high wages, whereas small (peripheral) regions have high unemployment rates and low wages.
a) Comparison with the wage curve theory

It is now worth discussing how our approach compares to the basic wage curve models from chapter C. We have named some fundamental criticisms with respect to the models of B/O (1996) and Blien (2001), which we took as the motivation to formulate an own approach. These critical points were:

- Regional disparities in the two model can not develop endogenously, but are due to exogenous assumptions, e.g. on sectoral specialization patterns.

- Regions were essentially identified with industries, and regional disparities were implicitly explained solely by sectoral factors.

- The existing models do not take into account regional agglomeration, which is one of the most salient features of economic landscapes and should thus not be neglected in regional approaches.

- Labour mobility leads to an erosion of the wage curve, and the foundation of the wage curve as a long-run equilibrium relation must rely on restrictive ad-hoc assumptions.

Our model approach leads to fundamentally different conclusions with respect to all these four points. In our model with a technology incorporating increasing returns to scale, regional asymmetries can develop completely endogenously. There is an endogenous tendency towards regional divergence transmitted through the market mechanisms alone. Furthermore, sectoral specialization patterns play no critical role in our model. In our two-region model, both locations are engaged in production activities within the same sector. Differences in the production structure exist only insofar that the larger region can produce a higher number of industrial intermediates. But all regional differentiation, and all interregional trade is of an intra-industry type. Through the working with localised scale economies, our model is able to account for forces that drive an economy towards spatial agglomeration. We have taken one of the classical arguments from Marshall (1890) that were discussed in chapter D (market linkages), and incorporated it explicitly in the analysis. It has been argued that such an approach in general seems much more plausible on a regional level than a conventional approach with constant returns to scale and perfect competition.

What about labour mobility? In Blien (2001), it is explicitly said that labour migration leads to the gradual erosion of the wage curve, i.e. to a gradual equilibration of regional wage and unemployment levels. For the model of B/O (1996), principally the same result holds. They assume the existence of regional disparities through an assignment of single regions to the production of specific distinct goods. The region that produces the “better” good ultimately has the higher equi-
librium wage rate and the lower unemployment rate. Workers from the region that produces the “bad” commodity therefore have an incentive to emigrate. But B/O assume that intrinsic regional preferences prevent migration from one region to the other, because individuals in the region with the unfavourable economic values are compensated by regional amenities, e.g. the lack of congestion.

In our model, we also have adopted “ad-hoc” regional preferences, which per se seem quite plausible. In one way, the construction is thus quite similar in spirit to B/O: The existence of intrinsic preferences slows down or even excludes migration in response to interregional wage and unemployment disparities. However, their role in our model is still fundamentally different. In B/O, the lack of migration prevents the erosion of the wage curve, or the convergence of regional wage levels and unemployment rates. In our model, the opposite is true. Here, the sluggish migration prevents further regional divergence. In absence of intrinsic regional preferences, the B/O-model would imply total convergence of regions, whereas our model (see section E3.3b) would imply complete concentration of all economic activity in only one location, i.e. “complete divergence”.

This conclusion becomes most obvious when comparing the two central diagrams of the B/O-model (fig. C2) and of our approach (fig. E3). As can be seen, the wage curve locus in both diagrams is identical, and was similarly rationalized on the basis of the shirking approach of efficiency wage theory (Shapiro/Stiglitz, 1984). The product market equilibrium curves in the B/O-model under constant returns to scale are just horizontal lines on an exogenously given level in the wage/unemployment-space. It has been shown in chapter C that this shape is the outcome of a limitational production function under constant returns to scale. The product market equilibrium curves in our model are instead downward sloping and non-linear. This reflects the scale effect that is incorporated in our technology. The lower is unemployment, the better can the scale economies be exploited and the higher is the equilibrium wage rate that is consistent with product market equilibrium. The figures C2 and E3 are constructed for given sizes of the regional labour forces. Introducing labour mobility would shift the two curves in fig.C2 closer together as workers from region 2 migrate to region 1. In fig. E3 on the other hand, the two curves would shift further apart. In other words, in the B/O-model the wage curve-type relation has an inherent tendency of vanishing. The implication of our model is fundamentally different: The wage curve is a stable relation that is even strengthened through labour mobility.

---

4 If a straight neo-classical production function with diminishing marginal returns would be assumed, the respective product market equilibrium curve would be upward sloping in the diagram. With low unemployment, the marginal productivity (=wages) falls. The higher is unemployment, the scarcer is labour and the higher is the MPL and the wage. For a further discussion see section B3.3.), where this case has been discussed in the vein of the PS curve of the ELM.
Hence, if one works with an increasing returns technology, the theoretical case for the existence of a wage curve is even stronger than it has been argued by B/O themselves. They are preoccupied to view the wage curve as a stable equilibrium relation over time. With their constant returns to scale technology, they can only rationalize this perception with the “ad-hoc” construction of intrinsic preferences. The use of an increasing returns to scale-technology, however, quite naturally implies the long-run stability of the wage curve as a basic outcome. The “ad hoc” preferences in our model put boundaries on the occurrence of regional divergence, but their use is not critically needed to prevent the erosion of the wage curve. All in all, the use of an increasing returns considerably improves the general equilibrium wage curve theory. Mechanisms for the emergence of regional disparities (which after all are of central interest also to the wage curve literature) develop endogenously and do not need to be exogenously assumed. Agglomeration is explicitly accounted for, and the wage curve is truly a long-run equilibrium relation due only to deep technological factors.

b) Comparison with NEG-models
How does our model relate to the NTT- and NEG-literature? Our approach was build on one specific NTT-model by Matusz (1996). We have argued that this framework is useful, because it combines a labour market equilibrium curve that can be interpreted as a wage curve with a production technology exhibiting increasing returns to scale.

We have extended the Matusz-model and introduced transportation costs and labour mobility. Thereby the analysis was brought closer in spirit to the NEG-literature. Nevertheless, the approach from section E3 differs in some important respects from a “typical” NEG-model. Recall from the discussion in chapter D that the core question of NEG is how the landscape of an economic area shapes as a function of opposing centripetal and centrifugal forces. Usually, the relative strength of either force is determined by exogenous parameters like the level of transportation costs.

This is different in our approach. Technologically there are only centripetal forces. With perfectly mobile labour, there would be complete agglomeration in order to fully exploit the increasing returns in production. This agglomeration would be socially optimal as it maximizes total national income and minimizes national unemployment. As a quasi-centrifugal force we have added intrinsic regional preferences (an “attachment to home” parameter) that led to imperfectly mobile labour. In the terminology of section D4, this type of mechanism should be labelled a “status-quo bias” rather than a “pure” centrifugal force.

Therefore, one can criticise our approach as not offering a fully endogenous explanation for the emergence of a geographical equilibrium structure. Our first reaction to this reproach would be to admit that our model is indeed not a “typical”
NEG-model. But we feel that there are good reasons for having taken our simpler approach. We like to comment now on these reasons.

We have argued in the introduction of this chapter that the motivation of our own model approach is twofold. The first objective was to formulate a general equilibrium model of the wage curve with a technology that allows for agglomeration forces. As argued in E4.1a), we feel that this objective has been achieved. We have formulated our model such that the analogy with the approach of B/O (1996) becomes as clear as possible, which enabled us to point also as clear as possible to differences and common features. As the second objective we have formulated that unemployment should be integrated as an element to the new regional agglomeration theories. This class of models is wider than just “typical” NEG-models in spirit of Krugman (1991a) or Venables (1996). Given that our model from section E3 also explicitly accounts for the role of space and for regional agglomeration, it surely belongs to this broad class of models in the literature, without being a NEG-model in the strictest meaning of the word.

But why did we not use a “typical” NEG-model? This is mostly due to the intractability of these models. Recall from chapter D that NEG, even without the analysis of unemployment, usually has to rely on numerical solution techniques. With unemployment added, these problems will be exacerbated, and instructive and intuitive analytical solutions will be even more out of reach. After all, the important question is whether working with a “typical” NEG-model would have offered additional insights that are not available with our relatively more simple approach. To assess this question, it is important to again consider the basic mechanisms e.g. of the seminal Krugman-model. The centripetal force in that model is structurally very similar to our approach. The main difference is that Krugman’s model entails an endogenous centrifugal force. This stem from the fact that he assumes local immobility of the agricultural workers. In our model, the centrifugal tendency is absent because we do not assume \textit{prima facie} immobility of parts of the workforce. This immobility of consumption demand in combination with transportation costs drives the results of the Krugman-model and leads to the non-linear interplay between agglomeration and dispersion forces. But recall that in the long-run equilibrium, the economy is either operating in complete symmetry, or parameters are such that a process of regional divergence is triggered at the end of which the whole mobile labour force is concentrated in one region.

We have abstracted from the particular centrifugal force of the Krugman model, because our main interest is not to derive the equilibrium geographical structure of the economy as a function of parameters, but rather to draw implications from the degree of agglomeration for the magnitude of regional unemployment and wage

---

5 The only (not crucially important) difference here is that the endogenous linkage effects in the Krugman model are between the firm and the consumption sector, whereas in our model they accrue on the production side alone.
disparities. If the underlying model can only predict either complete symmetry or complete asymmetry, the analysis of regional disparities becomes somehow meaningless. With our set-up, we can specifically look at the impact of the variable $\lambda$ (that reflects the degree of spatial concentration) on differences between region $r$ and region $s$. With the Krugman-model, such an analysis of gradual differences would not have been possible.

There are of course NEG-models that also allow for imperfect labour mobility and can thus predict intermediate degrees of agglomeration (e.g. Ludema/Wooton, 2000). In their model, there are parameter ranges where complete symmetry prevails. But then eventually, spatial asymmetries emerge as transport costs fall. But the change in the economic landscape is smooth and intermediate equilibria are feasible. The model of Ludema/Wooton would thus be applicable also for our purposes, since one could try to introduce unemployment in their model and see what the degree of agglomeration implies for spatial unemployment differences. But one has to acknowledge that using their (much more complicated) set-up would not add significant new insights compared to our approach. As long as transport costs are such that the economy operates under symmetry, by definition there can be no regional unemployment disparities. Once centripetal forces prevail, the model of Ludema/Wooton would likewise predict that the degree of spatial unemployment disparities is larger, the higher is the concentration of the increasing returns sector in one region. In our simpler approach, there is no parameter range of transportation costs for which there is symmetry. But for the “interesting” case with regional disparities, our model qualitatively behaves in the same way as a more complicated NEG-model would do.

Hence, we feel confident that our model set-up, where technologically only centripetal tendencies exist and quasi-centrifugal forces are entered “ad-hoc”, is sufficient to make our basic point that regional unemployment rates depend (negatively) on regional agglomeration. We now want to turn to a discussion of our theoretical result in view of the available empirical evidence.

**E4.2.) Discussion of the model from an empirical point of view**

In section A3.1.) we have shown that regional unemployment rates in the European Union (EU-15) closely resemble the core-periphery-structure of regional GDP per capita. Low unemployment is centred around the “European Banana”, medium income regions usually belong to the group with intermediate unemployment rates, whereas the poor “objective 1”-regions mostly have very high unemployment rates. We have also shown that the relation between regional GDP per capita and the regional unemployment rate is not “one-by-one”. But all in all,
the spatial distribution of unemployment rates closely follows the spatial agglomeration pattern.

a) The theoretical model and the stylised empirical facts from the EU

The theoretical model from section E3) is consistent with this observed stylised fact. The model predicts that unemployment is low in regions with high real wages (= real income levels) and high output levels (equation (E.21)). In the model, scale effects play a crucial role. The output and income level of a core region is high, because a large number of workers is concentrated in this location. The evidence presented in section A4.2) is also consistent with this theoretical result. The rich regions from the “European Banana” with low unemployment rates are also characterised by a high population density, which is a good proxy for saying that these regions are large. The population density (the “size”) in “objective 1”-regions is considerably lower. Low unemployment and high income levels in the EU-15 thus coincide with a high spatial concentration of population (and workers). This stylised fact is broadly consistent with the prevalence of technological scale effects.

Empirical support for the negative effect of economic agglomeration on the unemployment rate is provided by Blien et. al. (2002). The authors analyse the determinants of employment growth across East German regions for the time period 1993-2000. The partial effect of agglomeration is found to be significantly positive for explaining local employment growth. In view of the strong correlation between employment growth and falling unemployment rates, this empirical finding can also be interpreted such that economic agglomeration in any region has a negative impact on the local unemployment rate.

The next stylised fact we have presented in section A4.2) is that “objective 1”-regions have tended to loose population to the rich core regions through migration. Our model implies that this will lead to further regional divergence. And again, the evidence presented in section A2.2) is at least not inconsistent with this claim. There has been regional divergence on a regional level in the second half of the 1990s, and the regions that faced emigration during this period (e.g. Southern Italy) also tended to fall back with respect to relative income figures.

7 Of course one has to use population densities rather than population sizes in absolute terms, because the surface of NUTS2-regions is utterly unequal. If the territorial size of a region is not taken into account, one can come to quite misleading conclusions. E.g. Andalucia (ES) is one of Europe’s largest NUTS2-regions, which in the vein of our model would imply that the regional unemployment rate should be quite low. Andalucia, however, has one of Europe’s highest unemployment rates (about 26%). Yet, Andalucia is also one of Europe’s largest regions in terms of surface, and a view at map A6 indicates that the population density of Andalucia ranges somewhat below the European average. Hence, our theoretical results look much more reasonable again if a meaningful measure of the “size” of the respective regional labour force is used.
When looking at the correlation between regional agglomeration and regional unemployment rates, there are of course single counter-examples that contradict our general conclusion: Portuguese regions e.g. tend to be small and sparsely populated, GDP per capita and particularly GDP per employed worker is low. Still, the unemployment rates in these regions are also below the European average. However, Portugal constitutes an exception to a rule. All in all, this single case is insufficient to falsify the general spatial pattern that emerges from the maps in chapter A, which can be reasonably well explained by our theoretical model.

Based on these considerations, we conclude that the approach presented in this chapter is well consistent with the empirical evidence on the regional dimension of economic activity in the EU-15. It is therefore not only a contribution from a theoretical point of view, as it combines and extends previously unrelated strings of literature. The model also yields empirically relevant conclusions.

b) Intra- versus international disparities and the role of institutions

It is important to note that regional unemployment rates differ in our theoretical model, even though there are no differences in the underlying structural or institutional characteristics of the regional labour markets. We have argued above that the two parameters \( e \) and \( \gamma \) in principle can be thought of as reflecting structural characteristics of the labour markets in the two locations. But purposely we have assumed that \( e \) and \( \gamma \) do not differ, i.e. we have assumed that both regions face the same labour market institutions. Still, unemployment disparities can arise. But they are solely driven by other factors, namely the agglomeration economies that drive the location of labour demand, which in a secondary step determines the geographical configuration of unemployment.

In other words, our model neglects institutional differences as an explanation for spatial unemployment disparities. In this respect, our approach offers a perspective about the phenomenon of unemployment that is quite different from the conventional viewpoint in economic research, which usually attributes the observed cross-sectional variation of unemployment rates across different administrative units to the impact of institutional arrangements, most notably to various forms of labour market rigidities (e.g. Nickell, 1997). This conventional type of analysis is widely known. Put bluntly, it views the unemployment rate of any given jurisdiction as a measure of the malfunctioning of the labour market, somehow in the vein of Friedman’s natural rate of unemployment (see section B2.3.). Arguments in this spirit are normally concerned with unemployment in its national dimension. There is of course some degree of institutional heterogeneity across the single national labour markets. E.g., the Danish and the Dutch labour market institutions are quite different from the French and German ones. Typically, these institutional differences are argued to explain fairly well the differences between the national unemployment rates (Nickell, 1997).
However, when looking at the evidence presented in chapter A, one can have serious doubts whether it is really useful to predominantly think about the phenomenon of unemployment along national boarders. We have shown in quite some detail that dramatic spatial unemployment disparities exist within almost each EU member country. On an intra-national basis, however, institutional differences are almost negligible. In Germany e.g., the institutional rigidities that are frequently argued to have a significant adverse impact on the performance of the labour market apply to the entire nation without any regional differentiation. This concerns labour laws, the tax regime, the welfare state arrangements, the high degree of union power etc. The same set of institutions can bring about utterly different unemployment rate outcomes on a regional level, ranging from around 3.5 per cent in Oberbayern to roughly 20 per cent in Halle. The case of East Germany might be subject to very special influence factors due to the particular historical circumstances of division and reunification. But think of Italy or Spain. The “classical” labour market rigidities are also valid nationwide. Yet, unemployment differences of more than 20 percentage points exist and show no tendency to vanish. In practically all EU countries, except for the Netherlands, regional unemployment disparities are present, which can not be attributed to differences in labour market institutions across the single regions. Moreover, we have argued in chapter A – following Overman/Puga (2002) – that there exists a more meaningful grouping scheme for regional unemployment rates other than the pure assignment of regions to different member countries. There exist trans-national unemployment clusters that do not obey to national boarders. The spatial configuration of unemployment rates across the EU as a whole was shown to follow a quite distinct spatial pattern (European banana, intermediates, objective 1), in which national boarders have played a minor role. What do these empirical facts imply for the role of labour market institutions on unemployment? First of all it is clear that it is far too simple to interpret a regional unemployment rate as being determined at any point in time by the prevailing institutional characteristics of the local labour market. It is inappropriate to work with a simple function

\[ U_r = f(V_r) \]  

(E.27)

that describes the unemployment rate of region \( r \) only as a function of institutional variables \( V_r \). With a relation like (E.27), there could be no intra-national unemployment disparities, since there is no notable intra-national variation in labour market institutions.

---

8 Regional institutional differences in Germany exist at best with respect to the impact of collective bargaining, because East Germany is covered to a far lesser extent by union minimum wages. But unemployment in the East is far above Western levels.
But it would be precipitate to conclude that the presence of intra-national unemployment disparities implies that institutions are entirely irrelevant. This point can be illustrated by going back to the theoretical analysis from section E3. Let us consider two cases. Firstly suppose that the two regions r and s belong to the same nation, in which case the two institutional parameters e and γ (by definition) are identical in both regions. And secondly, assume that the two regions belong to different nations, and the parameters can be different in r and s. The first case has been analysed in section E3. It was shown that the assumption of no variation in the parameters e and γ implies that both regions face the same wage curve locus VV. Full equilibrium in figure E3 is obtained at the intersection points with the respective product market equilibrium curve BB. Due to the agglomeration economies, it is possible that spatial unemployment and wage disparities open up. They are entirely driven by the two BB-curves.

In the second case with institutional variation, the wage curve loci VV do not necessarily have to be identical in the two locations. Suppose that the work effort parameter e is higher in region r than in region s (e_r > e_s). Recall that the parameter e reflects the perceived disutility from working, and it can be understood as a real reservation wage. Maybe this effective wage floor is higher in region r because of a more generous welfare state regime. What are the consequences of this institutional difference? The wage curves of region r and s now run strictly parallel, and the curve for region r is located further apart from the origin. With the BB curves as given in figure E3, we might reach a final equilibrium where the unemployment rate in region r is lower (and wages higher) than in region s. But compared to the benchmark case with no institutional variation, the spatial disparities are now smaller.

In other words, institutions do matter in our approach insofar as they determine the labour market equilibrium locus VV, which is “one half” of the full equilibrium. But institutions are not the only influence factor for determining the regional unemployment rates. The second half, that is emphasised in our approach and (as we see it) often neglected in the literature, is the influence from the product market, with the agglomeration economies and the location of labour demand.

With these considerations in mind, we can more specifically demonstrate the empirical relevance of our model. In the version with no institutional variation, the model is able to explain why regional unemployment rates can be so utterly different within the same nation, notwithstanding the prevalence of the same institutions across regions. But the model is relevant also in a wider sense. The main motivation for our theoretical approach was to explain the geographical structure of unemployment in the EU as a whole, i.e. also across the single nations. The geography of unemployment closely resembles the overall core-periphery structure of economic activity. We have argued above that our model predicts precisely this resemblance, and is thus well suited to explain the regional dimension of unemployment in the EU. But we have also mentioned a few outliers. For example,
unemployment rates are very low in Portugal even though all Portuguese regions are located in the economic periphery. Supposedly this is due to the particular labour market institutions that render a wage curve in the Portuguese regions that is located closer to the origin as in other European countries. In other words, it might be the case that the VV curves are different in the single European regions, depending on national affiliation. The regional unemployment rate and the real wage level in equilibrium are determined jointly by the wage curve and the product market equilibrium curve, i.e. by institutions and agglomeration together. Institutional rigidities are by no means irrelevant for the determination of unemployment rates. But all in all, the influence of international differences in labour market institutions on unemployment can not be all to large. The geographical structure of unemployment on the basis of regions in the EU corresponds very little to national boarders. If institutional differences were very relevant, country boarders should be much more visible in a map like A3. Since this is not the case, we conclude that the spatial configuration of unemployment is mainly driven by product market agglomeration.

E5) Further issues
In section E4 we have tried to convince the reader that the theoretical model presented in this chapter is both an innovation from the point of view of economic theory, and an empirically relevant model. In the final section of this chapter we want to raise two additional issues that shall be further discussed in the next chapter.

Firstly, we think that a comprehensive discussion about the role of labour mobility is needed. We have argued that any regional model must take into account geographical factor mobility and must say something about the causes and the consequences of it. Labour mobility is a crucial element in the interregional general equilibrium of the wage curve literature, it is essentially the distinguishing feature between NTT and NEG, and it also plays a very prominent role in our theoretical model. The next chapter F will therefore be devoted to a systematic assessment about the role of labour mobility in view of the different theoretical models we have discussed so far. The general issue at stake is if labour migration is an equilibrating force that ceteris paribus leads to an erosion of regional disparities, or if labour migration rather perpetuates regional asymmetries.

During this discussion, we will highlight that the type of the underlying technology is an important determinant of the spatial effects of labour migration. But we will also raise another issue that is crucial for the consequences of internal labour migration, but that has not been explicitly spelled out yet. Namely, we will explicitly distinguish between different types of labour, skilled and unskilled labour, and analyse the impacts of selective labour migration. We show that migration will not lead to regional convergence, but rather to a strengthening of regional disparities, particularly if only the high skilled workers are regionally mobile. This is surely
the case if the underlying technology is subject to agglomeration economies, but even under a very neoclassical set-up with constant returns to scale. The theoretical analysis in the next chapter, however, will also be interesting from another perspective. In the approach of this chapter, we have introduced unemployment through a wage curve, which has been grounded on efficiency wages. In the model in the next chapter, unemployment will occur, and regional unemployment disparities will arise, because of union wage setting. Since unionisation is one of the most salient features of European labour markets, this alternative approach seems interesting for its own sake already. But further insights will be gained by that model. We will consider one particular form of wage setting that seems very relevant for continental European economies. Namely, we will take the case where a national union sets national minimum wages irrespective of regional productivity differences. This already makes clear that the models derived in the next chapter will be specifically concerned with the phenomenon of intra-national unemployment disparities, not with the trans-national unemployment clusters that have been discussed in this chapter. The models we derive in the next chapter are no wage curve models, as will be discussed in detail below. But we reach conclusions that are consistent with the wage curve literature. The insights from the model approach of this chapter are thus complemented and extended.