

# The Spatial Economy of the EU: Patterns and Trends

Mark Bagley  
Department of Economics  
Supervisor: Joakim Gullstrand  
Date: 2<sup>nd</sup> September 2010

*Abstract:*

*This paper uses a New Economic Geography and Complex Systems framework to explore the nature of agglomeration economies on a sector by sector basis, as well as dropping many of the assumptions that are present with many mainstream schools of thought. By drawing upon a model of self-reinforcing behaviour, and examining the effects of self-organisation, path dependence, and discontinuous change, the trends of each sector are compared by using a spatial lag model. Causes of such trends are then inferred.*

## Table of Contents

<b>I. Introduction</b>	<b>4</b>
<b>II. Theoretical Background</b>	<b>4</b>
Defining Agglomeration Economies	5
Agglomeration Economies and Mainstream Schools of Thought	5
Neoclassical Theory (NCT) and New Trade Theory	5
Expanding on the Mainstream Schools of Thought	6
New Economic Geography and Complexity Economics	7
New Economic Geography (NEG)	7
Mechanisms of the Spatial Economy as a Complex Adaptive System	8
Critical Mass and the ‘Racetrack’ Model	10
The Schelling (1971) Simulation of Self-Organisation	13
Causes of Localisation	16
Before the Localisation Process: The Base Theory Hypothesis	17
<b>III. Methodology</b>	<b>19</b>
The Model	19
Data	20
Geographical Units	20
Employment and Income by Sector	21
Monetary and Instrumental Dummies	23
Descriptive Empirics	24
A-B: Primary Sector	24
C-E: Secondary Sector	24
G-I: Wholesale, Logistic, and Catering Sectors	24
J-K: Finance and Communication Sectors	25
L-P: Public, Technical, and Real Estate Sectors	25
Percentage Change in Total Labour	25
Percentage Change in GDP/Capita	26
Method	26
Moran’s <i>I</i> Statistic	26
Spatial Cluster Analysis	27
<b>IV. Results</b>	<b>28</b>
Moran’s <i>I</i> and Weight Choice	28
Regression Results	29
A-B: Primary Sector	29
C-E: Secondary Sector	30
G-I: Wholesale, Logistic, and Catering Sectors	32
J-K: Finance and Communication Sectors	32
L-P: Public, Technical, and Real Estate Sectors	33
<b>V. Conclusion</b>	<b>34</b>
<b>Bibliography</b>	<b>37</b>
<b>Appendix 1: Figures</b>	<b>40</b>
<b>Appendix 2: Variables</b>	<b>46</b>

## List of Figures

Figure 1: The Spatial Economy as a Complex Adaptive System	9
Figure 2: Local Economies and Critical Mass	11
Figure 3: The ‘Racetrack’ Model	12
Figure 4: A Schelling (1971) Simulation Using Three Types of Nodes	14
Figure 5: Percentage Change in Labour by NUTS-3 Region (A-B)	40
Figure 6: Sector Share in Labour (1997) by NUTS-3 Region (A-B)	40
Figure 7: Percentage Change in Labour by NUTS-3 Region (C-E)	41
Figure 8: Sector Share in Labour (1997) by NUTS-3 Region (C-E)	41
Figure 9: Percentage Change in Labour by NUTS-3 Region (G-I)	42
Figure 10: Sector Share in Labour (1997) by NUTS-3 Region (G-I)	42
Figure 11: Percentage Change in Labour by NUTS-3 Region (J-K)	43
Figure 12: Sector Share in Labour (1997) by NUTS-3 Region (J-K)	43
Figure 13: Percentage Change in Labour by NUTS-3 Region (L-P)	44
Figure 14: Sector Share in Labour (1997) by NUTS-3 Region (L-P)	44
Figure 15: Percentage Change in Total Labour by NUTS-3 Region	45
Figure 16: Percentage Change in GDP/Capita by NUTS-3 Region	45

## List of Tables

Table 1: Evolving Behaviours of the Spatial Economy	18
Table 2: NUTS Classification and Institutional Structures of Selected Countries	21
Table 3: Local Economies and Critical Mass	22
Table 4: Moran’s $I$ , % Change in Labour by Sector by Queen Continuity	28
Table 5: Moran’s $I$ , Initial Sector Share of Labour by Queen Continuity	29
Table 6: Moran’s $I$ by EU Expansion Zones and Sector, % Change in Labour	29
Table 7: Comparing Expansion Zones, Sectors A-B (Spatial Lag)	31
Table 8: Comparing Expansion Zones, Sectors C-E (Spatial Lag)	31
Table 9: Comparing Expansion Zones, Sectors G-I (Spatial Lag)	32
Table 10: Comparing Expansion Zones, Sectors J-K (Spatial Lag)	33
Table 11: Comparing Expansion Zones, Sectors L-P (Spatial Lag)	34
Table 12: Countries with missing data	46
Table 13: List of Eliminated Languages and Reasoning	49

## **I. Introduction**

This paper will explore the topic of the economies of agglomeration, and specifically the behaviour of location of firms on a sector-by-sector basis. A model of self-organization within a New Economic Geography framework is adopted, while using the ideas of path-dependence and discontinuous change in order to construct a model of self-reinforcing behaviour that is characterised by a mechanism of positive feedback loops. Economic activity may concentrate due to the incentives gained on the firm level, including knowledge spillovers. Such incentives are proxied by labour data obtained by Eurostat (2010) on a geographical basis. The NUTS-3 classification of European regions is used in conducting a spatial regression that captures sector behaviour. By using such an analysis the behaviour of each sector can then be explored, and specifically by stages of European expansion. The years 1997-2007 are used for this study and specifically for the percentage change in labour for a given NUTS-3 area and sector. By analysing sector behaviour one can then assess any relevant policy considerations in the EU with regard to industrial location policies.

This paper is structured as follows. A theoretical overview of the agglomeration process will first be put forward, specifically by outlining the main themes of the more mainstream schools of thought while introducing some newer concepts relating to New Economic Geography and Complex Systems. Specific models are drawn upon, such as that of Schelling (1971) and the ‘Racetrack Model’ as proposed by Krugman (1993b). These models set up the theoretical framework of this paper, and it is hypothesised that many industrial sectors, after developing their core economic base, then *localise* as economies become integrated. The next section sets out a testable model based on this theoretical framework, as well as introducing the data and analysing the descriptive empirics. The spatial lag model (for statistical testing) and Moran’s  $I$  (for cluster analysis) are also explained. After a section detailing the results, the ultimate section concludes.

## **II. Theoretical Background**

After introducing both the definition of the agglomeration process and walking through the mainstream school of thought, this section will introduce ideas in new Economic Geography and Complexity Economics. Three models will be introduced, namely the ‘Racetrack’ model, Schelling’s model of self-

organization, as well as critical mass. These models are used to form three central concepts that will be used in later sections.

### **Defining Agglomeration Economies**

Malmberg *et al* (1996: 87), following on from Lloyd and Dicken (1997), point out that an ‘Agglomeration Economy’ is often used in one of two contexts. The first type, often called *urbanization economies*, is the general economies of regional and urban concentration that apply to all firms and industries in a given location. The savings from such large scale operations of the urbanization economy as a whole is then passed on to individual firms as a result of the given economies of scale and scope. A result is the emergence of industrial core regions, such as manufacturing belts and metropolitan regions. The second type of agglomeration economies described in economic literature involve that of *Localisation economies*, which are the specific activities that involve firms engaged in similar or inter-linked industries. This in turn leads to the emergence of spatial structures of related firms.

### **Agglomeration Economies and Mainstream Schools of Thought**

#### *Neoclassical Theory (NCT) and New Trade Theory (NTT)*

With regard to the economics of location Neoclassical Theory (NCT) was built upon Ricardo (1817) and subsequently followed up by Heckscher (1919), Ohlin (1933), Weber (1909) and Vanek (1968). It is characterized by perfect competition, homogeneous products, and non-increasing returns to scale. The location of the firm is determined exogenously, in what Krugman (1993a) termed ‘first nature’, which is the given spatial distributions, natural endowments, technology, and factors of production. It is the spread or concentration of such underlying features that lead to the spread or concentration of economic activity. Such geographical spreads and concentrations thus result from inter-industry specialization, i.e. where different sectors settle in a given location with a matching comparative advantage. If one were to assume zero trade costs, the location of production is not affected as the spatial distributions of demand only affect the pattern of trade. However, if we were to assume trade costs, then the locational dispersion of activity will be positively correlated with the level of trade costs. Mobile factors therefore find it more profitable to move towards regions with a higher return. Thus, if trade costs were prohibitive, then the perfect dispersion of industries, following the geographical distribution of demand, would result.

New Trade Theory (NTT) was pioneered by Krugman (1979, 1980, 1981), Dixit and Norman (1980), Helpman and Krugman (1985), and Weder (1995). Natural resources and factors of production are not uniformly distributed across regions, and it is this unevenness that most of NTT is built upon (Fujita and Thisse 2002: 6). Unlike NCT, all exogenous 'first nature' elements are dropped except for that of market size. NTT stipulates that market size is determined by the size of the labour force in a particular country, which in turn is immobile across countries. 'Second nature' features are introduced such as imperfect competition, differentiated products, and increasing returns to scale. Such features are therefore independent of natural advantage and are entirely endogenous. The outcome of spatial distributions is divided into two layers. The first is that of *inter*-industry specialization, in which sectors cluster in locations which offer the best access to product markets. The second layer is that of *intra*-industry specialization, where each firm produces a unique and horizontally differentiated variety of the industry's product. As trade costs tend to zero, economic activity will tend to cluster near the core market, and the intra-industry trade between the core and periphery disappears (Brühlhart 1998: 777).

Both neoclassical theory and new trade theory can then be combined in order to explain the possible effects of expansion of economic integration. Integration may cause the restructuring of industry and thus provide a more efficient allocation of resources. If factors are immobile, and barriers prevent trade, then production must take place locally despite any potential gains from agglomeration. If goods and factor prices are immobile, then the forces of agglomeration and dispersion come into play. Industrial location patterns are then subject to the outcome of two forces. Agglomeration forces encourage firms to concentrate geographically as a result of localized external economies of scale, and dispersion forces encourage economic activity to spread out because production uses natural resources and other immobile factors of production. EU integration is expected to change both good and factor mobility; one would therefore expect a scenario that results from the forces of agglomeration (Midelfart-Knarvik and Overman 2002: 325, 326).

#### *Expanding on the Mainstream Schools of Thought*

With regard to NCT and NTT, agglomeration economies are a result of a process where links between firms, institutions, and infrastructures within a geographic area give rise to economies of scale and scope. Specifically, this is due to the development of general labour markets and pools of specialized skills, the enhanced interaction between local suppliers and

customers, and shared infrastructure, amongst others (Malmberg *et al* 1996: 87). Despite this focus on the efficiency and intensity between agglomeration firms, McCann (1995) found that such linkages tend to be empirically weak. In reality, a large proportion of firms have few or no trading links with other firms in the same industry, even in the presence of spatial clustering.

While some elements of New Trade Theory do capture some phenomena that are either covered or supported in this paper, such as path dependence and network effects, there has been some criticism of the principles that the first two schools of thought are based upon. They are heavily dependent on the ideas proposed by Walras (1874/1877), who stipulated that equilibria in economics could be mathematically applied. Most of these ideas were based on the ideas of 19<sup>th</sup> century thermodynamics, specifically that of the first law of thermodynamics (or the *conservation of energy principle*), which stipulates that energy cannot be created or destroyed. As argued by Beinhocker (2006: 67), such ideas of economic equilibria ignore the second law, that of *entropy*. Entropy in the physical sciences involves the inevitable decay of all matter in the universe, leading eventually to disorder and chaos. By ignoring the second law, such ideas of steady-state equilibria come into question.

Criticism has also been applied with the many assumptions that are made with both NCT and NTT. Economics is a social science that exhibits an enormous level complexity and variety in its agents. It can thus be understandably difficult to test economic models without making such general assumptions. An essay by Friedman (1953) defended the use of simplistic assumptions by arguing that if the economy behaves “as if” people were perfectly rational, then it doesn’t matter what the input is. However, it may be said the purpose of scientific argument isn’t to *predict* things, but to *explain* them, as argued by Beinhocker (2006: 49-50). Granted, the many assumptions used in the mainstream schools exist due to the lack of tools available to social scientists of the day. While not writing off the ideas the mainstream schools put forward, it may still be beneficial to incorporate many of their basic ideas and principles and *expanding* on them.

## **New Economic Geography and Complexity Economics**

### *New Economic Geography (NEG)*

Based on ideas proposed by Marshall (1920), and formalised by Krugman and Venables (1995a, 1995b) and Fujita et al. (1998), New Economic Geography models stipulate that

location becomes entirely endogenous, and ‘second nature’ determines everything, as factors and firms are now assumed to be mobile. Models of NEG analytically start with an abstract two- or three-dimensional space with uniformly distributed labour and output of a single industry. Due to the ‘second nature’ characteristics of the economy, such as market size externalities and input-output linkages, such a uniform distribution tends to be unstable. It is due to such ‘second nature’ characteristics that a self-reinforcing agglomeration process results. A new locational equilibrium arises from disturbances of the initial distribution of labour and output, and the specific location of any new equilibrium depends on the nature of the disturbance, as well as any sector- or industry specific characteristics. Furthermore, such centripetal forces may conflict with given centrifugal forces such as the increase in prices of immobile factors. The end result is that the economy becomes most spatially polarized at intermediate trade costs, and thus agglomeration economies relate non-monotonically to economic integration (Brühlhart 1998: 779).

Researchers have found it difficult to devise rigorous tests for New Economic Geography models. In what follows is thus a description of a more simplistic and abstract model of self-organization and self-enforcement that may lead the way to a more testable model, while using many of the core principles that lie behind NEG.

#### *Mechanisms of the Spatial Economy as a Complex Adaptive System*

Complexity Economics illustrates how various types of microeconomic structures lead to a particular aggregate phenomena. It is characterised by feedbacks between the environment as a whole and individual actors. A complex system is usually understood as an evolving process rather than that of a system residing in a ‘steady state’, and such a process involves the interrelationships of many actors who possess limited information about the beliefs, preferences and opportunities of others and such beliefs evolve from the feedback from the environment as a whole. Furthermore, the idea of ‘path dependence’, which refers to the permanence of seemingly random innovations, can have permanent consequences (Durlauf 1998: 158-160). To quote Krugman (1997), who summarized many of the ideas of complexity science with respect to spatial economies:



“The spatial economy is, self-evidently, a self-organizing system characterized by path dependence; it is a domain in which the interaction of individual decisions produces unexpected emergent behaviour at the aggregate level; its dynamic landscapes are typically rugged, and the evolution of the spatial economy typically involves “punctuated equilibria”, in which gradual change in the driving variables leads to occasional discontinuous change in the resulting behaviour”

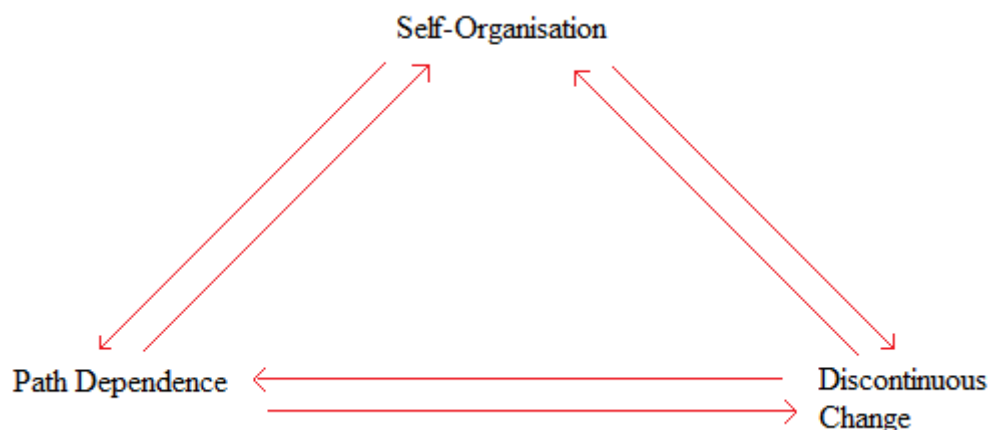
-Krugman (1997: 240)

Three central concepts regarding a spatial economic system will be backed up by three models. These three concepts are as follows:

1. Self-Organisation: Small-scale interactions among individual agents can spontaneously generate large scale order, or even the *appearance* of order.
2. Path Dependence: Initial conditions always have an effect on the long-run configuration.
3. Discontinuous change: Only small quantities in the underlying forces that drive locational behaviour can lead to large macro-level effect.

The concept of self-organisation is derived from the Schelling (1971) model. Both path dependence and discontinuous change is captured by Krugman’s (1993b) Racetrack model, and discontinuous change alone is a feature that is captured by the idea of critical mass. Schelling’s (1971) model illustrates how like-firms have the propensity to be close to one another even in the face of mild preferences. It should also be noted that the three points made above each have a mutual influence on each other, i.e. taking away the force of one will diminish the power of the other two, as seen in figure 1. Together they produce a *complex adaptive system*, a term coined by the Santa Fe Institute to describe a system that is *complex* as it is diverse and made up of multiple interconnecting elements, and *adaptive* in that it has the ability to learn from experiences and change.

*Figure 1: The Spatial Economy as a Complex Adaptive System*



It is therefore hoped that the above mechanism can be tested, by using each of the three tenets as representative of one or more independent variables. Additionally, it is aimed that such a test be as assumption free as possible, and by testing for the *properties* of agglomeration economies as a complex adaptive system instead of casual *causes* of agglomeration economies, many assumptions may be omitted.

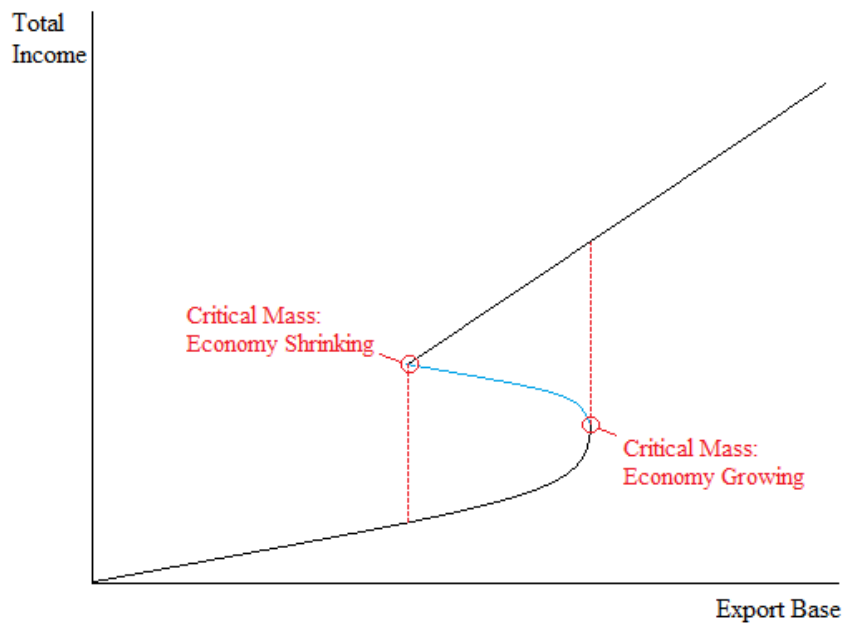
### *Critical Mass and the 'Racetrack' Model*

An agglomeration economy will be one that exhibits a large clustering of similar firms, which as an aggregate will have a larger export output than the majority of other regions, by comparison. This may be captured by the idea of critical mass. In such a model, a region has two types of activities: the 'export-base', i.e. activities that produce goods and services which are sold to consumers elsewhere, and the 'non-base' activities, i.e. those that produce goods and services to local residents. It is assumed that a fixed fraction of a region's income is spent on such locally produced goods and services. This gives a proportional relationship as follows:

$$Y = X + aX + a^2X + a^3X + \dots a^nX = \frac{X}{1 - a}$$

where  $a$  is the fraction of income spent locally,  $X$  is the income earned in the export sector, and  $Y$  is the region's total income. What this equation demonstrates is that the income earned in the export sector generates a local spending of  $aX$ , a fraction of this income in turn is spent locally giving rise to  $a^2X$ , and so on (Krugman 1997: 243). When introducing economies of scale, as performed by Pred (1966), it was proposed that when the export base reaches a critical mass a cumulative process of growth is set off. The expanding local economy attracts a wider range of import substituting activities, which causes the local economy to expand further, and so on. This may be illustrated in figure 2. There are in fact two critical masses, depending on whether the economy is growing or shrinking. Thus, we get both an implosion as well as an explosion (Krugman 1997: 243-245).

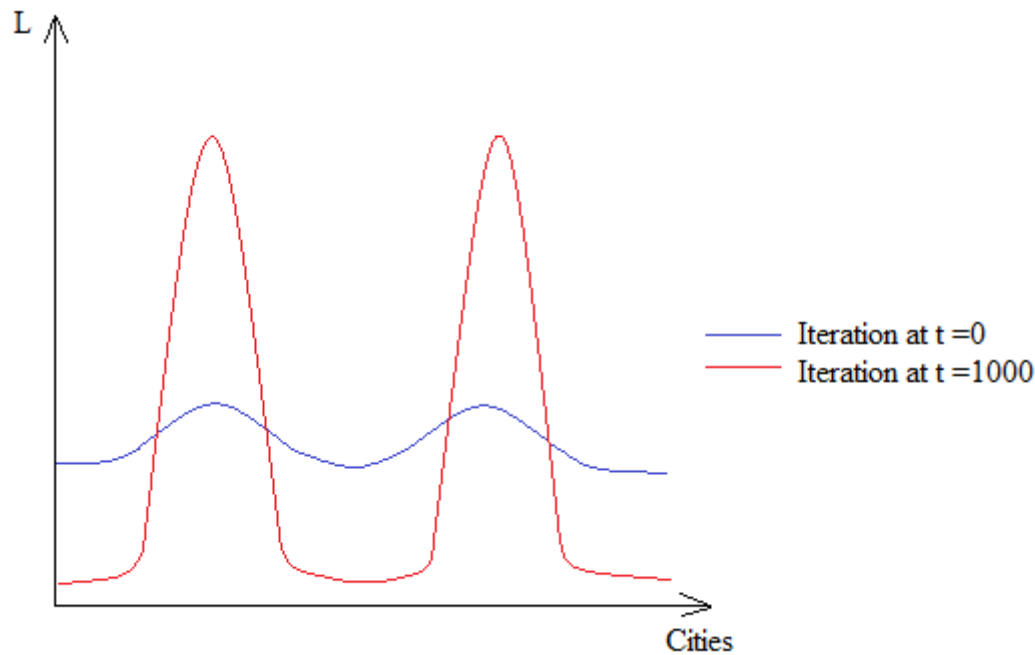
Figure 2: Local Economies and Critical Mass



Source: Krugman 1997: 245, with some basic modifications

The idea of critical mass was combined into Krugman's (1993b) 'Racetrack Model'. The model gets its name as it used 12 locations evenly spaced around a circle. The model focused more on the urbanization process and the formation of cities using two sectors, agricultural and manufacturing. The key concepts of the model may however be expanded upon to analyse the process of localisation. Manufacturers tend to clump together because of backward and forward linkages because of markets and supplies they provide to their workers and each other (Krugman 1993b: 294). The key result of the model's findings was that if the spatial economy starts off from a *near* uniform distribution of economic activity, some components of the series will grow, thereby producing an increasingly uneven spatial distribution. Moreover, the distribution becomes dominated by whatever location has grown the most rapidly. This may be seen in figure 3.

Figure 3: The 'Racetrack' Model



Often, using a twelve city approach, two or three agglomerations would often result. Thus, the economy will in effect experience an evolution in the form of self-organisation as a result of small underlying forces or differences. The locations that started off with a slightly larger level of manufacturing at the beginning of the simulation eventually ended up with one of the largest concentrations of manufacturing, and even sucked in the smaller manufacturing levels from neighbouring localities. Indeed, Fujita and Thisse (2002: 298) note that with regards to the spatial organization of a multiregional economy, one of the most striking results is that a “small initial advantage may lead to the emergence of a strongly polarized space once we explicitly account for the existence of localized production externalities, natural amenities, or both. This effect is magnified when the mobility of factors or the transportability of products are high, or both”. However, both the ideas of critical mass and the racetrack model give little explanation on the idea of *localisation* and are more geared towards the general *urbanisation* of industry as a whole. Ideas that take into account the differentiation of sectors and their individual traits are needed in order to incorporate ideas of localisation in a complexity setting.

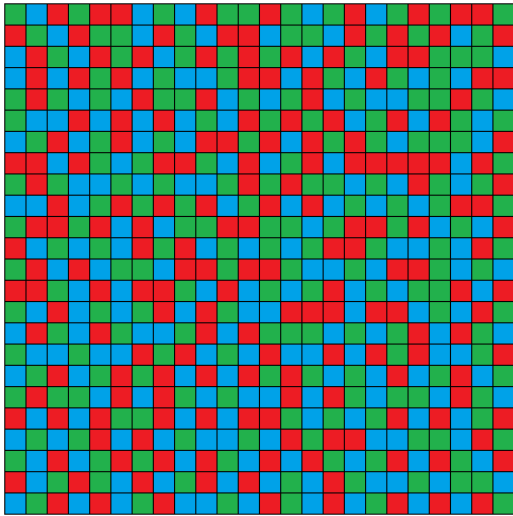
### *The Schelling (1971) Simulation of Self-Organization*

An early model of self-reinforcing behaviour is one devised by Schelling (1971), in what is often regarded as the first paper in economic complexity. Schelling's (1971) model originally looked at how ethnically segregated neighbourhoods could emerge in an unplanned fashion, and from a group of individual that exhibit relatively weak preferences when it comes to a neighbourhoods' ethnic composition.

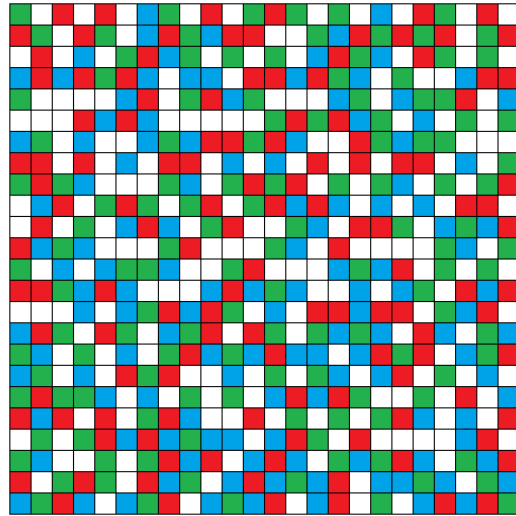
Schelling's (1971) model featured two nodes: black and white, in order to simulate an American city. However, the simulation may be expanded to include three different nodes: blue, green, and red; which may be thought of as three different sectors of the economy who exhibit some potential of getting along. Each node occupies a certain area on a 24x24 grid. The three-sector simulation was performed as follows, and may be seen in figure 4. First, 192 nodes of each colour were randomly assigned a place on the grid, thereby completely filling the grid (4.1). Then, 64 nodes of each colour were then removed from the grid, thereby giving way to empty cells and thus freedom for movement (4.2).

Figure 4: A Schelling (1971) Simulation Using Three Types of Nodes

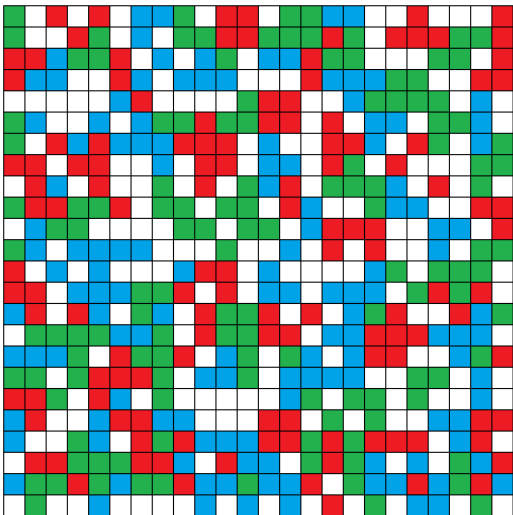
4.1: Initial Layout



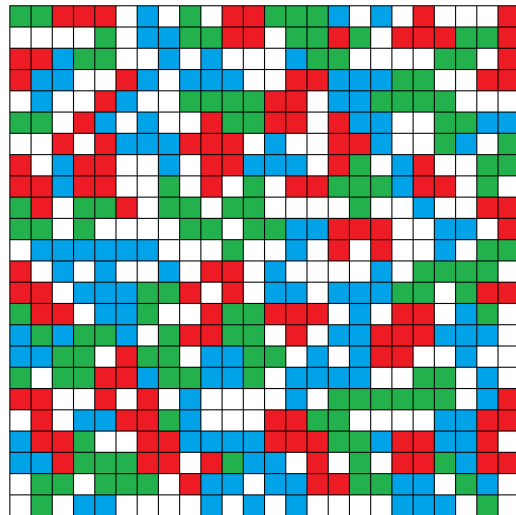
4.2: Nodes Removed



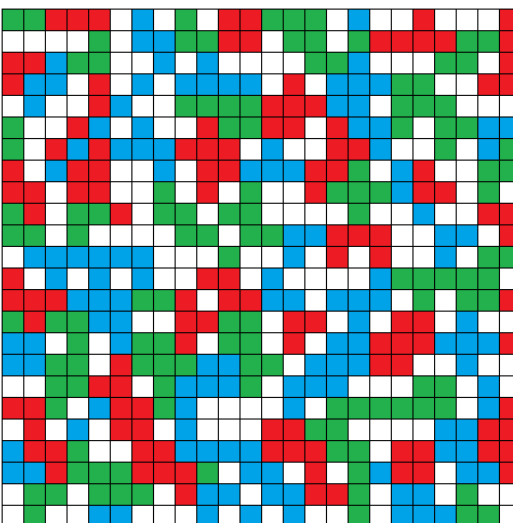
4.3: First Iteration



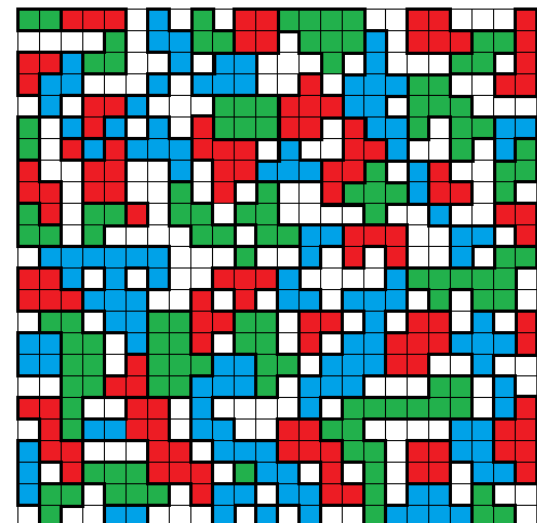
4.4: Second Iteration



4.5: Third Iteration



4.6: Final Iteration



It is assumed that each node cares about the types of her neighbours, which is defined as all of the 8 adjacent neighbouring squares to each node. Each node requires that some minimum fraction of these neighbours be the same type. If not, then the following occurs, starting from the top left square and moving down each row in a left to right fashion:

1. The node moves to an adjacent space which satisfies its preferences.  
If this is unavailable:
2. The node moves to an adjacent space that offers the most neighbours of the same type.  
If this is unavailable:
3. Any other empty space.  
If this is unavailable:
4. Does not move.

Furthermore, each node has the following preferences:

- A. 1 or 2 neighbours → 1 neighbour must be of the same type
- B. 3 to 5 neighbours → 2 neighbours must be of the same type
- C. 6 to 8 neighbours → 3 neighbours must be of the same type

This equates to the nodes being satisfied as long as at least 37.5% of their neighbours are of the same type as themselves. As seen in (4.2), many nodes appear to be in an acceptable location. However, when one node moves, a larger knock-on effect arises as seen in the first iteration (4.3). As the grid evolves, it turns out that when one node moves, this makes some other node unhappy, either due to the fact that her departure tips the balance in his old neighbourhood too far against her own type, or because of her arrival in a new location tips the balance there too far against nodes of the other type. The last iteration (more are possible) is seen in (4.6), when the grid exhibits an obviously emergent pattern. However it may be seen that the casual pattern emerges after even the first iteration. Such a pattern appears to be that of individuals with strong preferences with regard to their locational choice. What Schelling (1971) intended to show, however, was that *mild* preferences, which are on the surface compatible with an integrated structure, typically lead to a cumulative process of (in Schelling's context) segregation.

Krugman (1997: 241-243) notes two additional aspects to Schelling's (1971) model. First is the way in which short-range interactions produce a large-scale structure. Individuals care only about the type of their intermediate neighbours yet the end result is large and segregated neighbourhoods. The second aspect is the way in which a large-scale structure emerges from the model. The final details of the emergent structure are heavily dependent on the initial conditions. Thus, the final configuration in fig (4.6) is completely dependent on the initial

configuration (with empty spaces) in fig (4.2). In the context of this paper, it is ideal to think of each node as one firm that has a propensity for a similar neighbour. Yet that neighbour also has the same desires, and so on. The end result is a clustering of like-firms, or the localisation of sectors.

### *Causes of Localisation*

So far it has been explained how agglomeration economies may *grow* but it has been ignored how they may *begin* growing in the first place. Firms aim to undertake optimal strategies in order to enjoy the lower costs inside clusters or establish their plants in spatially separated markets in order to enjoy more market power. Drawing back to NCT, agglomeration forces tend to be especially strong when transport costs between regions are low, products are highly differentiated, or both. Agglomeration therefore occurs because firms are able to sell a large fraction of their output to distant markets. This is due to the fact that a higher degree of product differentiation in turn allows firms to relax price competition (Fujita and Thisse 2002: 269, 298). This is reinforced by common, customised infrastructure and a common, customised pool of labour.

In the context of knowledge-based firms such as research or finance sectors, this begins with the process of self-organisation, which in turn starts off the positive feedback loop as seen in figure 3. The main ingredient here is *knowledge spillover effects*, which involves the exchange of business information, know-how, and technological expertise, be it traded or non-traded. According to Malmberg (1996: 88-89), “industrial systems are not just fixed flows of goods and services but rather dynamic arrangements based on knowledge creation...Spatial clustering is argued to reinforce mutual commitments and moderate inclinations toward opportunism where trust may be viewed as a localized non-tradable input. Furthermore, knowledge-spillovers are accelerated in spatial clusters of related industries, and tacit knowledge tends to become embedded in local milieu”. According to Freeman (1991), there are three inter-related characteristics which are important for understanding the spatial clustering of related firms.

The first of these three characteristics is the need for the reduction of technical and economic uncertainty. Although uncertainty varies with the type of innovation in question, most technical aspects of innovation are derived by the means of trial and error. This in turn leads to a need for interaction, via both informal networks and formal co-operative agreements. Such interactions involve related firms directly or indirectly such as education



centres or research councils. Secondly, the ideas that lead to the actual development of manufacturing work often originate outside the firm in question. This may involve ideas originating from suppliers or the joint-development work of many firms. This in turn makes the innovation process highly interactive. Such interaction, coupled with the transfer of sensitive information, gives way to the need for trust and the importance of long-term business relationships. The third characteristic is the need for face-to-face contacts in the exchange and creation of new knowledge. Informal and oral communications of market opportunities often lead to innovation. According to Utterback (1974), it is often the unanticipated or unplanned personal encounters that lead to the most productive outcomes.

#### *Before the Localisation Process: The Base Theory Hypothesis*

One final note on the agglomeration process is that of the possibility of stages in local growth. The theories discussed so far have focussed on the localisation process. However, there is likely a preceding stage when the core activities of a sector have yet to develop. The distinction between the localisation and pre-localisation are characterised by the ratio of a region's output compared with its local demand. This may be explained using economic base theory, which stipulates that if there are more people engaged in a given activity than would expect under ratio comparisons of local employment to total employment in an industry compared to the local population and national population, the activity is 'basic' (Greenhut 1959: 71). The local industry is thus presumed to serve people beyond its locality. Expanding on this concept, the agglomeration process may be thought of as a two-stage one, where urbanization first takes place and then Localisation takes place afterward. It is possible that this may be reflected when comparisons are made between two different expansion zones of the European Union (i.e. EU10 and EU27). An area and its surrounding regions in, say, Romania may exhibit rampant growth in a certain industry that are not due to effects outlined in figure 3 but rather due to the rapid introduction of a sector that was previously un-established. To sum up, the effects in figure 3 may in fact of the *opposite* effect in such regions, i.e. not positive but negative. This will be explored in later sections.

The spatial economy may thus be thought of as on an evolutionary course. After an industrial sector has fully realised its non-basic potential, the process of localisation then occurs if a particular industry benefits from knowledge spillover effects or localized external economies of scale. Table 1 shows the relationships between the concepts discussed so far and their inferred effects on a region's labour growth. A region that is undergoing the pre-

localisation process will deceptively display indications of self-organisation, however this positive relationship is only indicative of a region undergoing similar levels of labour growth and not that of clustering. Income levels will be low and negatively correlated with labour growth. Low sector specialisation in regions will be associated with high levels of labour growth, and hence a negative correlation with path dependence exists. There would be little evidence of discontinuous change.

A region undergoing localisation will display a self-organization process that is a result of knowledge-spillovers in the case of technical industries or localized economies of scale in industries that benefit from shared infrastructure. Income may be positively correlated with labour growth due to the pull of higher wages that are a result of a local specialisation. As sector specialization and initial labour numbers have a long run impact on the specialization of the local economy, both path dependence and discontinuous change would be positively correlated.

After localisation, a sector may be seen to peak as a region's share of a given industrial sector has been exhausted. Path dependence therefore stops, while all other variables will remain positively correlated. Finally, an industry may be seen to perish in a given locality if growth in labour is negatively correlated with all of the three mechanisms of a complex adaptive system, as well as levels of income, which in this case will be comparatively high in the area of study as a whole. Such a phenomenon would occur if firms relocate their practices to areas of cheaper labour. Industrial flight in one region may also correspond to a localisation process in another region as a result of economic integration.

*Table 1: Evolving Behaviours of the Spatial Economy*

	Pre-Localisation	Localisation	Local Peaking	Industrial Flight
Self-Organisation	No	Yes	Yes	Yes or No
Path Dependence	No	Yes	No	No
Disc. Change	Yes or No	Yes	Yes	No
Income	No, low	Yes	Yes	No, high

### III. Methodology

In this section the three concepts of the spatial economy as a complex adaptive system are proxied by data obtained by Eurostat (2010). A basic model will first be introduced, and after some descriptive empirics, a final version of a testable model will be put forward. The spatial lag model is used to test for spatial clustering.

#### The Model

Overman and Puga (2002) examine geographical unemployment clusters that do not respect national boundaries. They find that during the polarization of unemployment rates in the 1980s, nearby regions tended to share similar outcomes due to related changes in labour demand. Such changes in labour demand were due in part to initial clustering of low skilled regions, while a significant neighbour effect remains even after controlling for such clustering. This paper uses a similar model to assess the clustering of different sectors.

An ideal model thus captures the effects of self-organization, path dependence, and discontinuous change given the constraints of the available data. To capture self-organization in a growth setting, it would be logical that the same variable is both dependent and independent. Growth in labour in a sector, in a given region, thus fits such conditions. The percentage change in labour in region  $i$  in sector  $c$ ,  $\% \Delta L_{ic}$ , is therefore assigned as the independent variable. The corresponding dependent variable, percentage change in labour in area  $j$  in sector  $c$ ,  $\% \Delta L_{jc}$ , captures the effects of changes in labour in regions neighbouring region  $i$ . Thus, small-scale interaction between areas  $i$  and  $j$  capture the idea of self-organization and thus the appearance of order. This spatial interdependence therefore reflects the location decisions of firms, with high levels of spatial interdependence existing in industries that benefit from knowledge-spillovers. The exact level of continuity between regions  $i$  and  $j$  will be determined in later sections by, and by using an index that measures global clustering (Moran's  $D$ ).

Path dependence is captured by the percentage of a region's total labour force in a given sector, in 1997,  $SHARE_{ic}^{1997}$ . As path dependence involves initial conditions having effects on long-run configuration, the 1997 figure is used, calculated as a percentage of total labour in a given area. A high coefficient for  $SHARE_{ic}^{1997}$  will occur if path dependence leads to further labour growth, and a negative coefficient may give rise to an explanation that a region's labour growth has already peaked due to 'saturation' of a given industry. Similarly, discontinuous

change is captured by the 1997 variable for initial labour force, in raw numbers, for area  $i$  and sector  $c$ . The discontinuous change element is the ‘magnitude’ variable, and it is hoped that small changes in the log of such numbers will result in larger percentage changes in labour in a given sector. This variable is thus denoted as  $lnL_{ic}^{1997}$ .

Further variables are used, including a given region’s initial income,  $lnInc_i^{1997}$ , in log form. This variable is included for two reasons. Firstly, it may be thought that the higher a region’s income, the more likely it will attract additional amounts of labour for a given sector upon EU expansion, considering an area’s large sector share in a given industry. Conversely, the indicator may give rise to the base theory hypothesis and that the lower an area’s income, the more likely that it will exhibit larger levels of growth in a given industry.

Finally, dummy variables are included to capture institutional similarities between regions. The first of these is whether or not a country uses the Euro, denoted by  $EURO$ , for all countries that joined the Eurozone by 2001. It is theorized that a common currency between two countries will facilitate the ease of business transactions. Secondly, language dummies,  $[LANG]$ , are used in order to capture the ease of cross-area trade due to the ease of communication. Rauch (1999) correlated the ease of language with networking ties.

The basic model therefore may so far be expressed as:

$$\begin{aligned} \% \Delta L_{ic} = & \beta_1 + \beta_2 \% \Delta L_{jc} + \beta_3 SHARE_{ic}^{1997} + \beta_4 lnL_{ic}^{1997} + \beta_5 lnInc_i^{1997} \\ & + \beta_6 EURO + \beta_N [LANG] \end{aligned}$$

The language dummy is for now written in parenthesis and will be expanded upon later.

## Data

### *Geographical Units*

This paper explores a spatial analysis of agglomeration economies, and specifically a spatial analysis that explores the economic relationships of the smallest geographical unit as possible. It is generally assumed that the larger the statistical unit, the greater the loss of inter-regional information. Eurostat uses the NUTS (*Nomenclature d’unités territoriales statistiques*) classification as a hierarchical system for dividing up the economic territory of the EU for the purpose of socio-economic analyses of regions. The hierarchical classification divides each EU state into a number of NUTS-1, with each number of NUTS-1 divided into a number of NUTS-2, and so on. Most data from Eurostat regard both the NUTS-1 and NUTS-2

levels, and very occasionally for the NUTS-3 level. Each NUTS classification corresponds to institutional and historical territories and provinces, which is summarised in table 2.

*Table 2: NUTS Classification and Institutional Structures of Selected Countries*

	<b>Sweden</b>	<b>UK</b>	<b>France</b>	<b>Germany</b>	<b>Spain</b>
<b>NUTS-1</b>	Landsdelar	Regions	Z.E.A.T	Länder	Groups of Comunidades
<b>NUTS-2</b>	Riksområden	Counties	Régions	Regierungsbezirke	Comunidades
<b>NUTS-3</b>	Län	Unitary Authority	Départements	Kreise	Provincias

Source: Eurostat: 2010

As the classification is largely based on existing divisions, there is thus no consistent standardization across the EU (Combes and Overman 2003: 3). Thus, when using such data, problems arise with regard to normalisation with respect to area and population sizes. In addition, areas that are defined as NUTS-3 areas may have a population that ranges from 150,00 to 800,000, a sizeable range.

A total of 919 observations were included for the EU10, 1073 for the EU15, 1217 for the EU25, and 1287 observations for the EU27

#### *Employment and Income by Sector*

The majority of data concerning the percentage change in labour, sector share, and 1997 employment for NUTS-3 areas were obtained from raw employment numbers from Eurostat (2010). Some of the data was incomplete, however, especially for some of the earlier years. In such cases data was extrapolated using a least squares method, the specifics of which may be summarised in the appendix. In total, such estimations had to be performed for 9 countries, and for a differing number of NUTS-3 areas. The most recent data is for the year 2009 and the oldest stretches back to 1995. For industrial sectors, Eurostat uses NACE (*Nomenclature statistique des activités économiques dans la Communauté européenne*) codes in order to group economic activity into classifications. For data regarding employment numbers, these codes were available as follows:

Table 3: NACE Code Classification for Employment Numbers

Sector	NACE Code
1. Primary	A - Agriculture, forestry and fishing B - Mining and quarrying
2. Secondary	C – Manufacturing D - Electricity, gas, steam and air conditioning supply E - Water supply; sewerage; waste management and remediation activities
3. Wholesale, logistic and catering	G – Wholesale and retail trade; repair of motor vehicles and motorcycles H – Transporting and storage I – Accommodation and food service activities
4. Finance and Communication	J – Information and communication K – Financial and insurance activities
5. Public, technical, and real estate	L – Real estate activities M – Professional, scientific and technical activities N – Administrative and support service activities O – Public administration and defence; compulsory social security P – Education

Admittedly, the grouping of the NACE codes as performed by Eurostat is not ideal, especially with regard to the topic of this paper. Many NACE sectors, for example *K-Financial and insurance activities* and *L-Real Estate Services* would be better off in the same category. Furthermore it would have been beneficial to have *M-Professional, scientific and technical activities* in a grouping that is apart from public sector activities. However, given that Eurostat does not offer any options that are more detailed than groupings in table 2, there is no option but to use what is given.

Due to the absence of employment data for parts of Scotland and Northern Ireland, difference sources were used as opposed to Eurostat. For Scotland, data was obtained from the NOMIS database, and for Northern Ireland, the Northern Ireland Neighbourhood Information Service (NINIS). The geographical areas used in these two databases are not in the format of NUTS-3 as used by Eurostat, with County Councils or groups of County

Councils overlapping or cumulating into a NUTS-3 unit. Thus, estimations were required with many of the NUTS-3 units in these areas, the method of which may be found in the appendix.

Figures for income in each NUTS-3 region were obtained from Eurostat (2010). No estimations were required the data was complete for all regions.

### *Monetary and Institutional Dummies*

Dummy variables for the use of the Euro were given to all countries according to the situation in 2002 only. For language dummies, was obtained from the European Commission (2006). Language fluency was marked on a NUTS-3 level, as many languages overlap national borders. On a national level, a country was marked as ‘speaker’ of English, German, or Russian if over 50% of the population of that country was fluent in the language, in addition to countries that have such languages as an official language. Czech and Slovak were treated as mutually intelligible languages. Irish was not used as only 2% of Ireland’s population are native speakers, and in a multitude of vastly dissimilar dialects. Finally, in order to obtain the same specification or each EU regression, each language was assigned into one of four ‘pots’. The details of these ‘pots’ are explained in the appendix, and were constructed in such a way that when running regressions for each of the EU expansions, it would be impossible to arrive at an ill-defined matrix.

Given this method of dealing with the language dummies, the specification of the equation to estimate may now be re-written as:

$$\begin{aligned} \% \Delta L_{ic} = & \beta_1 + \beta_2 \mathbf{W} \% \Delta L_{ic} + \beta_3 SHARE_{ic}^{1997} + \beta_4 \ln L_{ic}^{1997} + \beta_5 \ln Inc_i^{1997} \\ & + \beta_6 EURO + \beta_7 LANG_1 + \beta_8 LANG_2 + \beta_9 LANG_3 + \beta_{10} LANG_4 \end{aligned}$$

Each of the four ‘pots’ of language dummies are now included in the specification.  $\mathbf{W} \% \Delta L_{ic}$  is the dependent variable multiplied by the weights matrix. The weights matrix is used in a spatial lag model, and is explained later.

## **Descriptive Empirics**

The following maps were made by using data from Eurostat (2010) and compiling it using ArcGIS. All maps may be found in the appendix.

### *A-B: Primary Sector*

Figure 4 illustrates the percentage change in labour by NUTS-3 region. As expected, there seems to be relatively little in the way of change across the EU27, with pockets of increases found in the UK, Spain, Bulgaria, and Hungary. Any increase is likely due to the expansion of farming and mining in such areas, which in turn is completely dependent on a region's availability of raw materials or fertility. Figure 5, illustrating sector share in labour for the primary sector, likely points to a region's reliance on labour instead of its mechanisation. The primary sector is highly immobile, but is however included in this study in order to offer a comparison between sectors A and B and the rest of the sectors of interest.

### *C-E: Secondary Sector*

Figure 6 shows the percentage change in labour by NUTS-3 region for the secondary sector. Most of the increases may be found in Spain, Greece, as well as Central and Eastern Europe. Figure 7, illustrating the secondary sector's share in labour, displays an obvious case of clustering, especially around the former Czechoslovakia, Northern Italy, Hungary, the Ruhr, Bavaria, and southern parts of Scandinavia. Thus, some evidence of a 'manufacturing belt' may exist. More dependent on infrastructure and transport links, this may indeed be an explanatory factor. The increases in labour numbers typically take place outside of these manufacturing belts, which may either be due to the 'base theory hypothesis' as discussed earlier or the saturation of existing areas of manufacturing, thereby forcing new or expanded entrants to establish themselves elsewhere. With regards to Spain, the establishment of a network of 'dry ports' during the past 15 years may have led to the rise of manufacturing in the majority of Spain's NUTS-3 areas.

### *G-I: Wholesale, Logistic, and Catering Sectors*

This sector classification is ambiguous with regards to the subject of this paper. The catering industry is not considered to be particularly mobile, while the wholesale and logistic sectors are considered to be slightly more mobile. When examining the percentage change in labour (figure 8), there is no clear pattern with the exceptions of Spain, Ireland, and Latvia as



a whole. Outside these countries, there appears to be a small scattering of high increases in Central and Eastern Europe, and to a lesser extent, Western Europe. With regard to specialization (figure 9), it is apparent that the majority of clustering is centred either around ports (for logistic) and tourist destination (catering). A weak correlation is therefore expected for sectors G-I.

#### *J-K: Finance and Communication Sectors*

These sectors are expected to see the highest evidence of clustering, due to the high degree of mobility of the finance sector as well as the complementarity between the finance and communication sectors. Figure 11, illustrating sector share, shows a high concentration in large cities, and an especially apparent pattern of specialisation that spans from the south-east of the UK (capturing London and the Home Counties), to the entirety of The Netherlands and the south of Belgium. Increases in the percentage change in labour (figure 10) occur entirely outside this region, indicating the possibility of saturation in these established NUTS-3 areas. A negative relationship may then be expected.

#### *L-P: Public, Technical, and Real Estate Sectors*

Similar to sectors G-I, this is a poorly-defined category. The growth in the public sector is heavily dependent on country-specific policies, and the real estate sector may encompass both public and private-sector activities. It is also linked to general economic growth of an area. The technical sector, if private-based, is the only sector that should exhibit agglomeration-like behaviour. Figure 13 illustrates the country-wide specificity of sectors J-K. Scandinavia, the UK, and France exhibit the highest sector shares, whereas Central and Eastern Europe exhibit the lowest. This is not only country-specific but also due to the mass-privatisation policies seen in the former Soviet satellite states. Growth in sectors L-K (figure 12) is again subject to country specific policies, as seen in Ireland, Spain, Poland, and Greece. Like sectors A-B, category L-P has been included for comparison purposes.

#### *Percentage Change in Total Labour*

For illustrative purposes, percentage change in total labour by NUTS-3 region may be seen in figure 4. The largest increases occur in Spain and Ireland. Little may be interpreted out of this; however, as this paper focuses on individual sectors specifically and not general increases in labour. A view is taken that the percentage change in total labour is a result of the cumulative increases in individual sectors, and not the other way around.

## *Percentage Change in GDP/Capita*

Figure 15, percentage change in GDP per capita by NUTS-3 region, illustrates unsurprisingly that the majority of large increases in income lie within the newer EU member states, as well as Spain and Ireland. All other areas of Europe have either fallen or risen slightly over the 10 year period. The areas with large increases in income may be linked to the base theory hypotheses when analysing labour growth in specific sectors: a NUTS-3 region with high increases in GDP/Capita may also exhibit non-base growth within certain sectors. Thus, the higher the percentage change in GDP/Capita, the lower the likelihood of sector or neighbour effects in a given region. This will be explored by means of comparing regression outcomes in later sections.

## **Method**

### *Moran's I Statistic*

Moran's *I* statistic, as developed by Moran (1950), is an indicator that detects global clustering. It detects whether nearby areas have similar or dissimilar attributes overall, i.e. positive or negative correlation (Wang 2006: 172). It is calculated as:

$$I = \frac{N \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_i \sum_j w_{ij}) \sum_i (x_i - \bar{x})^2}$$

Where  $N$  is the total number of areas,  $x_i$  and  $x_j$  are the attribute values for areas  $i$  and  $j$ ,  $\bar{x}$  is the mean of attribute values, and  $w_{ij}$  are the spatial weights. The spatial weights are essentially the nearby neighbours, which may be compiled by means of distance or directly bordered neighbours. In the case of this paper, the latter was chosen as when using distance with regard to spatial weights, areas that are not physically linked may still be regarded as neighbours, such as islands or any two areas divided by bodies of water. This goes against the theoretically underpinnings of this paper. A disadvantage of using directly bordered weights, however, is that results may be affected by the differing areas of each NUTS-3 zone, coupled with the fact that a NUTS-3 area may be defined in the relatively long range of 150 to 800 thousand inhabitants.

Moran's *I* can thus be interpreted as the correlation between a variable and its spatial lag. The Spatial lag for variable  $x$  is the average value for  $x$  in neighbouring areas  $j$  and is defined as:

$$x_{i,-1} = \frac{\sum_j w_{ij} x_j}{\sum_j w_{ij}}$$

Moran's  $I$  varies between -1 and 1. A value near 1 indicates that similar attributes are clustered, a value near -1 indicates dissimilar attributes are clustered, and a value of 0 indicates a random pattern or absence of spatial autocorrelation (Wang 2006: 173).

### *Spatial Cluster Analysis*

A Moran's  $I$  that is close to 1 or -1 gives reasoning to apply an analysis that takes into account spatial autocorrelation. In the absence of such, OLS can be used. However, when spatial dependence is present, the residuals are no longer independent from each other, and therefore OLS is no longer applicable. There are two types of method that exist when taking into account spatial autocorrelation, both of which are maximum likelihood estimators. The first of these is the *spatial error* model, which assumes that the *residuals* explanatory variables of nearby areas are correlated with each other. The other type is the *spatial lag* model, as used by Baller et al. (2001), which includes the mean of the dependent variable in neighbouring areas (i.e. the spatial lag) as an extra explanatory variable. It is the spatial lag model that is used in this paper as the effects of the explanatory variable (percentage change in labour in neighbouring areas) is of interest. Spatial lag is suggestive of a possible diffusion process, i.e. events in one place predict an increased likelihood of similar events in neighbouring areas. The spatial lag of  $\mathbf{y}$  is written as  $\mathbf{W}\mathbf{y}$ , where  $\mathbf{W}$  is the weights matrix. This may be expressed as:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Where  $\rho$  is the regression coefficient for the spatial lag. Rearranging:

$$(\mathbf{I} - \rho\mathbf{W})\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Assuming  $(\mathbf{I} - \rho\mathbf{W})$  is invertible, we have:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

This final form shows that the value of  $y_i$  at each location  $i$  is determined not only by  $x_i$  at that location (like in the OLS model) but also by the  $x_j$  at other locations through the spatial multiplier (Wang 2006: 181). Instead of a t-statistic, a z-value is obtained in spatial lag

estimates, which, upon rejecting the null, also tell us that a given variable exhibits spatial correlation.

## IV. Results

### Moran's $I$ and Weight Choice

Table 4 shows the value for Moran's  $I$ , when examining percentage change in labour by sector, using three different queen values. Queen 1 is an order of continuity that indicates neighbours are included one border away from an area of question, as well as any touching vertices. Queen 2 denotes two borders away and includes all of queen 1, and so on. It may be seen that Moran's  $I$  significantly falls in each sector from queen 1 to queen 2, and then gradually falls for most sectors from queen 2 to queen 3. The queen order with the highest value of Moran's  $I$  is thus included in the regressions, in this case queen 1. Indeed it may also be shown that using this order of continuity produces the highest  $R^2$  values. Moran's  $I$  also gives some clue to the degree of clustering, with sectors L-P and G-I exhibiting the highest level of clustering while sectors A-B exhibiting near randomness. A different picture is seen when exploring Moran's  $I$  with respect to sector share (table 5), where A-B and J-K exhibit the highest rates. This may be explained by the fact that sectors A-B have already 'topped out' with regards to labour growth, but are more likely to be located next to a NUTS-3 area that also specializes in primary sector activities. In sum, high levels of clustering are seen in all sectors be it high-high or low-low clustering.

*Table 4: Moran's  $I$ , Percentage Change in Labour by Sector by Queen Continuity*

	<b>Moran's <math>I</math> (% Change in Labour by Sector, EU27)</b>		
	Queen 1	Queen 2	Queen 3
Primary	-0.0684	-0.0124	-0.0051
Secondary	0.2605	0.1387	0.1612
Wholesale, Logistic, and Catering	0.3326	0.2350	0.2375
Finance and Communication	0.2031	0.1757	0.1536
Public, Technical, and Real Estate	0.3812	0.2956	0.2569

Table 5: Moran's *I*, Initial Sector Share of Labour by Queen Continuity

<b>Moran's <i>I</i> (Sector Share, 1997, EU27)</b>			
	Queen 1	Queen 2	Queen 3
Primary	0.7466	0.6887	0.6394
Secondary	0.4728	0.3783	0.3037
Wholesale, Logistic, and Catering	0.4253	0.3637	0.3220
Finance and Communication	0.6168	0.5544	0.5034
Public, Technical, and Real Estate	0.5506	0.5155	0.4754

Table 6 depicts the change in Moran's *I* upon each EU expansion, using queen 1. The primary sector undergoes very little change upon EU expansion as expected. Sectors C-E, G-I, and J-K exhibit increasing levels of clustering as the European Union shrinks to the EU15, and then falls again when only including countries from the EU10. These changes may point either to the base theory hypothesis, especially at the first EU expansion for the EU15, or more likely that existing specialized regions within the new European expansion (Austria, Finland, Sweden, Spain and Portugal) were already witnessing growth in labour in those sectors. Indeed, it is likely that sector L-P increases in Moran's *I* upon each EU expansion for this very reason, as a result of country-specific policies.

Table 6: Moran's *I* by EU Expansion Zones and Sector, Percentage Change in Labour

<b>Moran's <i>I</i> (% Change in Labour by Sector, Queen 1)</b>				
	EU27	EU25	EU15	EU10
Primary	-0.0684	-0.0686	-0.0687	-0.0688
Secondary	0.2605	0.2667	0.2667	0.2340
Wholesale, Logistic, and Catering	0.3326	0.3842	0.3868	0.2660
Finance and Communication	0.2031	0.2482	0.2754	0.2354
Public, Technical, and Real Estate	0.3812	0.3734	0.3612	0.3220

## Regression Results

### *A-B: Primary Sector*

As noted, queen 1 was used as a weight for all regressions. As expected, there is little in the way of meaningful results in the primary sector. As mentioned, the primary sector was included in this study in order to act as a control. As the agriculture and mining sectors are completely immobile due to the availability of farmable land and raw materials, very little in the way of agglomeration is expected. Indeed, the only coefficient that is rejected as being

zero is the percentage of labour of an area's neighbour. This coefficient undergoes negligible change upon each phase of EU expansion. Coupled with low values for  $R^2$ , and insignificant coefficients for all other variable, any spatial pattern in the primary sector is indeed meagre. Brülhart and Traeger (2005: 614) find that the agricultural sector is highly concentrated with respect to areas of low population density. This is not disputed as table 5 shows that the agricultural sector is indeed the most concentrated. However, this paper explores concentration patterns with regards labour changes and not sector share.

#### *C-E: Secondary Sector*

The secondary sector exhibits a much clearer scenario. The neighbour effect increases upon each subsequent EU expansion zone. Initial labour numbers have a negative impact on the percentage change in labour numbers, which steadily converges towards zero as the zone expands. Thus, the higher amount of initial raw numbers in the secondary sector, the less likely the sector will expand. As this has a larger effect the smaller the EU zone, there is an indication that with regards to manufacturing, there is obvious shrinkage in the sector as a whole, which is especially more pronounced in the older EU members. Convergence is thus not likely. The spread of manufacturing to Asia may explain this. This may also be reflected in the negative and mostly significant coefficients for income, which adds to the observation that manufacturing is partial to inexpensive labour. Finally, it may be noted that, upon each zonal expansion, sector specialization has an increasingly negative effect on the growth of labour. We may speculate that areas that specialize in manufacturing are not only prone to 'topping out' but are also affected by the gradual transformation of industrial organization in the newer EU members.

Table 7: Comparing Expansion Zones, Sectors A-B (Spatial Lag)

	EU27	EU25	EU15	EU10
% $\Delta$ Labour, Neighbour	-0.235*** (0.043) <i>(-5.481)</i>	-0.235*** (0.044) <i>(-5.357)</i>	-0.232*** (0.046) <i>(-5.019)</i>	-0.231*** (0.049) <i>(-4.687)</i>
Raw Labour, 1997	12.580 (21.980) <i>(0.572)</i>	17.222 (23.616) <i>(0.729)</i>	19.822 (26.382) <i>(0.751)</i>	24.502 (32.621) <i>(0.751)</i>
Sector Specialization	-1.913 (2.880) <i>(-0.664)</i>	-1.841 (3.696) <i>(-0.498)</i>	-1.873 (5.028) <i>(-0.373)</i>	-2.408 (7.345) <i>(-0.328)</i>
Income	-11.554 (81.629) <i>(0.034)</i>	12.147 (56.423) <i>(0.215)</i>	27.343 (89.760) <i>(0.305)</i>	31.529 (105.960) <i>(0.298)</i>
Euro Dummy	Yes	Yes	Yes	Yes
Language Dummies	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.03	0.03	0.03	0.03

Notes: \*\*\* denotes coefficient significantly different from zero with a 1% confidence level. Standard errors are reported in parenthesis and z-values are reported in parenthesis with italics.

Table 8: Comparing Expansion Zones, Sectors C-E (Spatial Lag)

	EU27	EU25	EU15	EU10
% $\Delta$ Labour, Neighbour	0.353*** (0.034) <i>(10.369)</i>	0.333*** (0.035) <i>(9.482)</i>	0.345*** (0.036) <i>(9.395)</i>	0.247*** (0.042) <i>(5.909)</i>
Initial Labour, 1997	-2.383*** (0.654) <i>(-3.544)</i>	-2.637*** (0.664) <i>(-3.974)</i>	-3.316*** (0.709) <i>(-4.577)</i>	-4.574*** (0.876) <i>(-5.222)</i>
Sector Specialization	-0.237*** (0.071) <i>(-3.348)</i>	-0.191*** (0.072) <i>(-2.643)</i>	-0.118 (0.079) <i>(-1.504)</i>	-0.003 (0.088) <i>(-0.031)</i>
Income	-4.404*** (0.919) <i>(-4.789)</i>	-6.623*** (1.239) <i>(-5.346)</i>	-7.446*** (1.937) <i>(-3.844)</i>	-3.653 (2.340) <i>(-1.561)</i>
Euro Dummy	Yes	Yes	Yes	Yes
Language Dummies	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.22	0.23	0.23	0.22

Notes: \*\*\* denotes coefficient significantly different from zero with a 1% confidence level. Standard errors are reported in parenthesis and z-values are reported in parenthesis with italics.

*G-I: Wholesale, Logistic, and Catering Sectors*

No clear pattern is seen in the relationship of percentage change in labour with regard to neighbour effects, upon comparing EU expansion zones. A positive relationship with initial labour numbers with regards to percentage increases in labour seem to point to the fact that, with regards to the logistic and wholesale sectors, agglomeration may indeed occur due the presence of innate infrastructure, and has therefore ‘locked-in’ the agglomeration of these industries with regards to a given area. With the catering industry, the high reliance of labour coupled with the location-specific nature of the industry may also give rise to the positive correlation. This is not reflected with sector specialization; however, as the negative correlation may suggest that labour numbers in sectors G-E are undergoing an exodus out of established regions. The negative correlation with income further reinforces this idea. However, due to the poor categorization of sectors G-E, it is difficult to draw concrete deduction from these results.

*Table 9: Comparing Expansion Zones, Sectors G-I (Spatial Lag)*

	EU27	EU25	EU15	EU10
% $\Delta$ Labour, Neighbour	0.343*** (0.035) <i>(9.892)</i>	0.437*** (0.032) <i>(13.483)</i>	0.444*** (0.033) <i>(13.267)</i>	0.359*** (0.039) <i>(9.215)</i>
Initial Labour, 1997	1.420** (0.594) <i>(2.390)</i>	1.407*** (0.523) <i>(2.689)</i>	2.335*** (0.554) <i>(4.215)</i>	1.158* (0.658) <i>(1.761)</i>
Sector Specialization	-0.415*** (0.097) <i>(-4.288)</i>	-0.454*** (0.086) <i>(-5.253)</i>	-0.503*** (0.092) <i>(-5.475)</i>	-0.638*** (0.111) <i>(-5.747)</i>
Income	-4.482*** (0.883) <i>(-5.076)</i>	-5.466*** (0.985) <i>(-5.548)</i>	-10.088*** (1.472) <i>(-6.855)</i>	-5.705*** (1.701) <i>(-3.355)</i>
Euro Dummy	Yes	Yes	Yes	Yes
Language Dummies	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.22	0.30	0.32	0.21

*Notes:* \*\*\*, \*\*, and \* denote coefficients significantly different from zero with a 1%, 5%, and 10% confidence level, respectively. Standard errors are reported in parenthesis and z-values are reported in parenthesis with italics.

*J-K: Finance and Communication Sectors*

With positive coefficients for initial labour, the effect of discontinuous change is strong with the finance and communication sectors. This effect stays at comparably the same level



for each EU expansion zone. Positive neighbour effects are also present. Sector specialization is negatively correlated with labour growth, which may be explained that a) in the EU10 and EU15, the increasing use of information technology has allowed firms to ‘scatter’ their operations geographically, for example locating ‘back-office’ functions to sub-prime locations in town or even other NUTS-3 areas, and b), the base theory hypothesis may hold for EU25 and EU27 countries, where the changing landscape of industrial structure has led to the development of the finance sector in previously unspecialized areas. The latter is especially true for Warsaw, Prague, and Gdansk. Finally, there is a clear pattern that the income effect substantially increases as the EU expansion zone gets smaller. This is hardly surprising as, within the EU10, the main centres of finance are also amongst Europe’s richest. The low coefficients for income for the EU25 and EU27, coupled with the steady coefficients of sector specialization across all EU expansion zones further adds to points a) and b).

*Table 10: Comparing Expansion Zones, Sectors J-K (Spatial Lag)*

	EU27	EU25	EU15	EU10
%ΔLabour, Neighbour	0.250*** (0.036) (6.985)	0.286*** (0.036) (8.008)	0.257*** (0.038) (6.801)	0.199*** (0.042) (4.727)
Initial Labour, 1997	4.471*** (1.238) (3.613)	4.034*** (1.213) (3.326)	5.939*** (1.204) (4.933)	3.462** (1.466) (2.361)
Sector Specialization	-2.673*** (0.348) (-7.696)	-2.600*** (0.339) (-7.677)	-2.983*** (0.322) (-9.254)	-2.333*** (0.355) (-6.572)
Income	1.393 (1.871) (0.744)	3.488 (2.300) (1.516)	11.599*** (3.214) (3.609)	16.697*** (3.693) (4.521)
Euro Dummy	Yes	Yes	Yes	Yes
Language Dummies	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.15	0.18	0.23	0.21

*Notes:* \*\*\* and \*\* denote coefficients significantly different from zero with a 1% and 5% confidence level, respectively. Standard errors are reported in parenthesis and z-values are reported in parenthesis with italics.

#### *L-P: Public, Technical, and Real Estate Sectors*

As mentioned, not only is this category poorly defined but it is also policy specific, both initially as well as for the ten years in question. As a result, the coefficients for initial labour numbers have not been rejected for the most part. Policy-specificity may also lead us to ignore the neighbour effects, as any measures in labour will likely be seen in one country as a

whole. Sector specialization, although significant, are trivial in this case as it simply points to the fact that the expansion of the public sector has increased from 1997 to 2007. The effects of income are both significant and erratic. Across EU expansion zones this observation further compounds the idea that labour expansion in sectors L-P are country-specific policy-based.

*Table 11: Comparing Expansion Zones, Sectors L-P (Spatial Lag)*

	EU27	EU25	EU15	EU10
% $\Delta$ Labour, Neighbour	0.427*** (0.031) <i>(13.920)</i>	0.439*** (0.031) <i>(14.026)</i>	0.430*** (0.033) <i>(13.078)</i>	0.341*** (0.037) <i>(9.184)</i>
Initial Labour, 1997	0.724 (0.492) <i>(1.471)</i>	0.573 (0.496) <i>(1.155)</i>	0.975** (0.529) <i>(1.844)</i>	0.157 (0.630) <i>(0.249)</i>
Sector Specialization	-0.596*** (0.062) <i>(-9.671)</i>	-0.585*** (0.066) <i>(-8.931)</i>	-0.741*** (0.071) <i>(-10.489)</i>	-0.699*** (0.077) <i>(-9.026)</i>
Income	4.081*** (0.710) <i>(5.748)</i>	2.438*** (0.893) <i>(2.728)</i>	-4.018*** (1.386) <i>(-2.899)</i>	-1.157 (1.645) <i>(-0.704)</i>
Euro Dummy	Yes	Yes	Yes	Yes
Language Dummies	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.33	0.31	0.32	0.31

*Notes:* \*\*\* denotes coefficient significantly different from zero with a 1% confidence level. Standard errors are reported in parenthesis and z-values are reported in parenthesis with italics.

## V. Conclusion

This paper has examined the spatial economy as a self-reinforcing, complex dynamic system. The pre-existing conditions, namely path dependence and discontinuous change, combine with the ideas of self-organization to produce a pattern of spatial clustering or declustering based on the characteristics of the NACE sector in question. The behaviour of each sector, given that it is a sector based on knowledge-spillovers, take different directions, yet significant directions due to underlying forces that are specific to a certain industry. Despite the fact that Eurostat constructs a poorly categorized grouping of various industrial sectors, the results of spatial lag analysis, which tests for both significance of a coefficient as well as significance of high-high or low-low clustering, have proved noteworthy. Furthermore it has been argued that traditional schools of thought regarding agglomeration economies may be dropped in favour of amore simplistic approach as many assumptions that make up the

mainstream schools of thought have been dropped. By testing for the *behaviour* of specific sectors instead of testing for possible *causes* of agglomeration, one can then infer a cause of behaviour.

It was noted that the primary sector, due to the immobility of factors of production, result in a regression outcome that points little in the way of clustering with respect to income, discontinuous change, or path dependence. This was expected in a labour growth model, as labour in the primary sector is linked to the resources available for extraction, which in turn or finite in a static sense (in the mining sector) or a dynamic sense (in the agricultural sector). The secondary sector, although witnessing increases in labour in sector share in parts of Europe, is undergoing a process of industrial flight in many of the older EU countries, while seeing a process of pre-industrialization in the newer EU states. The wholesale, logistic, and catering sectors suffer from a poorly-defined category, and it is difficult to draw any concrete conclusions. The finance and communications sectors have undergone a process of localisation, and may have begun to peak in specific NUTS-3 areas. This is not to say that the Localisation process is over, as the large finance clusters in Europe are in fact ‘growing’ instead of localising on an individual NUTS-3 basis. This is not picked up by the specification of the model. Finally, the public, technical, and real estate sector, while not only suffering from being poorly categorized, is also influenced by country-specific policies rather than region-specific factors of localisation.

Studying the behaviour of industries on a sector-by-sector basis is of great interest to policy makers. People with the highest skills will benefit from agglomeration forces by moving to regions which offer them the highest rewards. However, as not all factors of production are equally mobile many people must stay in regions of decline. Because labour mobility in Europe is especially low, Policy makers often attempt to counter forces of agglomeration based on welfare reasons (Martin 1998: 765). However, the relationship between geography and the factors that affect it are not linear. Depending on the industry in question, projects such as the construction of new public infrastructure may not be affective as the concentration of firms in a region may be self-sustaining. Policy makers may often take advantage of certain areas that are subject to the impacts of new technologies, such as the finance and communication (J-K) category. These policies that are designed to target, for example, low-level manufacturing, are fraught with problems. Furthermore, caution should be applied when analysing changes occurring in the new EU members, who may well be

undergoing a process of 'non-basic' growth and in the process of urbanization and restructuring rather than localisation.

## Bibliography

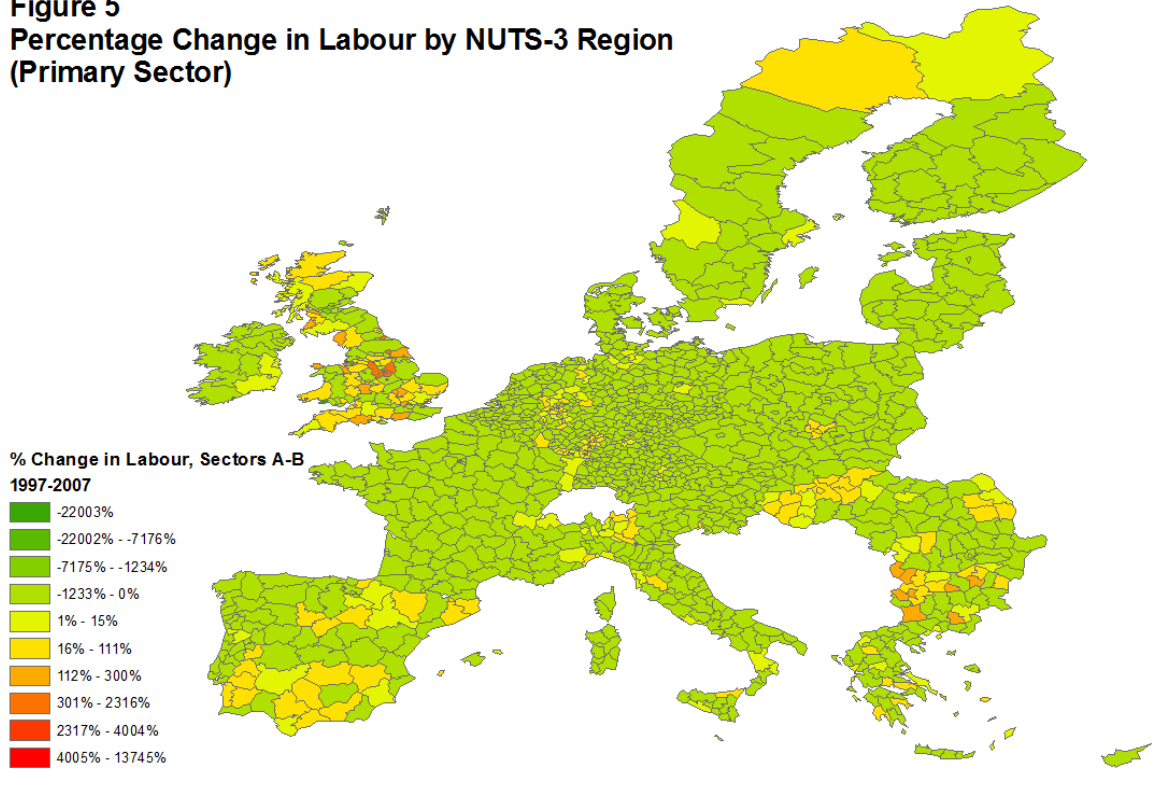
- Baller, R., L. Anselin, S. Messner, G. Deane, and D. Hawkins (2001) "Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects" *Criminology* 39(3): 561-590
- Beinhocker, E.D. (2006) *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics*. Boston, MA: Harvard Business School Press
- Brühlhart, M. (1998) "Economic Geography, Industry Location and Trade: The Evidence" *The World Economy* 21(6): 775-801
- Brühlhart, M. and R. Traeger (2005) "An account of geographic concentration patterns in Europe" *Regional Science and Urban Economics* 35(6): 597-624
- Combes, P-P. and H.G. Overman (2003) "The Spatial Distribution of Economic Activities in the European Union" *CEPR Discussion Paper No. 587*
- Dixit, A. and V. Norman (1980) *Theory of International Trade*. Cambridge: Cambridge University Press
- Durlauf, S.N. (1998) "What Should Policy Makers Know About Economic Complexity?" *The Washington Quarterly* 21(1): 155-165
- European Commission (2006) "Europeans and their Languages" *Special Eurobarometer 243*
- Eurostat (2010) *Regional Statistics*. Available online at:  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/region\\_cities/regional\\_statistics](http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics).  
Accessed June 18th 2010
- Freeman, C. (1991) "Networks of innovators: a synthesis of research issues" *Research Policy* 20(5): 499-514
- Freidman, M. (1953) "The Methodology of Positive Economics" in Freidman, M. (ed.) *Essays in Positive Economics* (1966) pp. 3-16, 30-43. Chicago: Chicago University Press
- Fujita, M., P. Krugman, and A.J. Venables (1998) *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press
- Fujita, M. and J.F. Thisse (2002) *Economics of Agglomeration*. Cambridge: Cambridge University Press
- Greenhut, M.L. (1959) "Comments on Economic Base Theory" *Land Economics* 35(1): 71-75
- Heckscher, E.F. (1919) "Utrikeshandelns verken på inkomstfördelningen" *Ekonomisk Tidskrift* 21(2): 1-32

- Helpman, E. and P. Krugman (1985) *Market Structure and Foreign Trade*. Cambridge, MA: MIT Press
- Krugman, P. (1979) "Increasing Returns, Monopolistic Competition, and International Trade" *Journal of International Economics* 9(4): 469-79
- Krugman, P. (1980) "Scale Economies, Product Differentiation, and the Pattern of Trade" *American Economic Review* 70(5): 950-9
- Krugman, P. (1981) "Intraindustry Specialization and the Gains from Trade" *Journal of Political Economy* 89(6): 959-73
- Krugman, P. (1993a) "First Nature, Second Nature, and Metropolitan Location" *Journal of Regional Science* 33(2): 129-44
- Krugman, P. (1993b) "On the Number and Location of Cities" *European Economic Review* 37(2): 293-298
- Krugman, P. (1997) "How the Economy Organizes Itself in Space: A Survey of the New Economic Geography" in Arthur, W.B, S.N. Durlauf, and D.A Lane (eds.) *The Economy as an Evolving Complex System II. Santa Fe Institute Studies in the Sciences of Complexity, Vol. XXVII* (1997) pp. 239-262. Reading, MA: Addison-Wesley
- Krugman, P. and A.J. Venables (1995a) "Globalization and the Inequality of Nations" *Quarterly Journal of Economics* 110(4): 857-80
- Krugman, P. and A.J. Venables (1995b) "The Seamless World: A Spatial Model of International Specialization" *NBER Discussion Paper No. 5220*
- Lloyd, P.E. and P. Dicken (1977) *Location in space. A Theoretical approach to economic geography*. 2<sup>nd</sup> Edition. London: Harper & Row
- Martin, P. (1998) "Can growth policies affect growth and geography in Europe?" *The World Economy* 21(6): 757-774
- McCann, P. (1995) "Rethinking the economics of location and agglomeration" *Urban Studies* 32(3): 563-577
- Malmberg, A, Ö. Sölvell, and I. Zander (1996) "Spatial Clustering, Local Accumulation of Knowledge and Firm Competitiveness" *Geografiska Annaler. Series B, Human Geography* 78(2): 85-97
- Marshall, A. (1920) *Principles of Economics*. London: Macmillan
- Midelfart-Knarvik, K.H. and H.G. Overman (2002) "Delocation and European integration: is structural spending justified?" *Economic Policy* 17(35): 321-359
- Moran, P.A.P. "Notes on Continuous Stochastic Phenomena" *Biometrika* 37(1): 17-23

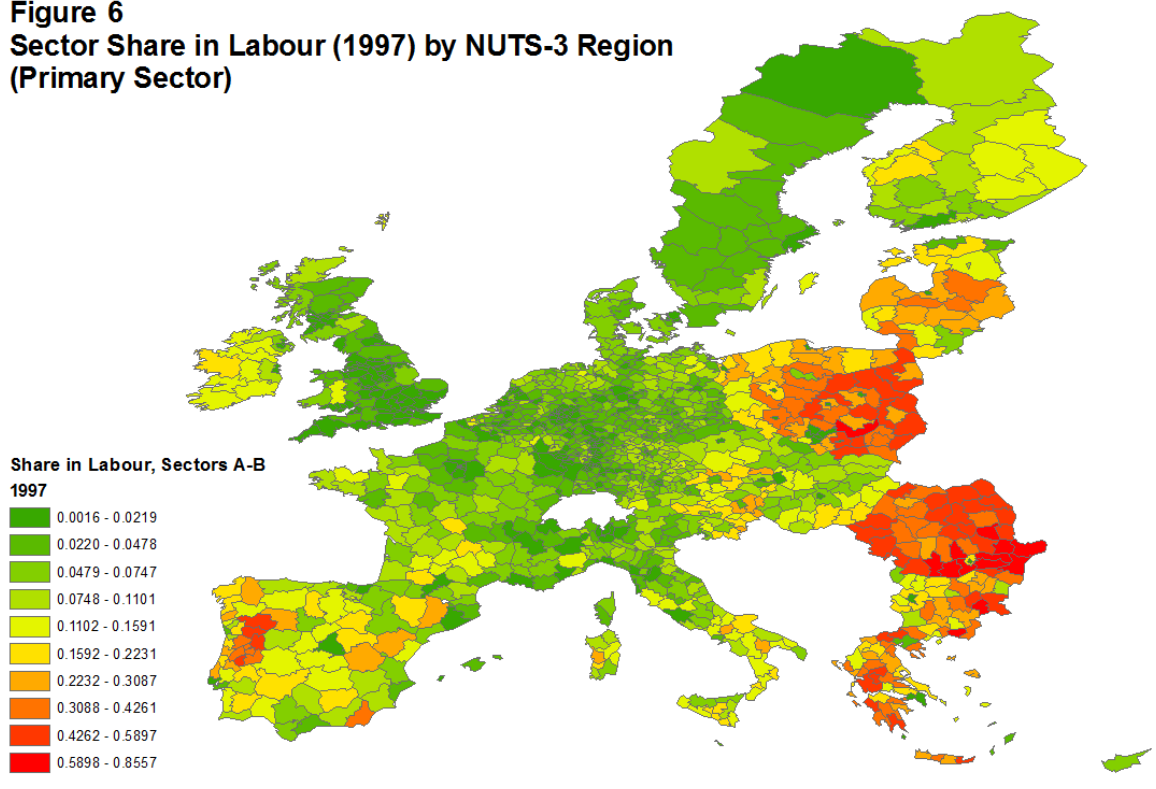
- NOMIS (2010) *Official Labour Market Statistics*. Available online at: <https://www.nomisweb.co.uk>. Accessed June 21<sup>st</sup> 2010
- NINIS (2010) *Labour Market Statistics*. Available online at: <http://www.ninis.nisra.gov.uk>. Accessed June 21<sup>st</sup> 2010
- Ohlin, B. (1933) *Interregional and International Trade*. Cambridge, MA: Harvard University Press
- Overman, H.G. and D. Puga (2002) "Unemployment Clusters across Europe's Regions and Countries" *Economic Policy* 17(34): 115-147
- Pred, A. (1966) *The Spatial Dynamics of US Urban-Industrial Growth*. Cambridge, MA: MIT Press
- Rauch, J.E. (1999) "Networks versus Markets in International Trade" *Journal of International Economics* 48(1): 7-35
- Ricardo, D. (1817) *On the Principles of Political Economy and Taxation*. Harmondsworth: Penguin
- Schelling, T. (1971) "Dynamic Models of Segregation" *Journal of Mathematical Sociology* 1(2): 143-86
- Utterback, J. (1974) "Innovation in industry and the diffusion of technology" *Science* 183(2): 658-662
- Vanek, J. (1968) "The Factor Proportions Theory: The N-Factor Case" *Kyklos* 21(4): 749-56
- Walras, L. (1874/1877) *Eléments d'économie politique pure*. Lausanne: Corbaz
- Wang, F. (2006) *Quantitative Methods and Applications in GIS*. Boca Raton: CRS
- Weber, A. (1909) *Über den Standort der Industrien*. Tübingen: J.C.B. Mohr
- Weder, R. (1909) "Linking Absolute and Comparative Advantage to Intra-Industry Trade Theory" *Review of International Economics* 3(3): 342-54

## APPENDIX 1: Figures

**Figure 5**  
**Percentage Change in Labour by NUTS-3 Region**  
**(Primary Sector)**

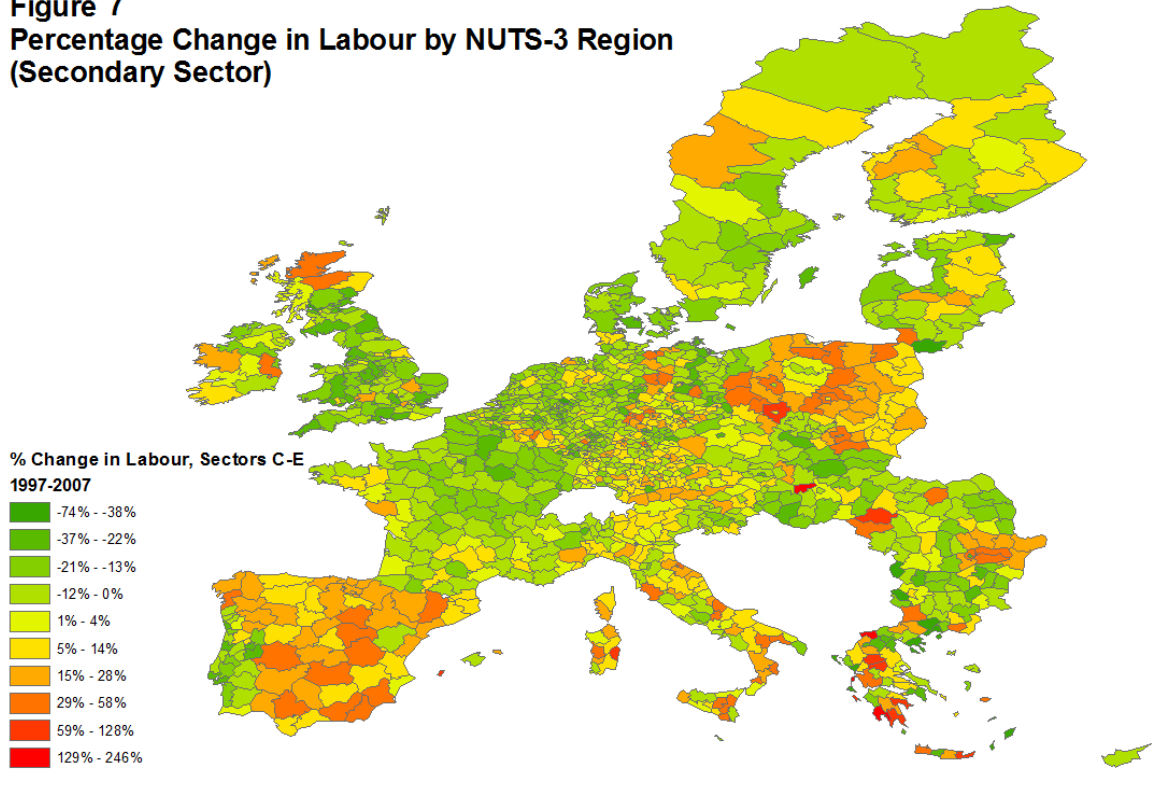


**Figure 6**  
**Sector Share in Labour (1997) by NUTS-3 Region**  
**(Primary Sector)**

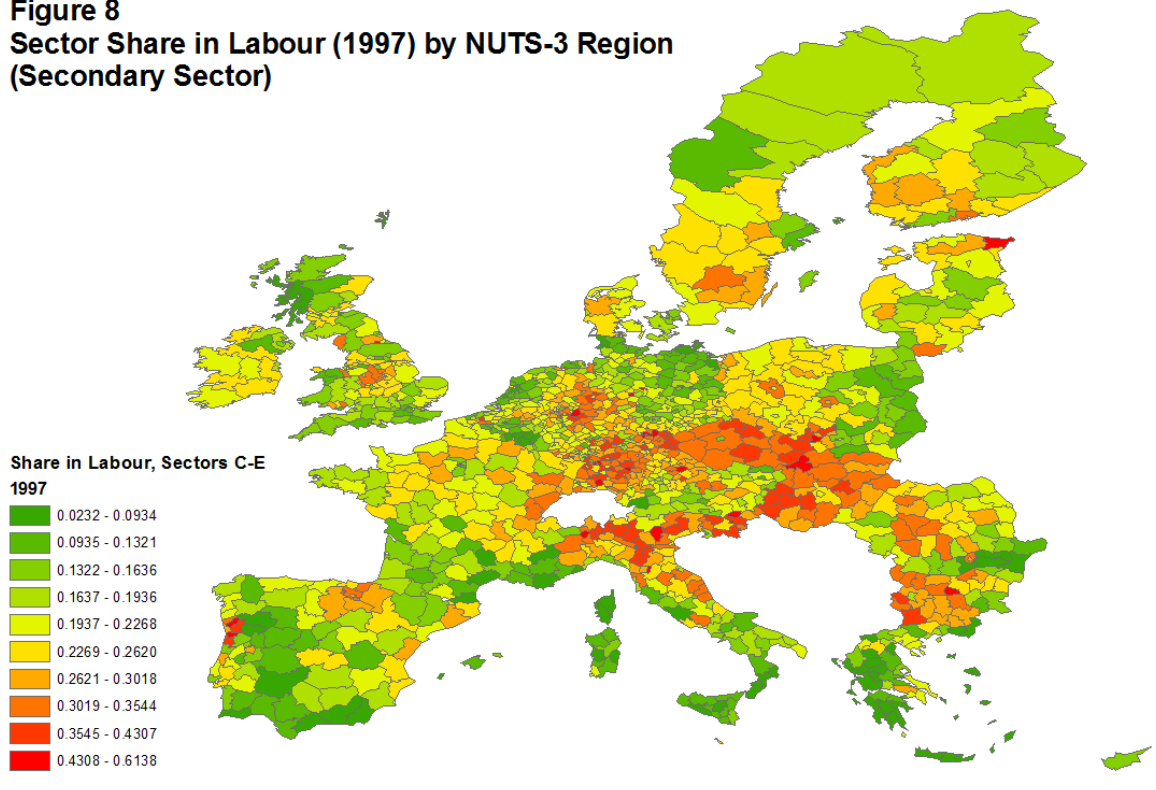




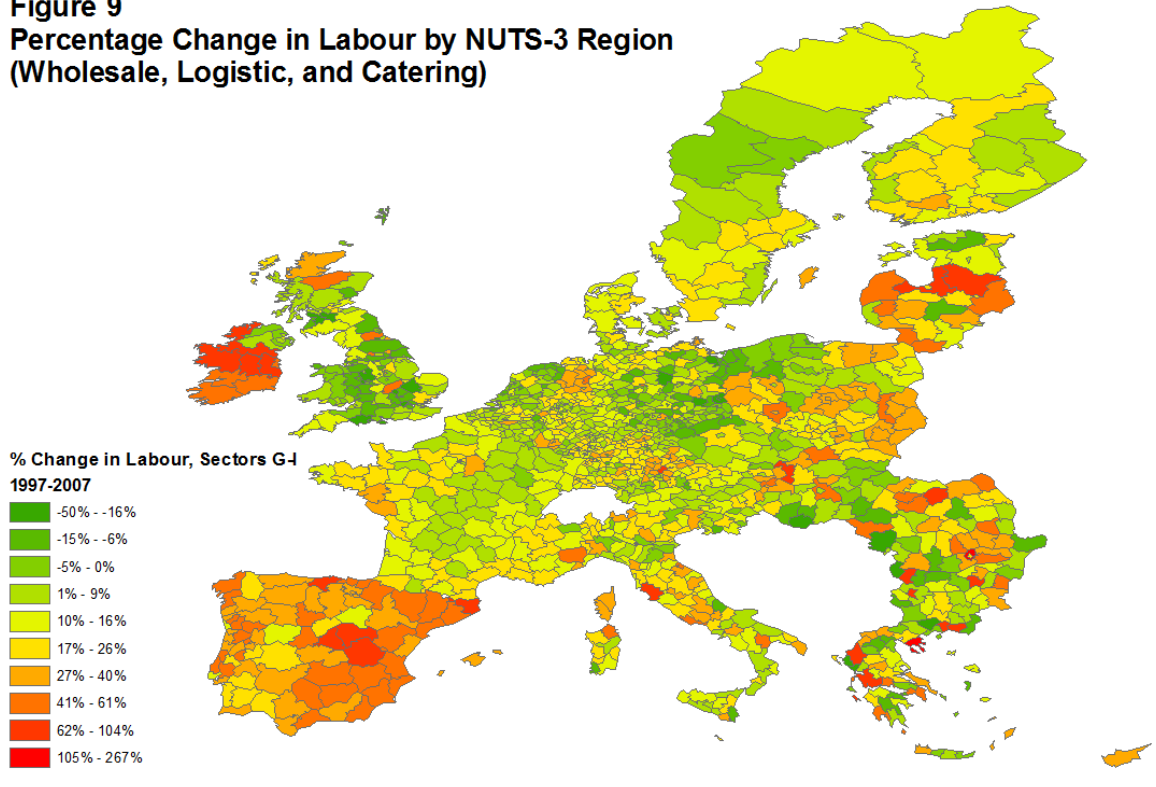
**Figure 7**  
**Percentage Change in Labour by NUTS-3 Region**  
**(Secondary Sector)**



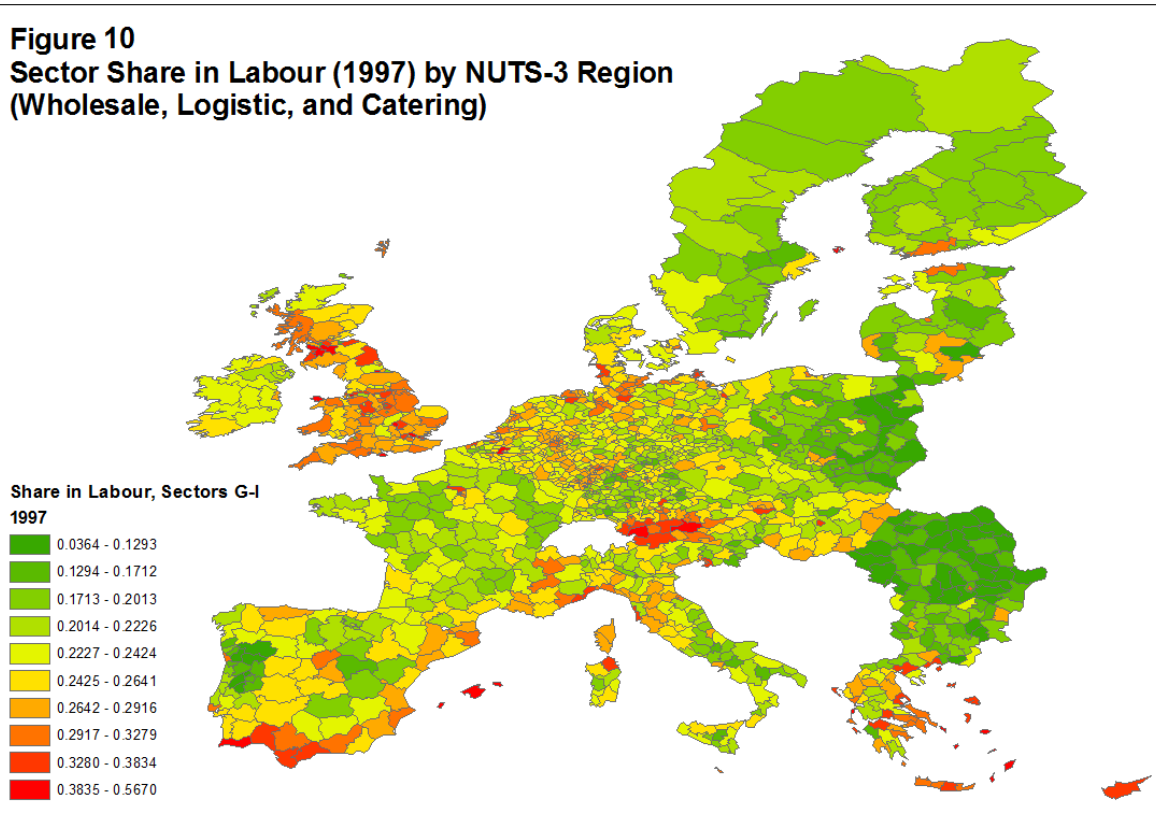
**Figure 8**  
**Sector Share in Labour (1997) by NUTS-3 Region**  
**(Secondary Sector)**



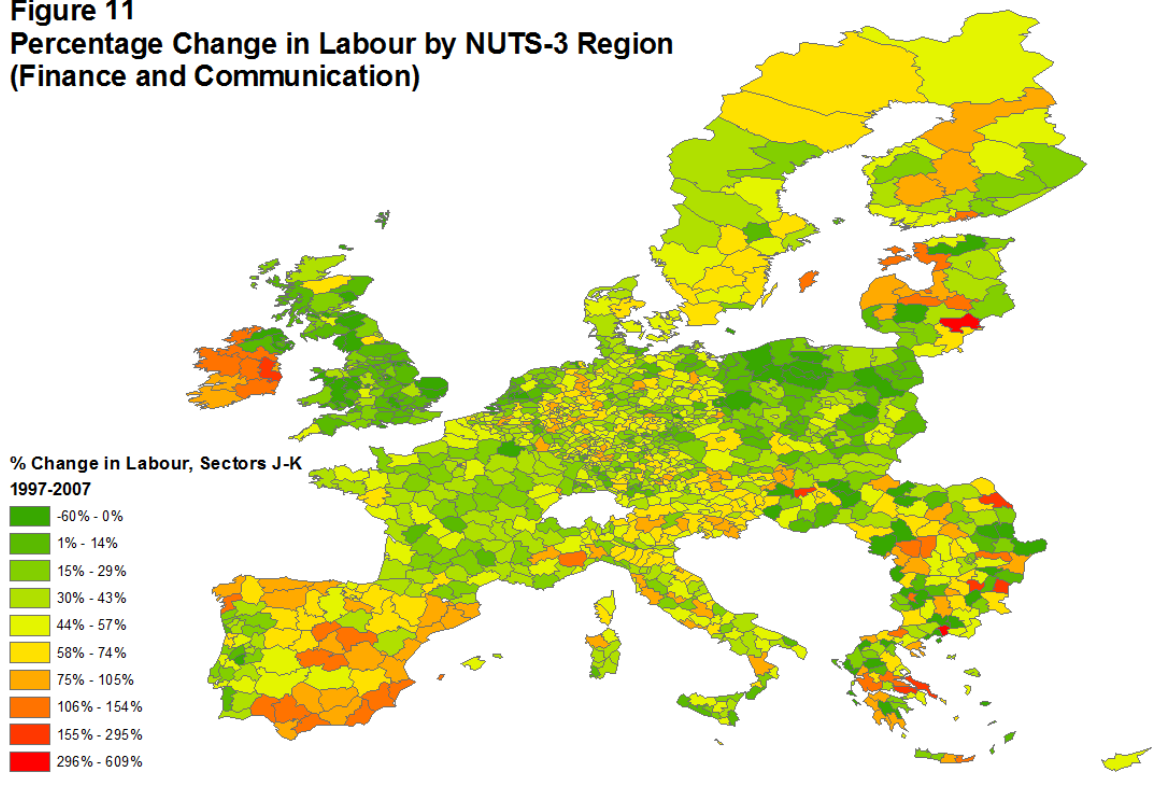
**Figure 9**  
**Percentage Change in Labour by NUTS-3 Region**  
**(Wholesale, Logistic, and Catering)**



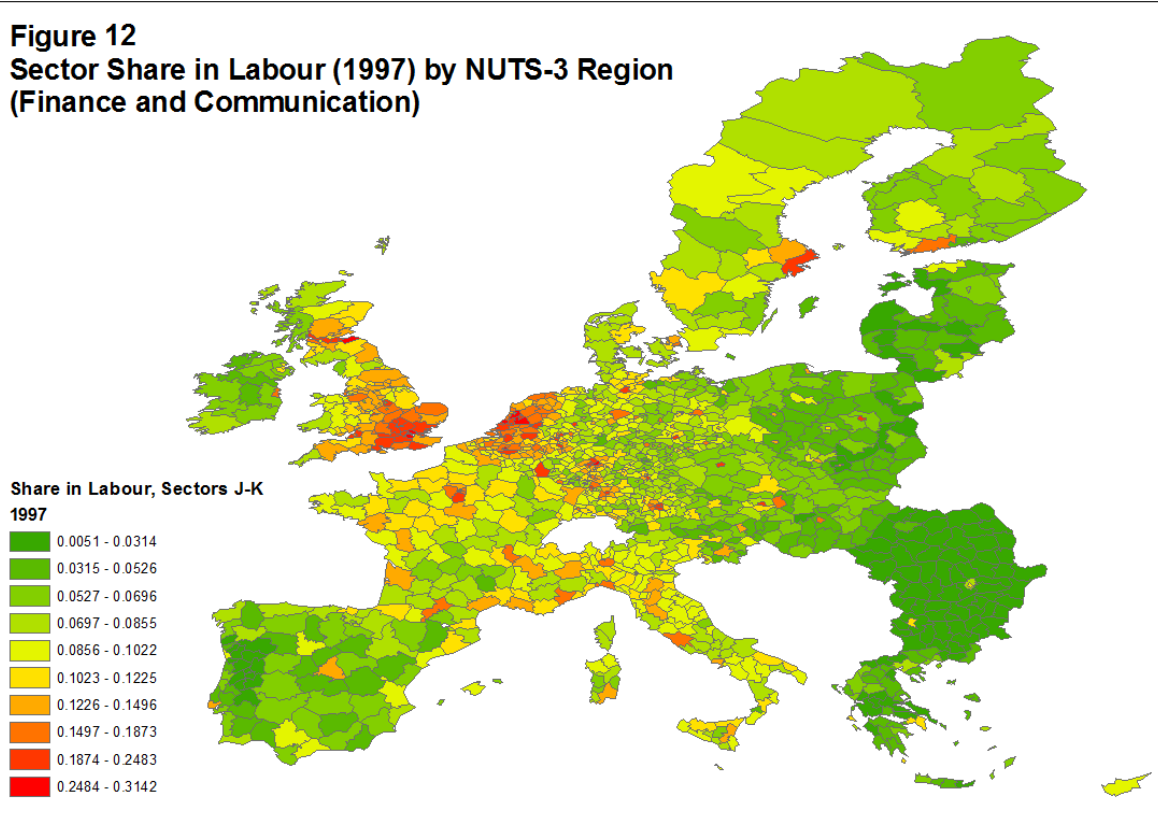
**Figure 10**  
**Sector Share in Labour (1997) by NUTS-3 Region**  
**(Wholesale, Logistic, and Catering)**



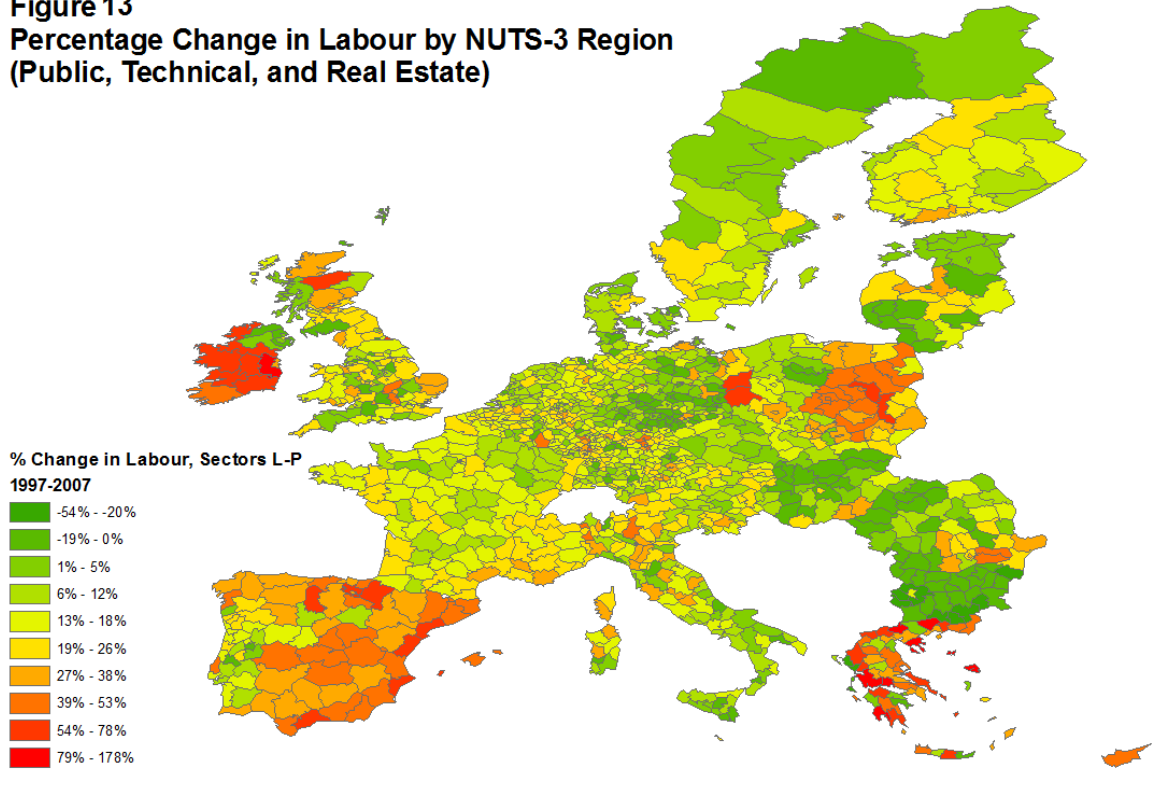
**Figure 11**  
**Percentage Change in Labour by NUTS-3 Region**  
**(Finance and Communication)**



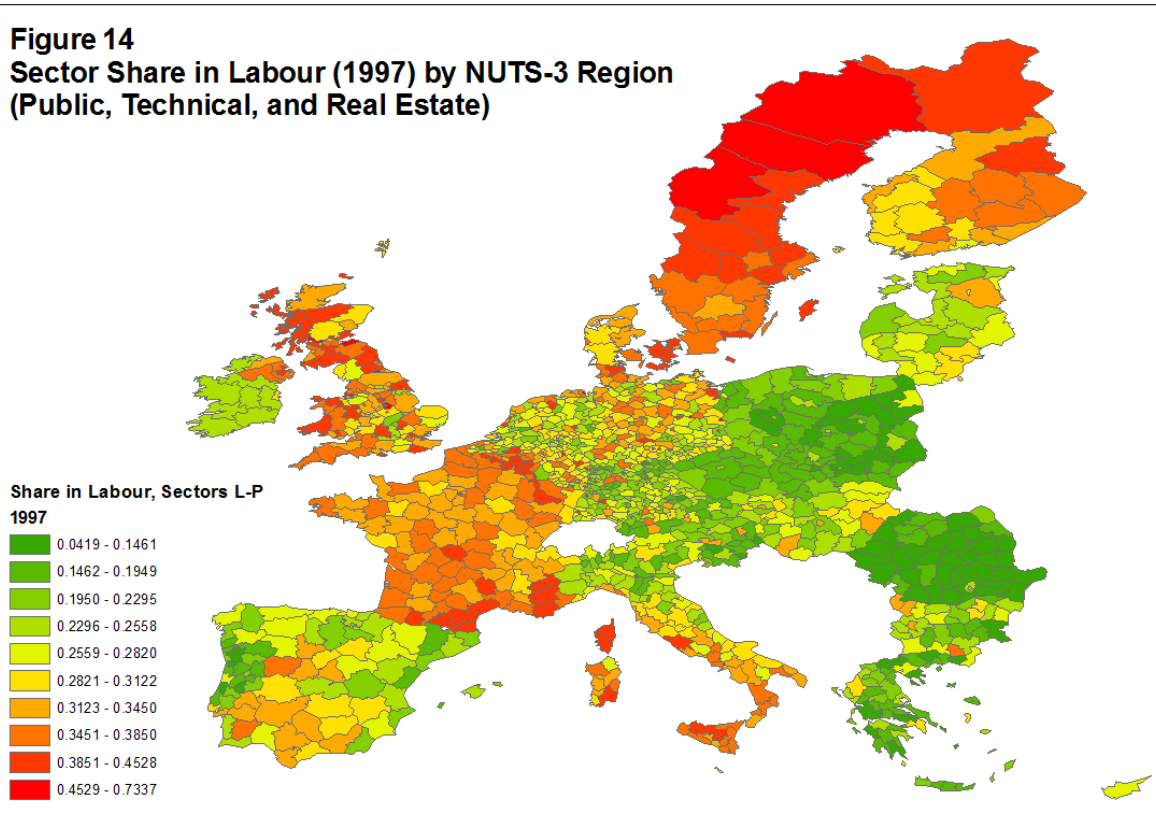
**Figure 12**  
**Sector Share in Labour (1997) by NUTS-3 Region**  
**(Finance and Communication)**



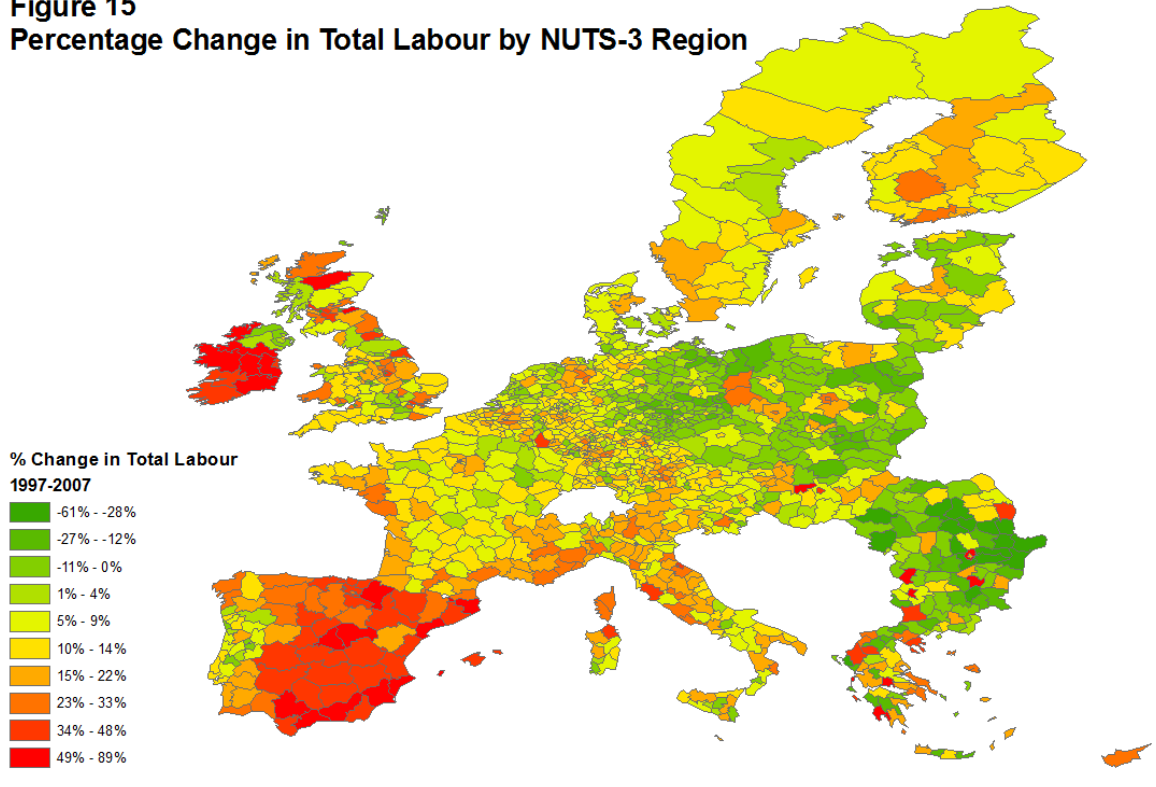
**Figure 13**  
**Percentage Change in Labour by NUTS-3 Region**  
**(Public, Technical, and Real Estate)**



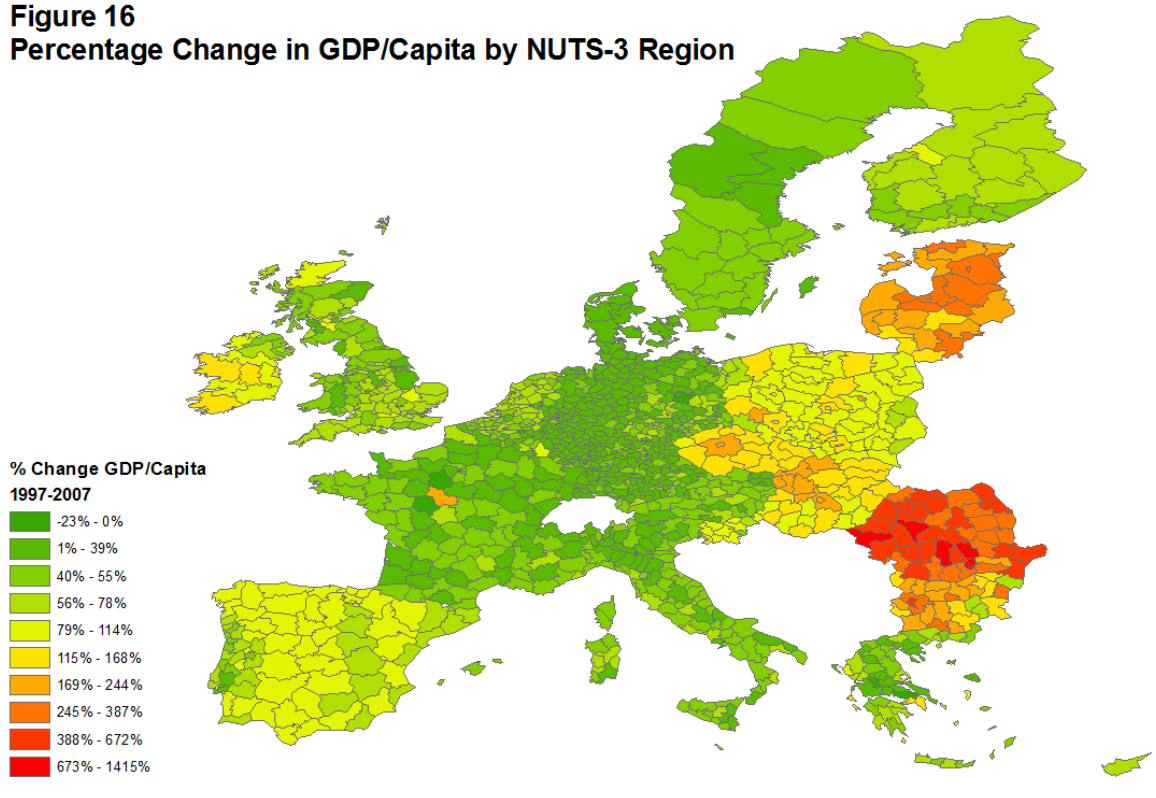
**Figure 14**  
**Sector Share in Labour (1997) by NUTS-3 Region**  
**(Public, Technical, and Real Estate)**



**Figure 15**  
**Percentage Change in Total Labour by NUTS-3 Region**



**Figure 16**  
**Percentage Change in GDP/Capita by NUTS-3 Region**





## APPENDIX 2: Variables

### A. Estimation of Labour Force for Missing Years

Countries that exhibited missing data may be summarised in table 11.

*Table 12: Countries with missing data*

Country	NACE Code					
	Total	A-B	C-E	G-I	J-K	L-P
Belgium	1997-99 2 NUTS	1997-99 2 NUTS	1997-99 2 NUTS	1997-99 2 NUTS	1997-99 2 NUTS	1997-99 2 NUTS
Greece	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS
Spain	1997-99 8 NUTS	1997-99 8 NUTS	1997-99 8 NUTS	1997-99 8 NUTS	1997-99 8 NUTS	1997-99 8 NUTS
Hungary	Complete	Complete	1997-98 all NUTS	1997-98 all NUTS	1997-98 all NUTS	1997-98 all NUTS
Malta	1997-98 all NUTS	1997-98 all NUTS	1997-98 all NUTS	1997-98 all NUTS	1997-98 all NUTS	1997-98 all NUTS
Netherlands	1997-2000 all NUTS	1997-2000 all NUTS	1997-2000 all NUTS	1997-2000 all NUTS	1997-2000 all NUTS	1997-2000 all NUTS
Poland	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS
Romania	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS	1997-99 all NUTS
UK (Excluding N. Ireland and parts of Scotland)	1997-2001 all NUTS	1997-2001 all NUTS	1997-2001 all NUTS	1997-2001 all NUTS	1997-2001 all NUTS	1997-2001 all NUTS

A missing year ( $t-1$ ) for the above countries were thus extrapolated using an estimate of a linear function via ordinary least squares and then by assigning a value on the estimated line using all  $n$  proceeding years. This  $n$  number of years was then used for all years before the first missing year. By using only one value for  $n$  it was hoped that there would be some consistency in estimating all missing years. This linear estimation, where  $t$  is a given year, may be formalised as:

$$L = a + bt$$

Where:

$$b = \frac{n \sum(Lt) - \sum L \sum t}{n \sum(t^2) - (\sum t)^2}$$

$$a = \frac{\sum L - b \sum t}{n}$$

## **B. Estimation of Missing Values for North of Scotland and Northern Ireland**

### **Consolidation of NUTS-3 areas**

For Scotland, data was taken from the NOMIS database. The areas of study conducted by NOMIS involve the use of local authorities as opposed to NUTS-3. As a result, many of these geographical areas either encompass many NUTS-3 or overlap them. Furthermore, NOMIS uses a different set of industrial sectors, which may or may not overlap with the NACE codes used by Eurostat.

*UKM50: Aberdeen City and Aberdeenshire*

Data was taken from the NOMIS database for Aberdeen City Council and Aberdeenshire County Council for the years 1997 and 2007, and added together to obtain a grand total.

$$\text{Aberdeen City} + \text{Aberdeenshire} = \text{UKM50}$$

Total workforce was taken from NOMIS for 1997.

*UKM61: Caithness & Sutherland and Ross & Comarty*

*UKM62: Inverness & Nairn and Moray, Badenoch & Strathspey*

*UKM63: Lochaber, Skye & Lochalsh, Arran & Cumbrae and Argyll & Bute*

Two county councils, Highland and Moray, were first added together to produce a grand total. Then, using 2007 data from Eurostat, proportions were calculated for total employment. This fraction was in turn used to estimate the total amount of the workforce in each NUTS-3.

$$\text{Highland} + \text{Moray} = (0.267)\text{UKM61} + (0.472)\text{UKM62} + (0.261)\text{UKM63}$$

*UKM64: Eilean Siar (Western Isles)*

*UKM65: Orkney Islands*

*UKM66: Shetland Islands*

These NUTS-3s are identical to their respective county councils.

*UKN01: Belfast*

*UKN02: Outer Belfast*

*UKN03: East of Northern Ireland*

*UKN04: North of Northern Ireland*

*UKN05: West and South of Northern Ireland*

The grand totals for 2007 were taken from the Northern Ireland Neighbourhood Information Service (NINIS) by adding up each local authority in order to obtain their respective NUTS-3 areas.

### **Estimation of Labour Force by Sector**

Both NOMIS and NINIS use a different classification methodology than Eurostat. In order to give some consistency for EU-wide data, estimations of the labour force for each industry

were made by using the proportions of the labour force for a given year. This was performed using the following methodology given the data available:

*For all sectors in the north of Scotland:*

$$L_{ic}^{1997} = \frac{L_{ic}^{2007}}{L_i^{2007}} \times L_i^{1997}$$

*For all sectors in Northern Ireland:*

$$L_{ic}^{2007} = \frac{L_{ic}^{2002}}{L_i^{2002}} \times L_i^{2007}$$

*For all sectors in Northern Ireland except for NACE codes A-B:*

$$L_{ic}^{1997} = \frac{L_{ic}^{2002}}{L_i^{2002}} \times L_i^{1997}$$

As proportions of industry distribution are kept fixed for the above estimation, and as this paper has focused on the raw increase in labour numbers, it is thus inferred that such estimations will only adversely affect the accuracy of inter-regional studies of labour migration within the north of Scotland and Northern Ireland, rather than that of intra-regional studies of the EU and these two areas.

### **C. Construction of Language Dummy ‘Pots’**

The four language dummies were constructed by assigning each language into a specific ‘pot’. Each was assigned in such a way that no two languages can border each other geographically, e.g. German and Polish. This would give rise to the false impression of two countries speaking the same language. Furthermore, it was assumed that, given one language, if every NUTS-3 area of that language speaks an additional language fluently (‘fluent’ in this case means more than 50% of the population is fluent speaker), then the first language is logically eliminated. For example, every NUTS-3 that speaks Dutch also speaks English, and thus Dutch was eliminated. This list of eliminated languages, along with the reasoning, may be seen in table 12. The pots were thus constructed as follows:

Pot 1: Spanish, Russian, English

Pot 2: Czech, Bulgarian, Finnish, Italian

Pot 3: Hungarian, Romanian, French

Pot 4: Polish, Portuguese, Slovene, Greek

Lastly, it was important to keep at least one EU10 language in each pot, for the sake of avoiding an ill-defined matrix. These are marked as underlined. The information of language fluency levels in each geographical area was obtained from the European Commission (2006).



Table 13: List of Eliminated Languages and Reasoning

Language	Reason for Exclusion
Dutch	All regions have high fluency of <i>English</i>
Danish	All regions have high fluency of <i>English</i>
Estonian	All regions have high fluency of <i>Russian</i>
German	All regions have high fluency of <i>English</i>
Irish	Low level of fluency
Latvian	All regions have high fluency of <i>Russian</i>
Lithuanian	All regions have high fluency of <i>Russian</i>
Maltese	All regions have high fluency of <i>English</i>
Slovak	Mutually intelligible with <i>Czech</i>
Swedish	All regions have high fluency of <i>English</i>