External effects of industrial clustering in Poland

Dorota Ciołek, Tomasz Brodzicki
University of Gdansk

38th MACROMODELS INTERNATIONAL CONFERENCE
Poznań 30.11-3.12.2011
Motivation

✓ Functioning industrial clusters - drivers of regional growth (Porter 1998)
✓ The actual impact of clusters unclear – Martin & Sunley (JEG, 2005)
✓ Lack of methodologically sophisticated analysis - for Poland and the EU
✓ First-attempt: Rodriguez-Pose and Comptur (2010) *Do clusters generate greater innovation and growth? An analysis of European regions*, WP IMDEA 2010/15 => the role of clusters in regional growth outweighed by the role of the social enviroment
✓ Cluster-based policy (CBP) adopted and implemented on the basis of partial evidence

The research project:
“Identyfikacja klastrów przemysłowych w Polsce. Próba oceny ich efektów ekonomicznych. Implikacje dla polityki rozwoju regionalnego”
(1649/B/H03/2010/38) granted by the Ministry of Science and Higher Education of the Republic of Poland
What are industrial clusters?

✓ „Geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (for example universities, standards agencies, and trade associations) in particular fields that compete but also co-operate.” / M. Porter 1998, p. 19/

✓ For the purpose of the present study clusters:

  spatial concentrations of interlinked firms in vertically or horizontally related areas of economic activity

✓ One has to note that:

  Clusters could be in different phases of cluster life cycle (CLC) – focus on mature – functioning clusters

  Cluster types: industrial districts (SMEs driven) vs. hub-and-spoke clusters (dominated by a small no of large enterprises)
What are industrial clusters?

Features of clusters:

- regional concentration of vertically/horizontally linked firms in interwoven sectors
- critical mass of market and non-market institutions (number of enterprises)
- dense network of formal and informal relations between market agents (spatially bounded)
- co-opetition (cooperative behavior) – fierce competition with simultaneous cooperation in some fields
- detectable benefits:
  - for the participants of the cluster (internal effects)
  - for the regional economy (external effects - positive externalities) => potential positive and statistically significant impact on the level of economic development (ln GDPpc) and/or the rate of growth
Cluster mapping – methodology

Mapping team: M. Tarkowski, T. Brodzicki, D. Ciołek.

Analysis:

✓ Spatial disaggregation: LAD 4 level – powiat
✓ Sectoral disaggregation: PKD 2004 – 3-digit NACE rev. 2
✓ Information on employment, no of enterprises, no of enterprises by employment size (data source: GUS)
✓ Spatial autocorrelation taken into account => spatial weighting matrix
✓ Static approach – analysis based on 2006 only (due to data availability)
Cluster mapping – methodology

Cluster sectors – bundling methods considered:

- Analysis of sectoral agglomeration patterns at different levels of spatial disaggregation (LAD 4, NUTS 3, NUTS 2)
- Analysis of input-output tables for Poland (methodology of Titze et al. 2010, NUTS3 – 2-digit NACE => vertical clustering)
- Analysis of co-localization patterns for Poland (horizontal clustering)
- Porter et al. (2003) – cluster sectors definitions – adopted by Europe Innova Cluster Observatory
  - General conclusion: patterns for Poland do not deviate significantly from American patterns (technological convergence)
- Porter et al. (2003) cluster sectors definitions adopted for comparability of results
Cluster mapping – methodology

Identification of clusters:

- Location quotients (LQ) for spatially-weighted employment in cluster sectors bundles (LQ_EW)
- Number of enterprises in core sectors of clusters, spatially-adjusted (N_core)
- Threshold levels adopted by the analysis of actual distribution of LQ_EW and N_core and the use of the Fisher-Jenks method
- Adjustments for clusters with low numbers of enterprises in the core cluster sectors (*hub-and-spoke clusters*)
- Adjustments for municipal and non-municipal (rural) local administrative districts – economic geography perspective
Cluster mapping – biopharmaceuticals (K3)
External effects of clusters

Main hypothesis: potential positive and statistically significant impact on:
- the level of economic development (GDPpc)
- the rate of growth

Features to be taken into account:
- Existence of clusters (agglomeration economies, MAR externalities, etc.)
- Number of clusters in a given region
- Specialization relative to national average
- Diversification of clusters in a given region (Jacobian externalities)
- Strength of regional clusters – the role played in the regional economy (how structurally clusterized is the regional economy?)
Clusterization index for regions

Modified EI Cluster Observatory indices

- Specialization quotient (SQ)
- Significance index
- Diversification index

- Results of cluster mapping and data aggregated from LAD 4 (powiats) to NUTS 3 (subregions) levels
Spatial data – observations about objects with additional information about their position in space $S$:

$$S = \{O, R\}$$

- $O$ – objects of the study (geometrical figures: points, lines or polygons in $\mathbb{R}^k$),
- $R$ – relations between objects.

Observed sample drawn from a Probability Model:

$$\Phi = \left\{ f_{x_{s_1}, x_{s_2}, \ldots, x_{s_n}} \left( x_{s_1}, x_{s_2}, \ldots, x_{s_n} : \theta \right) \mid s_i \in S, \theta \in \Theta \right\}$$

where: $x_{s_1}, x_{s_2}, \ldots, x_{s_n}$ random variables ordered with respect to their geographical location,

$$f_{x_{s_1}, x_{s_2}, \ldots, x_{s_n}} \left( x_{s_1}, x_{s_2}, \ldots, x_{s_n} ; \theta \right)$$ joint probability density function.
If we define:

\[ x \] – longitude of the geometric center of subregion, \\
\[ y \] – latitude of the geometric center of subregion.

we can estimate different surface trends:

\[ y_i = a_0 + a_1 x_i + a_2 y_i + \varepsilon_i \]

\[ y_i = a_0 + a_1 x_i + a_2 y_i + a_3 x_i y_i + a_4 x_i^2 y_i + a_5 x_i y_i^2 + a_5 x_i^2 y_i^2 + \varepsilon_i \]

in general:

\[ y_i = \sum_{r+s \leq p} a_{rs} x_i^r y_i^s + \varepsilon_i \]

\[ p \] - the order of the trend.
1) SAR – Spatial Autoregressive Model

\[ y = \rho W y + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I). \]

\(\rho\) - autoregressive parameter

\(W\) - standardized weight matrix of connectivity.

reduced to:

\[ (I - \rho W)y = X\beta + \varepsilon \Rightarrow \]
\[ y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon \Rightarrow \]
\[ y = (I - \rho W)^{-1} X\beta + \xi \]

where covariance matrix of disturbance:

\[ \text{E}(\xi \xi^T) = \sigma^2 (I - \rho W)^{-1} [(I - \rho W)^{-1}]^T \]
Linear Regression Models for Spatial Data

2) SEM – Spatial Error Model

\[ y = X\beta + \xi, \quad \xi = \lambda W\xi + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I). \]

\( \lambda \) - autocorrelation parameter

or

\[ y = X\beta + (I - \lambda W)^{-1} \epsilon, \quad \epsilon \sim N(0, \sigma^2 I). \]

3) SCM – Spatial Cross-regressive Model

\[ y = X\beta + WX\gamma + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I). \]
4) SARAR – Spatial Autoregressive with Autoregressive disturbances

\[ y = \rho W_1 Y + X\beta + \xi, \quad \xi = \lambda W_2 \xi + \epsilon, \]

reduced form:

\[ \epsilon \sim N(0, \sigma^2 I). \]

\[ y = (I - \rho W_1)^{-1} X\beta + (I - \rho W_1)^{-1} (I - \lambda W_2)^{-1} \epsilon, \]

\[ \epsilon \sim N(0, \sigma^2 I). \]

\( W_1 \) and \( W_2 \) – two different types of connectivity matrixes
Specific features of spatial data

1) Spatial dependence - autoregression of variables

Moran I statistic:

\[ I_W = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{z^T W z}{z^T z} \]

where: \( W \) - connectivity matrix - binary matrix of elements \( w_{ij} \)

\[ w_{ij} = \begin{cases} 1 & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases} \]

\( N(i) \) – set of neighbors according to chosen definition.

Two different definitions of neighborhood:

- **W1** - matrix of first order neighborhood - common border between regions,
- **W2** – matrix of inversed distances (in km) between objects.
Specific features of spatial data

2) Heterogeneity – spatial nonstacionarity

where spatial stacionarity is defined as:

\[
E[X(s)] = E[X(s + \delta)] = \mu \\
E[X(s)^2] = E[X(s + \delta)^2] = \sigma^2 \\
E[X(s_i)X(s_j)] = \gamma(d_{ij})
\]

\(d_{ji}\) – distance between objects \(i\) and \(j\).

Sources of heterogeneity:
- Spatial heteroscedasticity of disturbances,
- Unstable parameters in space.
Specific features of spatial data

Heteroscedasticity should be tested together with autocorrelation in space:

**SEM:** \[ y = X\beta + \xi, \quad \xi = \lambda W\xi + \epsilon, \quad \epsilon \sim N(0, \Omega) \]

Diagonal elements of \( \Omega \): \[ \sigma_i^2 = \sigma^2 f(z^T \alpha) \] functions of additional variables \( Z \)

\[ H_0 : \lambda = 0, \alpha = 0 \]

Joint Lagrange Multiplier statistic (Anselin, 1988):

\[ LMJ = \frac{1}{2} \hat{f}^T Z (Z^T Z)^{-1} Z^T \hat{f} + \frac{1}{T_r} \left( \frac{e^T We}{S_e^2} \right)^2 \sim \chi^2_{(p+1)} \]

\( Z \) – additional variables,

\( T_r = tr \left[ (W^T + W)W \right] \)

\( \hat{f} \) – (Nx1) vector of elements

\[ \left( \frac{e_i^2}{S_e^2} - 1 \right) \]
Spatial model of external effects of clusters

1)  \[
\ln(\text{GDPpc}_{i,2006}) = \beta_0 + \rho W \ln(\text{GDPpc}_{i,2006})_i + \\
+ \beta_1 \text{Soc. filter}_i + \beta_2 \text{HC. filter}_i + \beta_3 \text{Cluster. filter}_i + \zeta_i \\
\zeta_i = \lambda W \zeta_i + \epsilon_i
\]

2)  \[
growth_i = \beta_0 + \rho W \text{growth}_i + \beta_1 \ln(\text{GDPpc}_{i,2003}) + \\
+ \beta_2 \text{Soc. filter}_i + \beta_3 \text{HC. filter}_i + \beta_4 \text{Cluster. filter}_i + \xi_i \\
\xi_i = \lambda W \xi_i + \epsilon_i
\]

\text{GDPpc}_{i,t} - \text{Gross Domestic Product per capita in subregion NUTS-3} \ i \ \text{in year} \ t \\
(\text{constant prices} \ 2003) \\
\ i=1, \ldots, 66

The average rate of growth in 2003-2008:

\[
growth_i = \ln\left(\frac{\text{GDPpc}_{i,2008}}{\text{GDPpc}_{i,2003}}\right)/5 \\
\ i=1, \ldots, 66
\]
Explanatory variables

Products of Principle Components Analysis:

1) Social Environment Filter:
   - Long-term unemployment rate
   - Share of employment in agriculture
   - Share of young people in total population (age 15-29)

2) Human Capital Filter:
   - Share of people with higher education in population of the subregion
   - School enrolment on the secondary level
   - School enrolment on the upper secondary level
   - Numbers of beds in hospitals per 1000 habitants

3) Cluster Filter:
   - Specialization quotient (SQ)
   - Significance index
   - Diversification index
Estimation results for $\ln(GDP \text{ per capita})$

**OLS estimation:**

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Explained</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\ln(GDP_{pc_2006})$</td>
<td></td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>OLS</td>
<td></td>
</tr>
<tr>
<td><strong>const.</strong></td>
<td>10.3112</td>
<td>[0.0000]</td>
</tr>
<tr>
<td><strong>Soc.filter</strong></td>
<td>0.12606</td>
<td>[0.0018]</td>
</tr>
<tr>
<td><strong>HC.filter</strong></td>
<td>0.1336</td>
<td>[0.0000]</td>
</tr>
<tr>
<td><strong>Cluster.filter</strong></td>
<td>0.05240</td>
<td>[0.1254]</td>
</tr>
<tr>
<td><strong>Adj. $R^2$</strong></td>
<td>0.6942</td>
<td></td>
</tr>
<tr>
<td><strong>Wald Statistic</strong></td>
<td>50.17</td>
<td>[0.0000]</td>
</tr>
<tr>
<td><strong>AIC</strong></td>
<td>-34.759</td>
<td></td>
</tr>
</tbody>
</table>

Moran I statistic: 0.18419  
$p$-value = 0.00786
Estimation results for ln(GDP per capita)

<table>
<thead>
<tr>
<th>Model</th>
<th>SARAR</th>
<th>SARAR</th>
<th>SARAR</th>
<th>SARAR</th>
<th>SAR</th>
<th>SAR</th>
<th>SEM</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix of neighborhood</td>
<td>W1, W1</td>
<td>W2, W2</td>
<td>W1, W2</td>
<td>W2, W1</td>
<td>W1</td>
<td>W2</td>
<td>W1</td>
<td>W2</td>
</tr>
<tr>
<td><strong>Explain: lnGDPpc_2006</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$ - autoregression</td>
<td>0.1749 [0.4641]</td>
<td>0.4442 [0.3496]</td>
<td>0.3228 [0.1519]</td>
<td>0.3930 [0.2222]</td>
<td>0.3768 [0.0081]</td>
<td>0.5535 [0.0600]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$ - autocorrelation</td>
<td>0.2642 [0.2133]</td>
<td>0.4815 [0.3683]</td>
<td>0.4123 [0.4300]</td>
<td>0.1736 [0.0167]</td>
<td></td>
<td></td>
<td>0.4541 [0.0040]</td>
<td>0.62886 [0.0706]</td>
</tr>
<tr>
<td>LR test</td>
<td>8.9514 [0.0114]</td>
<td>4.8280 [0.0894]</td>
<td>7.4571 [0.0240]</td>
<td>9.596 [0.0167]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald Statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10,305 [0.0013]</td>
<td>6,369 [0.0116]</td>
<td>11,417 [0.0007]</td>
<td>7,9254 [0.0049]</td>
</tr>
</tbody>
</table>
Estimation results for growth

OLS estimation:

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Explained growth</th>
<th>Moran I statistic: 0.24877 p-value = 0.00089</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>OLS</td>
<td></td>
</tr>
<tr>
<td>Const.</td>
<td>-0.0911 [0.3789]</td>
<td></td>
</tr>
<tr>
<td>Soc.filter</td>
<td>-0.0028 [0.3573]</td>
<td></td>
</tr>
<tr>
<td>HC.filter</td>
<td>-0.0034 [0.1572]</td>
<td></td>
</tr>
<tr>
<td>Cluster.filter</td>
<td>0.0049 [0.0429]</td>
<td></td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.1537</td>
<td></td>
</tr>
<tr>
<td>Wald Statistic</td>
<td>2.77 [0.0350]</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>-386.02</td>
<td></td>
</tr>
</tbody>
</table>
**Estimation results for GROWTH**

<table>
<thead>
<tr>
<th>Model</th>
<th>SARAR</th>
<th>SARAR</th>
<th>SARAR</th>
<th>SARAR</th>
<th>SAR</th>
<th>SAR</th>
<th>SEM</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix of neighborhood</td>
<td>W1, W1</td>
<td>W2, W2</td>
<td>W1, W2</td>
<td>W2, W1</td>
<td>W1</td>
<td>W2</td>
<td>W1</td>
<td>W2</td>
</tr>
<tr>
<td>Explained: growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$ - autoregression</td>
<td>0.23588 [0.6693]</td>
<td>0.40793 [0.8039]</td>
<td>0.3327 [0.1066]</td>
<td>0.22982 [0.6512]</td>
<td><strong>0.3539 [0.0286]</strong></td>
<td>0.54475 [0.1432]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\lambda$ - autocorrelation</td>
<td>0.1635 [0.7835]</td>
<td>0.3409 [0.8479]</td>
<td>0.1244 [0.8197]</td>
<td>0.3184 [0.1441]</td>
<td>-</td>
<td>-</td>
<td><strong>0.3622 [0.0342]</strong></td>
<td>0.5332 [0.1712]</td>
</tr>
<tr>
<td>LR test</td>
<td>5.0109 [0.0816]</td>
<td>2.5705 [0.2766]</td>
<td>4.3373 [0.0890]</td>
<td>4.6845 [0.0961]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wald Statistic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td><strong>6.3055 [0.0120]</strong></td>
<td>4.5087 [0.0337]</td>
<td><strong>6.2028 [0.0127]</strong></td>
<td>4.0712 [0.0436]</td>
</tr>
<tr>
<td>AIC</td>
<td>-387.03</td>
<td>-384.59</td>
<td>-386.86</td>
<td>-386.70</td>
<td><strong>-388.81</strong></td>
<td>-386.16</td>
<td><strong>-388.5</strong></td>
<td>-385.89</td>
</tr>
</tbody>
</table>
### Estimation results

<table>
<thead>
<tr>
<th>Covariates</th>
<th>SEM (W1)</th>
<th>SAR (W1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGDPpc_2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>growth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>const.</td>
<td>10.3339</td>
<td>-0.0923</td>
</tr>
<tr>
<td></td>
<td>[0.0000]</td>
<td>[0.3252]</td>
</tr>
<tr>
<td>lnGDPpc2003</td>
<td></td>
<td>0.0124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.1799]</td>
</tr>
<tr>
<td>Soc.filter</td>
<td>0.1079</td>
<td>-0.0023</td>
</tr>
<tr>
<td></td>
<td>[0.0042]</td>
<td>[0.4005]</td>
</tr>
<tr>
<td>HC.filter</td>
<td>0.14539</td>
<td>-0.0034</td>
</tr>
<tr>
<td></td>
<td>[0.0000]</td>
<td>[0.1162]</td>
</tr>
<tr>
<td>Cluster.filter</td>
<td>0.0568</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>[0.0905]</td>
<td>[0.0794]</td>
</tr>
</tbody>
</table>
Methodology:
- Spatial dependence is **significant** in analyses of the regional level of economic development and the economic growth.
- The character of the dependence could be different for different variables.

Economic:
- We got same support of thesis that functioning of clusters in the region significantly **influences the level of economic development**.
- No exact evidence of influence on economic growth – probably time analysis is necessary.
What could be done more?

- More analysis required
  - Longer data set – panel data analysis
  - Higher/lower levels of spatial aggregation => NUTS 2 (1995 – 2010)
  - Model applied to NUTS 2 regions of EU27 (CRP 2010 with spatial interactions)

- Other possibilities
  - => theoretical regional development model leading to clear structural equation (restrictions on parameters in the empirical equation)
  - => NEG modelling