6. Econometric Analysis of Influences on Agglomeration and Regional Specialization

The current chapter 6 will analyse by means of regression analysis the part of the hypothesis about the influences working on industry agglomeration and those relevant for regional specialization in Hungary. The results obtained so far for agglomeration and regional development in Hungary point to a trade-off between the overall catching-up process at the country-level due to increasing trade integration under the Europe agreement, on the one hand, and the increase of regional disparities in the pre- and post-accession phase, as shown by the results from regional specialization and other indicators, on the other hand.

In view of the factors influencing agglomeration and regional specialization processes, two regressions have been generated after a longer modelling process. One will be on the question: Which influences worked on concentration of manufacturing industries in Hungary? This will be the subject of section 6.1 and subsections. The other regression will deal with the question: Which influences shaped the degree of regional specialization in Hungary? That will be the subject of section 6.2 and subsections. The analytic tool which has been used is the programme Eviews 6.0. The two regressions have been formed by way of panel estimation in order to have a sufficient number of observations as base.

6.1 The question for regression 1 on agglomeration

This research analyses agglomeration processes in the manufacturing sectors in Hungary during proceeding European integration, notably during the pre-accession phase governed by the trade provisions of the Europe agreement as well as in the post-accession phase five years after Eastern enlargement of the EU.

The question analysed by regression 1 regarding agglomeration is the following:

Which influences were relevant for an industry’s concentration in Hungary in context with industry characteristics and European integration?

As dependent variable for this regression for modelling agglomeration, manufacturing industry concentration as measured by the Krugman concentration index was chosen. This index - calculated based on employment figures - proved to be the most conservative estimate of concentration, as was shown in Chapter 4 in the
comparison of the 6 concentration measures (section 4.9). The industries were the 8 manufacturing industries of the Hungarian classification (based on TEÁOR and NACE) which were used throughout this study and for which data were available from HCSO since 1992 up to 2008:

- **Food**, beverages and tobacco;
- **Textiles**, wearing apparel, leather and fur products;
- **Wood**, paper and printing, publishing;
- **Chemicals** and chemical products;
- Other non-metallic **mineral** products;
- **Basic metals** and fabricated metal products;
- **Machinery and equipment** (n.e.c., electrical and optical equipment, transport equipment);
- **Other** manufacturing industries/ recycling.

The observation period for this regression was chosen as 1993 to 2008, as the data for 1992 proved to be somewhat blurred with respect to certain industry characteristics, which may be due to the break-down of the communist system and the transition process to a market economy still going on with a certain vigour.

The number of observations for regression 1 was 120\(^{155}\). This is in the normal range for regression analysis on CEECs due to the short period of availability of reliable data (e.g. Hildebrandt & Wörz (2004)).

### 6.1.1 Selection of independent variables and description of data

As the regression model, the following linear panel regression was formulated:

\[
\text{CONC}_{i,t} = a_0 + a_1 \text{Expr}_{i,t} + a_2 \text{Impr}_{i,t} + a_3 \text{FDI}_{i,t} + a_4 \text{Prod}_{i,t} + a_5 \text{Lent}_{i,t} \\
+ a_6 \text{Scal}_{i} + a_7 \text{Hitec}_{i} + a_8 \text{Medtec}_{i} + a_9 \text{Low}_{i} + a_{10} \text{Hiw}_{i} + \text{resid} \varepsilon_{i,t-1}
\]

In this section, the variables will be explained, the data for the variables which entered into the regression 1 shall be described, and an expectation for the signs of the coefficients of the independent variables shall be given in light of relevant theories.

\(\text{CONC}_{i,t}\) is the dependent variable. It stands for the degree of agglomeration of a manufacturing industry \(i\) at time \(t\), where an industry’s concentration is measured by the Krugman concentration index (4 digits after the comma).

---

\(^{155}\) 120 observations: concentration for 8 industries over 15 years (1994 to 2008), as 1994 became the start year due to the AR correction term which was lagged by 1 year.
The dependent variable had a mean of 0.4189 and a standard deviation of 0.125. The data calculated for the Krugman concentration index per industry have been described in detail and evaluated in the relevant theoretical and empirical context in chapter 4, section 4.3.

The independent variables:

\( C \) The constant was included in this linear multiple regression.

\( \text{Expri}_{it} \) This first variable for integration with the EU was an industry’s exports to the EU in that year, in million HUF. The EU as export destination was taken in its contemporary form for each year, i.e. EU-25 for 2004 until 2006, EU-27 for 2007 and 2008. The 50 relevant SITC categories were converted systematically into the 8 industries according to a conversion table developed by myself and based on the detailed industry specification in SITC and TEÁOR statistics (the latter has a NACE correspondence). This allowed to then put exports of industry i to the EU over output of industry i for a given year, in million HUF - thus any inflation built in over the years would disappear. The resulting percentage, e.g. 0.120, was multiplied by 100 and then entered into the regression (as 12.0).

It would also have been conceivable to use the logarithm of the value of an industry’s exports to the EU, or alternatively to use the share of exports to the EU in total exports to the world. The choice of exports to EU over output has been made as it is an even more concise parameter for integration with the EU than those other measures would have been.

The data for this variable showed a rise of the export rate to the EU (over output) for all industries over the years 1993 to 2008. Certain industries had low export rates, such as wood and paper (33% in 2008), while others such as machinery and equipment (including transport equipment) had high export rates (up to over 85%). A special phenomenon was going on in the textiles industry which showed export rates to the EU of up to 126% of national output. This was due to the outward processing trade with EU countries (see Badone & Sdogati 1997 on this) which was strongest in 2001 and declined subsequently, as Hungary lost its productivity advantage to other CEECs and to the Far East, mainly China.

Traditional trade theory predicts that industries are concentrated in the countries with a comparative advantage. In New Trade Theory, which includes product differentiation, a reduction in trade costs reduces concentration, as the home bias - originally in large countries - is removed (as in Helpman & Krugman (1985)). As increasing European integration went along with the reduction of trade costs,
and as the share of Hungary’s exports to the EU rose steadily during the process, the expected sign for the coefficient of the export variable with respect to concentration is negative.

\( \text{Impr}_{1,t} \)  

The second integration variable chosen is an industry’s imports from the EU in a year, in million HUF. This was converted from SITC statistics analogous to exports from the EU with the same conversion table, then set over output of the industry for that year; the resulting percentage entered multiplied by 100.

While an industry’s imports could be sourced worldwide, it is a fact that the share of imports from the EU in total Hungarian imports increased significantly and almost in parallel to exports over the period 1993 to 2008 (see graph in Fig. 21 in section 3.7 of Chapter 3). As this research is interested in the effects of integration with the EU, imports from the EU were chosen here as variable. That import rate (over output) was between 29.7% for minerals and 85.3% for metals and metal products in 2008. Again, the textiles industry as well as other industries including recycling showed percentages above 100% of national output due to outward processing - the latter is a hint to a new division of labour among West European and the new Central and East European member states after Eastern enlargement.

For the same theoretical reasons as given for \( \text{Expri}_{1,t} \), the expected sign for the coefficient of the import variable is the same as for \( \text{Expri}_{1,t} \), that is negative: more imports from the EU are expected to lower the degree of agglomeration in Hungary and encourage the dispersion in space.

\( \text{FDI}_{1,t} \)  

As a further variable for integration with the EU, foreign direct investment (FDI) in an industry was entered into the regression. This was taken as the FDI stock (from world) in million HUF present in an industry at time \( t \). The variable was entered taking the logarithm (lg) of this value (base 10). To take FDI stock (from world) is seen as appropriate here, as a very large proportion of that FDI in manufacturing - over 50% - originated from EU countries, most prominently among them Germany. Due to the sensitivity of FDI data for the Hungarian authorities, it was not possible to get chronological data by industry according to the countries of origin of the FDI.

The stock of FDI in Hungarian manufacturing industries rose significantly during the period 1992 to 2008\(^{156}\). This was shown in detail for the 8 sectors in chapter 3, section 3.8. FDI in Hungarian manufacturing overall rose from 332 billion HUF in 1993 to 5,311 billion HUF in 2008, with the machinery and equipment sector

\(^{156}\) For the FDI stock in 2007 and 2008, provisional data had to be used, as the final data were scheduled to be published after the time of writing.
holding the lion share. After taking the logarithm of the FDI stock (in million HUF), the data ranged from 3.7 for other industries and recycling in 1993 to a high of 6.44 for machinery and equipment in 2008.

The expected sign for the coefficient of the FDI variable is negative: more FDI stock in an industry is expected to lower the degree of agglomeration in Hungary and facilitate the dispersion in space and the location in many possible sites.

In context with FDI and given that the privatisation of the economy had a substantial influence on FDI inflows in Hungary during a certain period (Kalotay & Hunya 2000), I also thought about introducing a privatisation variable as a dummy, taking the value one after the sector was included into the state privatisation scheme. But while information on the Hungarian privatisation scheme was readily available in the mid 1990s (Hantke 1995), this is no longer the case now when the privatisation scheme of the transition to a market economy has been concluded and Hungary is a full member state of the EU.

\[ \text{Prod}_{it} \]

This variable was entered in order to model the effect of labour productivity in an industry on its concentration\(^{158}\). \text{Prod}_{it} is the ratio of output of the sector (in million HUF, at constant prices of 1992\(^{159}\)) over employees of the sector at time \( t \).

Labour productivity in Hungary improved significantly by about the double of the initial level on average for the manufacturing sector over the period 1993 to 2008. Productivity in the chemicals sector tripled, while that in the machinery and equipment sector increased fivefold. Sectors with below average increases were the wood, pulp and paper as well as the recycling and other industries sectors. This overall improvement was a clear sign of the restructuring process during the transition phase initially, until about 1995, and also due to the modernisation of production processes after privatisation and the large inflows of FDI in the latter part of the observation period. The highest productivity prevailed in the chemicals industry, the lowest in the textiles industry.

\(^{157}\) During my research in Budapest in the year 1994 for my publication Hantke (1995), officials of various governmental institutions readily provided information for potential investors during the ongoing privatisation.

\(^{158}\) The OECD defines labour productivity as “the ratio of a volume measure of output to a volume measure of input”. Volume measures of output are normally gross value added, expressed at constant prices, i.e. adjusted for inflation, measures of input include the number of employees (OECD 2002).

\(^{159}\) This producer price index is published by the HCSO as an index with previous year =100. In order to take out inflation, I corrected this by holding 1992=100 constant and using a producer price index in terms of prices of 1992 as correction.
In NEG models (Krugman 1991a, Ludema & Wooton 1997), industry tends to concentrate where labour productivity is higher than elsewhere. Therefore, the expected sign for the coefficient of Prod\textsubscript{it} is positive.

**Lent\textsubscript{it}** This is the number of enterprises\textsuperscript{160} of an industry in year t, of which the logarithm was taken. This variable was entered into the model to take account of the third dimension of concentration: where in space it took place. The number of enterprises - as a geographic illustration of the third dimension of concentration, namely where in space - has been described in section 3.4.3 of Chapter 3. This variable shall indirectly reflect the role of “history” - as described in NEG models - for determining industry location and concentration in certain areas, i.e. due to the past presence of certain industries in a production site, new industry is more likely to locate nearby than elsewhere. Further, this variable also implies to some extent the Hungarian government policy of industrial parks under the Széchenyi plan (see section 3.2.4 in Chapter 3) which encouraged industry to locate in these parks which were spread over the country (although two thirds are in the Western part and one third in the less developed Eastern part of the country).

The data for the number of enterprises per sector were lowest for the metals and metal products industry, and for the chemical industry, around 3,500 in 1993, and highest for the textiles industry and machinery and equipment, around 18,000 each. Comparing the year 1995, which showed the highest number of enterprises in all industries, with the end year 2008, a certain consolidation process took place in manufacturing. This process was most pronounced in the textiles industry which reduced active enterprises to less than 40% of that figure. In addition, a low number of enterprises would indicate a high degree of concentration. Therefore, the expected sign for this variable is negative.

To include a government variable such as tax benefits, though perhaps desirable, was not feasible due to language barriers in accessing Hungarian legislative texts; moreover, tax benefits most likely would not have been differentiated by manufacturing industry.

**Scal\textsubscript{i}** This is a dummy variable for the presence and importance of economies of scale in the industry. It takes the value 1 if economies of scale play a role and 0 otherwise. The classification of the OECD (OECD 1994) was followed, i.e. scale intensive were 3 industries in the Hungarian data set, namely wood, paper and printing, chemicals and chemical products, as well as metals and metal products, while the other 5 industries formed the reference category.

\textsuperscript{160} Enterprises in the statistics of HCSO: active corporations with 4 or more employees.
In a world with transportation costs, as analysed by the models of New Trade Theory, big countries gain a comparatively larger share in industries where product differentiation and internal or external economies of scale are important (Helpman & Krugman 1985). New Trade Theory postulates that even in the absence of differences in factor endowments, scale economies can induce specialization among countries (regions in this study) and thus affect relative concentration. Further, industries with higher economies of scale may tend to concentrate in relatively central locations or agglomerations (Krugman 1980; Midelfart-Knarvik et al. 2000). Therefore, the expected sign for the SCA coefficient is positive.

$\text{Hiteci, Medtec}_i$ These are dummy variables accounting for the importance and level of technology in an industry. Hitec stands for high technology, Medtec for medium technology respectively. They take the value 1 if the industry belongs to that category and 0 otherwise. Again, the classification of the OECD (OECD 1994) was followed. The machinery and equipment industry was categorized as high tech (due to electrical and optical equipment, computer equipment); 3 industries were medium tech, namely chemicals and chemical products, metal and metal products, and other manufacturing industries/recycling. The remaining industries formed the reference category (low tech). The OECD classification was judged more accurate and commonly acceptable than the estimates for the importance of scale economies by Forslid et al. (2002, Table 5) which were based on input/output matrices for the year 1992 and for the EEA due to the inappropriate country scope and reference year.

Differences in productivity levels between industries are implied to comprise technological differences and thus comparative advantages, which are at the heart of traditional Ricardian trade theory. Large differences in technology levels between industries are expected to have a positive influence on the concentration of an industry. Therefore, the expected signs for the coefficients of the Hiteci and Medtec variables are positive.

$\text{Low}_i, \text{Hiw}_i$ These are dummy variables accounting for the wage level in an industry. Low for low wage industries, Hiw for high wage industries. They take the value 1 if the industry belongs to that category and 0 otherwise. Here again, the classification of the OECD (OECD 1994) was followed. The chemicals and chemical products industry was high wage, while 3 industries were low wage, namely textiles, food, beverages and tobacco, and other industries and recycling. The remaining industries formed the reference category of being medium wage level.

---

161 The EEA (European Economic Area) comprised the EU-15 plus the EFTA countries, thus no country from Central and Eastern Europe at all.
According to the NEG model by Krugman (1991a), higher real wages prevail in the agglomeration and lower wages in the periphery. Therefore, the expected sign for the coefficient of Low\textsubscript{i} is negative, while the expected sign for the coefficient of Hiw\textsubscript{i} is positive.

**resid e\textsubscript{t-1}** This AR correction was necessary as in the original regression, errors were auto-correlated as indicated by the value for the Durbin Watson statistic which was lower than the low-sided boundary for that value (0.7236). Consequently, the 8 graphs of the residuals prior to that correction looked very similar to the graphs of the concentration index for each industry.

The significance of the independent variables was not altered by the AR-correction, i.e. those which were significant before remained significant, and those that were not remained insignificant. For one variable, the sign of the coefficient changed though\textsuperscript{162}. The value for the adjusted R-squared increased by this correction (as did the Durbin-Watson statistic).

Prior to including the AR correction, it had also been tried to include a linear trend into the regression; this remained insignificant, however, and did not alter the low result of the Durbin-Watson test. So that idea was dropped.

At this point, some lines shall be said about the modelling of trade costs. In most NEG models, the level of trade costs - falling during integration - plays a prominent role in explaining agglomeration processes. I therefore would have liked to include an explanatory variable for the level of tariffs and quotas under the Europe agreement. There were no data available, however, differentiated by industries and for all years, as they were kept secret by the EU administration.

Further, it is acknowledged that NTBs play an even more important role in shaping the level of bilateral trade and economic integration. Regrettably, there are no quantitative estimates on NTBs available for Hungary; just for the EU-15, a series of studies for certain industries was carried out for the EU Commission in context with evaluating the Single Market programme in the mid-1990s.\textsuperscript{163} Therefore, NTBs could not be included here either.

\textsuperscript{162} This variable was Lent, which had a negative sign for the coefficient prior to the AR correction, and a positive sign afterwards.

Finally, the time spent in passing borders (waiting time at border controls, and the filling in of customs formalities) could be conceivable as a measure of trade costs. 

Bröcker (1998) estimated this at 100 minutes and 60 minutes for the mid-1990s, and Niebuhr & Schlitte (2008) modelled integration by reducing these time penalties. As these are mere estimates or assumptions, they are not deemed appropriate, given the modernisation of border crossing points in accession countries, including Hungary, and the harmonisation of customs formalities which took place financed by EU-funds during the pre-accession period (i.e. under the Europe agreement). Other empirical estimates of trade costs worldwide have been presented in Section 2.2.3 (the latter part) of Chapter 2 but were not suitable to serve as explanatory variable in this model either.

6.1.2 Results of regression 1 - Agglomeration

The results of regression 1 regarding agglomeration shall be presented now. Table 25 shows the coefficients for the constant and the explanatory variables as well as the error correction term; the p-values are shown beneath in italics and parenthesis. The last column indicates whether the coefficient had the expected sign as well as the significance level.

Regression 1 as presented had an $R^2$ of 88.7% and an adjusted $R^2$ of 87.6%. The F-statistic was sufficiently high at 77.61. The constant had a positive coefficient and was significant at the 5% level. The export rate to the EU was highly significant, yet the coefficient was positive unlike the expectation. The coefficient of the import rate to the EU had the expected negative sign, but was not significant at all. The coefficient of the logarithm of the FDI stock was not significant, but had the expected negative sign. Labour productivity was highly significant, and the coefficient had the expected sign. The number of enterprises was not significant, but the coefficient had the expected sign.

The presence of scale economies was very highly significant, yet the coefficient did not have the expected positive sign. The fact that an industry was high tech was very highly significant, yet the coefficient did not have the expected sign. Being medium tech was not significant, nor did the coefficient have the expected sign. The coefficient for being a low-wage sector had the expected negative sign and was very highly significant. Being a high wage sector was highly significant and the coefficient had the expected positive sign. Finally, the residual error correction term was very highly significant.
Table 25: Results of regression 1 on manufacturing industry agglomeration in Hungary, 1993 to 2008

<table>
<thead>
<tr>
<th>Variable</th>
<th>for</th>
<th>Coefficient (p-value in italics)</th>
<th>had expected sign/ significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Constant</td>
<td>0.758351 (0.0208)</td>
<td>*</td>
</tr>
<tr>
<td>Exp_{i,t}</td>
<td>EU Exports/output ratio</td>
<td>0.000721 (0.0057)</td>
<td>**</td>
</tr>
<tr>
<td>Impr_{i,t}</td>
<td>EU Imports/output ratio</td>
<td>-1.29E-05 (0.9484)</td>
<td>✓</td>
</tr>
<tr>
<td>FDI_{i,t}</td>
<td>Lg of FDI stock</td>
<td>-0.00714 (0.7151)</td>
<td>✓</td>
</tr>
<tr>
<td>Prod_{i,t}</td>
<td>labour productivity</td>
<td>-0.011276 (0.0017)</td>
<td>**</td>
</tr>
<tr>
<td>Lent_{i,t}</td>
<td>Lg number of enterprises</td>
<td>-0.005097 (0.9442)</td>
<td>✓</td>
</tr>
<tr>
<td>Scal_{i}</td>
<td>Scale economies</td>
<td>-0.229913 (0.0001)</td>
<td>***</td>
</tr>
<tr>
<td>Hitec_{i}</td>
<td>High tech sector</td>
<td>-0.377893 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>Medtec_{i}</td>
<td>Medium tech sector</td>
<td>-0.031706 (0.072)</td>
<td></td>
</tr>
<tr>
<td>Low_{i}</td>
<td>Low wage sector</td>
<td>-0.322779 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>Hiw_{i}</td>
<td>High wage sector</td>
<td>0.137256 (0.0004)</td>
<td>***</td>
</tr>
<tr>
<td>Resid e_{k-1}</td>
<td>AR-correction factor</td>
<td>0.641621 (0.0000)</td>
<td>***</td>
</tr>
</tbody>
</table>

R²           | 0.887                      |
adjusted R²  | 0.876                      |
F-statistic  | 77.61                       |
Number of observations | 120                      |

* Significant (at the 5% level)
** Highly significant (at the 1% level)
*** Very highly significant (at the 0.1% level)

Source: Own calculations.
6.1.3 Evaluation of the results in view of the hypothesis, NEG models and previous empirical studies

The question for regression 1 on agglomeration was: Which influences were relevant for an industry’s concentration in Hungary in context with industry characteristics and European integration? The results show that of the integration variables, an industry’s export rate to EU over its output as well as the FDI stock, the exports to the EU had a very significant influence on the concentration of that industry in Hungary during the 1993 to 2008 period. The coefficient was positive, that is a higher export rate to the EU on average increased manufacturing concentration. This showed that the export reorientation to the EU turns out to be an important factor for manufacturing industry agglomeration in Hungary. While the FDI stock had the expected dispersing influence on concentration, the coefficient was not significant in this case. The import rate from the EU had the expected sign, but remained insignificant. The importance of trade integration in the sense of export orientation verifies the predictions by the model of Livas-Elizondo & Krugman (1996) which predicted agglomeration to form within the multi-region country in the course of integration processes.

Of the industry characteristics, labour productivity was highly significant for the concentration, and the coefficient had the expected sign. While the coefficient for the number of enterprises in a sector had the expected sign, it was not significant for concentration in this data set. Of the remaining industry characteristics all but one were very significant for the degree of industry concentration: the presence of scale economies - in the chemicals and metal products industries - was very highly significant; but contrary to what would be expected from NEG theories, this reduced industry concentration by -0.2299 on average, as compared with the reference category of no scale economies, controlling for other variables. Further, being a high tech sector was very significant. Contrary to what would be expected from NEG theories, this rather encouraged industry dispersion on average, however, as indicated by the negative coefficient. Being a high tech sector on average reduced concentration by -0.3778 as compared with the reference category low tech - food, textiles, minerals and paper industries -, controlling for other variables. Being a medium-tech sector reduced concentration as expected, but the coefficient just fell out of the 5% significance level.

Being a low wage sector had a strong centrifugal influence on industry location, spreading out the industries concerned more over the country: the textiles industry, food, beverages and tobacco, as well as other industries and recycling. This corresponds to the NEG theories, where higher real wages reinforce industry agglomeration (Krugman 1991a; Ludema & Wooton 1997; Livas-Elizondo & Krugman 1996). Being a low wage sector in Hungary reduced concentration
by -0.3227 as compared with the reference category medium-wage sectors, controlling for other variables. On the other hand, being a high-wage sector had a strong concentrating influence on the industry concerned, as would be predicted by NEG theories. In fact, the chemical industry had high values for the Krugman concentration index in Hungary over the entire period (see section 4.3.2 of Chapter 4). Being a high wage sector on average increased concentration by 0.1372 as compared with the reference category medium wage sectors, controlling for other variables.

The results of regression 1 shall be put now into context with the relevant literature. Some other authors have run regressions on industry concentration in EU or Central and East Europe countries previously. In their regression for relative industry concentration in 10 countries based on a modified form of the Hoover-Balassa index, Hildebrandt & Wörz (2004) also found a negative sign for the coefficient of scale economies, as in my regression 1, both when the concentration (dependent variable) was calculated based on output as well as when it was calculated with employment figures; but that scale variable was not significant in their data. They found a significant influence for the FDI stock, but with a positive sign for the coefficient (unlike in my regression 1). Their technology variable - which was taken as a difference from the mean - had a positive coefficient and was very highly significant for concentration, as in my regression 1. They had included a quadratic time trend to improve the fit of that regression, though, which does not make sense intuitively with respect to the research question. In their regressions for individual industries over the 10 countries, the exports to the EU (exports to the EU over exports to the world in percent) were highly significant for 4 out of 11 industries, the electronics, wood, paper, and food industries. The imports from the EU variable was not significant for the vast majority of industries, except for machinery and wood. Their regressions per industry were based on only 80 observations and had an $R^2$ around 0.20 for most industries, except for electronics with 0.81.

Brülhart & Torstensson (1998) compared industry Gini coefficients with industry centrality indices for the EU-15 and found a positive correlation between scale economies and industry bias towards the central EU in both 1980 and 1990. Midelfart-Knarvik et al. (2000) found that unskilled labour-intensive industries have become more concentrated in the EU from 1970 to 1997, but usually in peripheral low wage countries (thus, this would be a dispersion regarding the third dimension of concentration - where in space: centre versus periphery). During the same period, a number of high and medium tech industries have become more dispersed, i.e. the sign of the coefficient would correspond to the negative sign found in my regression 1.
Finally, in their analysis of concentration effects in context with the Single Market in the EU-15, Aiginger & Pfaffermayr (2004) conducted a panel regression on changes of entropy from 1985 to 1992, and from 1993 to 1998, choosing a set of industry variables. Their results showed that skill-intensive industries dispersed faster on average, and that highly globalized (by their exports to world) and capital-intensive industries exhibited a tendency of increasing concentration significantly in the second period only.

Thus, some of the few results of concentration regressions went in the same direction as the results of my regression 1: a negative sign for a scale coefficient, an imports variable not being significant, exports positively correlated with concentration and partly significant and low-wage industries positively correlated with concentration.

The next section 6.2 and subsections will look at regions instead of industries, analysing the subject of regional specialization by means of regression analysis.

## 6.2 Regression 2 on regional specialization

The part of the hypothesis about the development of regional specialization in Hungary during the period 1993 to 2008 will now be analysed by means of regression analysis. In particular, influences which were relevant for the specialization of a region in context of European integration and region-specific factors playing a role for the specialization of a region shall be the subject of analysis in this part. These questions shall be analysed in regression 2 by means of a panel analysis spanning over the 20 Hungarian regions and the period 1993 to 2008 (300 observations164).

The question of regression 2 - on regional specialization - addresses the first of the three dimensions of agglomeration: How strong was the effect of the ongoing manufacturing industry agglomeration on regional specialization in Hungary over the period 1993 to 2008? It does not address the geographic dimension where in space did it happen (due to the panel nature of the regression), nor which industries were involved.

---

164 The number of observations was reduced from 320 to 300 due to an AR-correction factor of the error term which shortened the period to 1994 to 2008 (15 years).
### 6.2.1 Descriptive statistics and trend regression

Before doing the entire regression analysis, some descriptive statistics and analysis of the dependent variable, specialization of a region as measured by the Krugman specialization index, shall be presented. Specialization was used in regression 2 over the period 1993 to 2008. The variable had a mean of 0.3831 and a standard deviation of 0.1150. The national average for regional specialization was calculated for 1993 - which was 0.3824 - and for 2008 respectively - which was 0.3790. Thus, on average, regional specialization declined by -0.0034, which is a rather slight change overall. As Chapter 5, section 5.3.1, showed, there had been a stronger rise to a peak around 1999 in between, and a second slighter peak around 2006.

**Figure 45:** Regional specialization in Hungary relative to the national average, 1993 and 2008 in comparison, Krugman specialization index based on manufacturing employment

Source: Own calculations; own presentation.

For each region, the regional specialization relative to the national average was calculated for the start year 1993 and the end year 2008. As the bar diagram in Figure 45 shows, the results are mixed. While 10 regions had a lower than average specialization in 1993 and 10 a higher than average, the result for 2008 was also 10 to 10. Comparing the direction of deviation from the national average, 9 regions had changed their sign from 1993 to 2008, while 11 remained specialised in the same direction. So this does not indicate any clear trend. When looking at the 5 regions with the highest export share, Fejer, Komarom, Győr, Vas, and Somogy, no distinct trend is clear either: while the first two regions had higher than average specialization in both years, Győr had a lower specialization in both years, whereas Vas and Somogy had changed the direction of deviation in opposite sides during the period.

Therefore, in order to get further information, a trend regression was run for the dependent variable, regional specialization. The equation for this regression is shown in the following:

\[
\text{SPEC}_{j,t} = C + a_1 \text{Trend} + a_2 \text{Resid}_{j,t-1}
\]

The residual correction was necessary as without it, errors were auto-correlated as indicated by the value for the Durbin Watson statistic which then was lower than the low-sided boundary for that value (0.3216).

**Table 26: Results of trend estimation for regional specialization in Hungary from 1993 to 1999**

<table>
<thead>
<tr>
<th>Variable</th>
<th>For</th>
<th>Coefficient (p-value in italics)</th>
<th>had expected sign/significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Constant</td>
<td>0.375719 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>Trend</td>
<td>Trend</td>
<td>0.007102 (0.0033)</td>
<td>✓</td>
</tr>
<tr>
<td>Resid ( \text{e}_{j,t-1} )</td>
<td>AR-correction factor</td>
<td>0.889334 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>0.7843</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td>140</td>
<td></td>
</tr>
</tbody>
</table>

* Significant (at the 5% level)
** Highly significant (at the 1% level)
*** Very highly significant (at the 0.1% level)

**Source:** Own calculations.
The trend regression was run firstly for the periods 1993 to 1999, during which an increase of regional specialization had been observed up to a peak in 1999 (see section 5.3.1 of chapter 5), secondly for the long period 1994-2008, and thirdly for the period 2004-2008, i.e. since the EU accession of Hungary (80 observations), during which a slight increase in regional specialization had been observed (ibid). The coefficient for the latter post-accession period was small and positive (0.0068), it remained insignificant, however, and is therefore not shown here. The results of the first two trend regressions are presented in Table 26 and Table 27.

The trend regression shown in Table 26 was run for regional specialization over the 20 regions and for the years 1993 to 1999, i.e. with 140 observations. The trend was small and positive as expected and highly significant, as was the constant. $R^2$ was 78.4%.

### Table 27: Results of trend estimation for regional specialization in Hungary from 1994 to 2008

<table>
<thead>
<tr>
<th>Variable</th>
<th>For</th>
<th>Coefficient (p-value in italics)</th>
<th>had expected sign/ significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Constant</td>
<td>0.404251 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>Trend</td>
<td>Trend</td>
<td>-0.002272 (0.0075)</td>
<td>✓</td>
</tr>
<tr>
<td>Resid $\epsilon_{j,t-1}$</td>
<td>AR-correction factor</td>
<td>0.851413 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.7067</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

* Significant (at the 5% level)
** Highly significant (at the 1% level)
*** Very highly significant (at the 0.1% level)

Source: Own calculations.

The trend regression shown in Table 27 was run for regional specialization over the 20 regions and for the years 1994 to 2008\(^{165}\), i.e. with 300 observations. The trend had a negative sign and a small coefficient, as was expected. It was highly significant, as was the constant. $R^2$ was 70.6%.

---

\(^{165}\) Initially 1993, but the start year was eliminated due to the AR-correction.
6.2.2 Selection of independent variables and description of data

The model of regression 2 aims at analysing which influences were significant for the degree of regional specialization of Hungarian regions during 1993 to 2008 in context with increasing European integration.

Which influences were relevant for the specialization of a region in context with European integration, and which region-specific factors did also play a role?

The dependent variable chosen was regional specialization, measured by the Krugman specialization index calculated with manufacturing employment data. The explanatory variables were chosen according to what would be desirable to include, on the one hand, and subject to the constraint which data were available from HCSO at the level of the 20 regions and for the entire period 1993 to 2008, on the other hand.

In this section, the selection of the independent variables which entered into the regression 2 shall be described, their data shall be analysed and an expectation as to the sign of the coefficients of the independent variables shall be given in light of relevant theories.

The following equation has been formulated for regression 2:

\[
\text{SPEC}_{j,t} = \alpha_0 + \alpha_1 \text{INT}_j + \alpha_2 \text{East}_j + \alpha_3 \text{LAct}_{j,t} + \alpha_4 \text{LFDI}_{j,t} + \\
\alpha_5 \text{Mig}_{j,t} + \alpha_6 \text{Exsh}_{j,t} + \alpha_7 \text{LGW}_{j,t} + \text{resid} \varepsilon_{j,t-1}
\]

\(\text{SPEC}_{j,t}\) is the dependent variable, a region's specialization in manufacturing industries as measured by the Krugman specialization index at time \(t\) (4 digits after the comma; see section 5.2.1 for the formula and data base).

The independent variables:

\(\text{C}\) The constant was included in this linear multiple regression.

\(\text{INT}_j\) This is a dummy variable which takes the value 1 if the region is an internal region - which is the case for 8 of the 20 regions -, and 0 otherwise. The reference category was thus being a border region, either with an external country (BEX), or with an accession country (BCE), or with a former EU-15 member state (BEU).
Being an internal region also indicates centrality - in the sense of NEG theories - with respect to the region’s location, i.e. closeness to the capital region Budapest with the international airport and other facilities. Therefore, the expected sign of the coefficient is positive, i.e. that internal regions have a higher specialization than border regions.

It had been tried to run the regression with three dummies for these three types of border regions (BEU, BCE, and BEX) and the internal regions as the reference category; these three variables remained insignificant, however, without improving the quality of the regression, therefore including \( \text{INT}_j \) was the best choice.

**East}_j** This is a dummy variable for the location of a region in the economically less developed Eastern part of the country,\(^\text{166}\) which takes the value 1 if the region is located there – which is the case for 7 of the 20 regions-, and 0 otherwise. The expectation for the sign of the coefficient is based on the empirical results derived for the specialization of Eastern regions in Chapter 5, section 5.4.1, namely that their specialization is slightly higher than that of Western regions, which was attributed to their thinner economic and manufacturing base. Therefore, the expected sign for the coefficient of \( \text{East}_j \) is positive.

**LAct}_j,t** This is the economically active population of the region in a year, of which the logarithm was taken. The economically active population was preferred to the population of the region as such due to its higher relevance for manufacturing employment, based on which the dependent variable - regional specialization - was calculated.

For some regions, the economically active population increased over the period 1993 to 2008, as was the case for Győr (+39.0% to 197,900), for some it decreased as for Budapest (-11.5% to 780,800). Taking the logarithm smoothed out the differences in endowment with active population among the regions, improving the results for this variable (not altering the sign of the coefficient though).

A larger number of economically active people are expected to allow for a broadening of the industrial base including manufacturing sectors, i.e. for a greater diversification of the region. Therefore, the expected sign for the coefficient with respect to regional specialization is negative.

\(^{166}\) Division into Eastern part and Western part according to László Faragó (1999), as in Chapter 5.
LFDI_{j,t} \hspace{1em} \text{This is the share of FDI in enterprises with FDI (minimum 10\%) in the region at time } t, \text{ in billion HUF, of which the logarithm was taken.}^{167} \text{ The FDI capital in enterprises with FDI of a region, in billion HUF, showed an increase from low to higher levels for all regions over the period 1993 to 2008. For Hungary as a whole, this increased from 662.9 to 11,795 billion HUF. The logarithm was taken to smooth values for this variable. A larger FDI stock is expected to allow for a broadening of the industrial base of the region; therefore, the expected sign for the coefficient is negative.}

Mig_{j,t} \hspace{1em} \text{This is the internal net migration of the region in a year (not in logarithm, as the numbers are sometimes positive and sometimes negative). The data for internal net migration of the 20 regions over the period 1993 to 2008 have been described in detail in section 3.5 and subsections of Chapter 3, and a link has been examined between the balance of internal net migration per region and the regional wage level, verifying what is described in the relevant NEG models (please, see also the tables, maps and graphs in that section). In particular, the model by } \textit{Ludema & Wooton (1997)} \text{ allows for (partial) internal migration and for regional preferences of manufacturing workers. In case 5 of the equilibria described for the model (see Chapter 2, section 2.1.2.3), which form in the course of integration processes, regional income differences prevail. Both regions have some industry, however, to a different extent.}

With a positive internal net migration - such as for Pest region and many Western regions-, regional specialization is expected to decline, whereas with a negative internal net migration, specialization is expected to rise. Therefore, the expected sign for the coefficient is negative.

Exsh_{j,t} \hspace{1em} \text{This is the share of exports (to world) in sales of industry located in region } j \text{ in year } t. \text{ Even though the variable is based on exports to the world, this points to the EU which had a rising share of around 80\% of Hungary’s total exports in the latter years of the period. The variable } Exsh_{j,t} \text{ indicates the degree of openness of a region, and also at which level its economy is internationally linked, to a large extent also through the division of labour with the EU.}

Over the period, the figures for the share of exports (to world) in sales of industry by region rose for all regions and for Hungary as a whole. For the year 1993, the export share ranged from 9.8 to 39.1\%, whereas for the year 2008, the range was from 28.1 to 89.4\%. Five regions have developed into strongly export-oriented

---

167 The HCSO calls this the “share” of FDI in its statistical publications, even though it should rather be called the foreign “capital”.
regions over the period: Fejer, Komarom-Esztergom, Győr-Moson-Sopron, Vas, and Somogy. All of these are in the Western part of Hungary, but only two have a common border with Austria (with the former EU-15).

In the models of New Trade Theory based on a world with transportation costs, big countries gain a comparatively larger share in industries where product differentiation and internal or external economies of scale are important (*Helpman & Krugman 1985*). New Trade Theory also postulates that even in the absence of differences in factor endowments, scale economies can induce specialization among countries (regions in this study) and affect relative concentration. This is complemented by the NEG model of *Livas-Elizondo & Krugman (1995)* who explicitly modelled a country with two (or several) regions. Therefore, the expected sign for the coefficient of the export variable is negative: a higher share of exports in sales is expected to lower the degree of regional specialization.

\[ \text{LGW}_{jt} \] This is the (average) monthly gross wage rate in manufacturing industry of the region in year t (in million HUF), corrected by the producer price index for the manufacturing sector\(^{168}\). The logarithm was taken of these values. These absolute numbers for the gross wage rates are similar to the regional profile which had been shown in section 3.5.2 of Chapter 3 (Figure 14 illustrated the monthly gross wages in manufacturing per region as percent of the national average for 2005 and 1992).

In the NEG model by *Krugman (1991a)*, higher real wages prevail in the agglomeration, which is the region with high specialization. In the *Ludema & Wooton (1997)* model with partial internal migration, case 5 of the equilibria in the course integration processes describes that regional income differences prevail. Both regions have some industry, however, to a different extent. Therefore, the expected sign for the coefficient of the \( \text{LGW}_{jt} \) variable with respect to regional specialization is negative.

\[ \text{resid } e_{jt-1} \] This AR correction was necessary as in the original regression, errors were auto-correlated as indicated by the value for the Durbin Watson statistic which was lower than the low-sided boundary for that value (0.3873). After this correction, the problem was eliminated and the quality of the regression improved \( (R^2 \) rose from around 20\% to 70\%).

---

\(^{168}\) This producer price index is published by the HCSO as an index with previous year =100. In order to take out inflation, I corrected this by holding 1992=100 constant and using a producer price index in terms of prices of 1992 as correction.
Prior to including the AR correction, it had also been tried to include a linear trend into the regression; this remained insignificant, however, and did not alter the low result of the Durbin-Watson test. So that idea was dropped.

As a variable for the quality of infrastructure of a region which was available from the HCSO data for the entire period, the number of major roads \( (RDS_{j,t}) \) per 100 km\(^2\) of a region in a year has been considered. The idea behind would be that better infrastructure encourages more industry into the region, fosters the region’s diversification and thus reduces its specialization. The data for this variable are such that for some regions, Pest for example (the region surrounding the capital), the figures are around 40, whereas for other regions, like Hajdu and Szolnok, the figures are around 25. The data show hardly any variation over time, i.e. they remain more or less the same, improving only slightly over the period 1993 to 2008. The coefficient had the expected negative sign with respect to regional specialization; but it remained (very) insignificant and did not improve the quality of the regression either. Therefore, it was eliminated in the manual process of stepwise regression.

6.2.3 Results of regression 2 - Regional specialization

The results of regression 2 on regional specialization in Hungary are presented in Table 28. The table shows the coefficients for the constant and the explanatory variables as well as the error correction term, and beneath each the p-values in italics and parenthesis. The last column indicates whether the coefficient had the expected sign as well as the significance level.

The regression 2 in its form as presented had an \( R^2 \) of 70.58% and an adjusted \( R^2 \) of 69.77%. The F-statistic was sufficiently high at 87.27. The constant had a positive coefficient and was significant at the 1% level. The dummy for being an internal region was very highly significant, and the coefficient had the expected positive sign. The dummy for being a region in the Eastern part had the expected sign, but was not significant. The variable for the economically active population of a region (in logarithm) was very highly significant, and the coefficient had the expected negative sign.

The FDI stock of a region (in logarithm) was significant, and the coefficient had the expected negative sign. Internal net migration was very highly significant, and the coefficient - though small due to the fact that the values were not in logarithm - had the expected sign. The coefficient for the export share of a region had the expected sign (was small though), but was not significant. The gross manufacturing wage of a region (in logarithm) had the expected negative sign, but was
not significant either. Finally, the residual error correction term was very highly significant.

Table 28: Results of regression 2 on regional specialization in Hungary from 1993 to 2008

<table>
<thead>
<tr>
<th>Variable</th>
<th>for</th>
<th>Coefficient (p-value in italics)</th>
<th>had expected sign/ significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Constant</td>
<td>0.959313 (0.0017)</td>
<td>***</td>
</tr>
<tr>
<td>Int$_{ij}$</td>
<td>Internal region</td>
<td>0.022274 (0.0076)</td>
<td>✓</td>
</tr>
<tr>
<td>East$_{ij}$</td>
<td>Eastern region</td>
<td>0.005529 (0.4933)</td>
<td>✓</td>
</tr>
<tr>
<td>LAct$_{t}$</td>
<td>Lg of active population</td>
<td>-0.097088 (0.0012)</td>
<td>✓</td>
</tr>
<tr>
<td>LFD$_{t}$</td>
<td>Lg of FDI stock</td>
<td>-0.028406 (0.0631)</td>
<td>*</td>
</tr>
<tr>
<td>Mig$_{t}$</td>
<td>Internal net migration</td>
<td>-6.08 E-06 (0.0000)</td>
<td>***</td>
</tr>
<tr>
<td>Exsh$_{t}$</td>
<td>Export share of the region</td>
<td>-5.24 E-05 (0.8328)</td>
<td>✓</td>
</tr>
<tr>
<td>LGW$_{t}$</td>
<td>Lg of gross manufacturing wage</td>
<td>-0.004222 (0.9360)</td>
<td>✓</td>
</tr>
<tr>
<td>Resid $\varepsilon_{t-1}$</td>
<td>AR-correction factor</td>
<td>0.808182 (0.0000)</td>
<td>***</td>
</tr>
</tbody>
</table>

R$^2$                | 0.7058                      |
adjusted R$^2$       | 0.6977                      |
F-statistic          | 87.27                       |

Number of observations | 300                          |

* Significant (at the 10% level)
** Highly significant (at the 5% level)
*** Very highly significant (at the 1% level)

Source: Own calculations.
6.2.4 Evaluation of results in view of the hypothesis, NEG models and previous empirical studies

The question for regression 2 on regional specialization was: Which influences were relevant for the specialization of a region in context with European integration, and which region-specific factors did also play a role?

The relevant NEG models which play a role for the interpretation shall be recalled. The model by Livas-Elizondo & Krugman (1996) for regional integration modelled in the context of international trade described that an (irreversible) agglomeration would form within the country consisting of more than one region, and that specialization would increase there. The model by Puga (1999) predicted an \( \Omega \)-shaped relationship between agglomeration and trade costs with proceeding integration, with regional specialization reaching its highest level when agglomeration is strongest. The model by Ludema & Wooton (1997) predicted (less than complete) agglomeration to form at some intermediate level of trade costs, followed by dispersion as trade costs decline even further in the course of proceeding integration. Finally, in the model by Krugman & Venables (1996), with labour assumed internationally immobile but moving between different sectors, agglomeration will become necessary for trade costs at very low levels, such that each region (country) will lose its presence in one of the industries; thus specialization is then increasing to high levels.

The results of regression 2 show that of the two integration variables - a region’s FDI stock and its export share - only the FDI stock was significant with the Hungarian data, reducing regional specialization and allowing for a broadening of the manufacturing industry base in a region.

Of the region-specific factors related to a region’s location - being an internal region, and being in the (economically backward) Eastern part - only the first was significant for regional specialization. An internal region had a regional specialization which was 0.0222 higher on average than that of a border region, controlling for other variables.

Of the three region-specific factors related to population, a region’s active population (in logarithm) had the largest influence on regional specialization, reducing it by allowing for the broadening of the manufacturing sectors of a region. Internal net migration had a negative influence on specialization. In Pest region, the region with the largest internal net migration (even in absolute terms), regional specialization decreased by -35% from 1993 to 2008, while the cumulative internal net migration was +242,000 (or almost 20% in terms of the region’s population in 2005). This would point to the NEG-model by Ludema & Wooton.
as the relevant model for the specialization of Hungary's regions in the observation period, as inter-regional migration did play a significant role for the degree of regional specialization, perhaps working through a whole chain of other influences such as the local housing markets. Finally, the gross manufacturing wage of a region did not have a significant influence on regional specialization, unlike the importance it is generally given in NEG models based on the original Krugman (1991a) model.

Few other authors have run regressions on regional specialization in the EU or in CEECs in previous empirical literature. While some studies exist on country-level specialization of EU member states, e.g. on decreasing trade specialization (Sapir 1996) and decreasing production specialization (Amiti 1997, Midelfart-Knarvik et al. 2000), few dealt with the specialization of regions by way of regression analysis. Hallet (2000) estimated a trend regression for regional specialization of EU-15 regions based on production data from 1980 to 1995. The trend line was a linear regression with $\beta = -0.3869$ (thus much stronger than in my trend regressions, due to the longer time span and the fact that the sample also included regions of large countries such as Germany) and an $R^2$ of 43.1%. He did not analyse influences on regional specialization though.

Based on specialization of regions by output, Traistaru et al. (2003) showed a trend regression for Hungary over the period 1990 to 1999. This had the very low $R^2$ within of 0.0074. The time trend had a small negative coefficient -0.0019, but was not significant. Instead of a regression about influences on regional specialization, the authors produced summary statistics for regional specialization and a number of economic indicators (GDP per capita, wages, and unemployment) relative to the national averages and for the years 1990 and 1999. They inferred that border regions (all three categories analysed, BEU, BCE and BEX) were found to be more specialised compared to the national average, while internal regions were less specialised. High specialization was associated with inferior economic performance, while regions with low specialization performed better than the national averages.

Compared with the attempts in previous literature, regression 2 has brought forward some interesting results about region-specific characteristics and the role of proceeding European integration for regional specialization. Based on the empirical results of regression 2 for regional specialization in Hungary during 1993 to 2008, the following general statement can be made: Regional specialization is an attribute of a region which is more transitory in nature than the degree of concentration of an industry is (regression 1). Regional specialization of a given region changes over time, and the direction for a given region can rarely be predicted (as shown in Figure 45 of section 6.2.1.). What can be predicted,
however, is the average degree of specialization for a group of regions during proceeding integration processes, based on the NEG models by Livas-Elizondo & Krugman (1996), Ludema & Wooton (1997), Puga (1999), and Krugman & Venables (1996), which have been chosen as the analytic framework for this part of the research. Past specialization of groups of Hungarian regions has been analysed empirically in Chapter 5, section 5.4 and subsections.

This concludes chapter 6 about regression analysis of manufacturing industry concentration and regional specialization in Hungary during the 1993 to 2008 period. The final chapter 7 will provide a summary of the results of this research and draw conclusions on the direct effects which pre-accession policies such as the Europe agreement had on agglomerations and regional development in Hungary. Further, it will highlight the role which agglomeration and regional specialization can play in the overall cohesion process. Future options for the reform of European regional policy will be discussed in the face of new needs after Eastern enlargement, based on results for the Hungarian regions. And finally, some policy recommendations will be given regarding future EU enlargement candidates.