Approaching regional openness through measures of specialization and spatial market shares: experimentations with micro-data on enterprises

Moritz Lennert*

Abstract

In this paper, I present the results of ongoing experimentations with micro-data data of firms in order to develop better insights into the economic specialization and openness to trade of regions. Using newly developed indicators, one based on Huff-like models and another on Marcon and Puech's M function, I explore the possibilities offered by building analyses on the basis of point data instead of the classic regional aggregations.

Keywords

Economic geography, Spatial concentration, Regional trade, Location quotient, Micro-data

Introduction

Classical approaches to estimating the share of regional production that is exported outside of the region are mainly based on notions of concentration, most notably on location quotients, or derivatives thereof, which, despite severe criticism (Isserman, 1980; Richardson, 1985), still are subject

* DGES, Université Libre de Bruxelles (ULB)

to ongoing research and use (Billings and Johnson, 2012; Crawley *et al.*, 2013; Strotebeck, 2010; Tian, 2013). Such location quotients are calculated based on data aggregated for spatial units which immediately raises a red flag concerning potential modifiable areal unit problems (Arbia, 2001a; Grasland, Claude and Madelin, 2006; Openshaw, 1983; Tiebout, 1956) In response to such problems, a growing body of literature proposes to move away from aggregated data and to work with precise location data, analysing that data in a continuous space (Arbia, 2001b; Duranton and Overman, 2005; Grasland *et al.*, 2000; Tobler, 1989).
Using firm data localised at the exact address, I experimented with some such techniques in continuous space. I develop two possible new indicators, based on existing work by Huff and by Marcon and Puech (Huff, 1964; Marcon and Puech, 2010).

Huff-like model

David Huff developed his model in the field of geo-marketing and retail geography (Huff, 1964). Even though the model was definitely not conceived for estimating exports by all industries, it does allow applying a similar logic as the location quotient, but to a continuous space, thus taking into account the notion of distance.
The approach presented here entails some very simplifying approximations such as the idea that all consumption – including intermediate – can be modeled by population and that one can determine a radius of sales of a production site. These simplifications will, at least partly, be replaced by more differentiated approaches in future work.

The following formula provides the central indicator of Huff's original model, i.e. the probability of a given point i in space to consume from a given point j of retail in a given sector:

$$P_{ij} = \frac{A_j / D_{ij}^{\gamma}}{\sum_j \ (A_j / D_{ij}^{\gamma})}$$

where A is the attractiveness of the point of retail, D is the distance between point i and retail point j, and γ is a parameter allowing to modify the importance of distance. This probability is calculated using all retail points of a similar sector.

In the enlarged framework of the estimation of exports of all types of production, this probability becomes the probability in any given point i in space of using the production of the unit j in a specific sector amongst the total offer of production in that sector.

One can then estimate the population in i that uses the production of j with the simple formula:

$$Pop_{ij} = Pop_i * P_{ij}$$

This formula postulates a closed economy and homogenous consumption across the entire reference space. Isserman proposed a series of solutions in the context of location quotients that can be applied here, i.e. modulate population by income and probability of consumption by the international trade balance within a given sector (Isserman, 1980):

$$C_{ij} = (Pop_i * I_i) \cdot (P_{ij} \cdot (1 - e_s))$$

where Cij is a per firm consumption estimator taking into account population (Pop_i) and income, with I_i a ratio of per pixel income to reference space income, and e_i defined as the

ratio between trade balance and output for sector i in the reference space.
For each unit of production j (or spatial unit j containing a certain number of production units) one can then sum the total consumption catered to:

$$C_j = \sum_i C_{ij}$$

In the case when this is useful, it then becomes possible to aggregate the population catered to by a specific spatial unit u, by summing total populations catered to by all production units localised within that spatial unit:

$$C_u = \sum_{j \in u} C_j$$

By doing this for each sector one can estimate whether the production units in the spatial unit produce for a population that is larger or smaller than the residential population of that spatial unit. In the first case, one can postulate exports, in the second case imports.

M function-based model

Marcon and Puech provide an overview of different indicators of concentration, many based on Ripley's K function (Ripley, 1976, Marcon and Puech, 2012). Their own M function differs from Ripley's K in that it provides the possibility of a relative measure – a location quotient in contiuous space - and that it allows the weighting of points (Marcon and Puech, 2010):

$$M_s(r) = \frac{\sum_i \left(\frac{\sum_{j \neq i} (c_s(i,j,r) \cdot w_j)}{\sum_{j \neq i} (c(i,j,r) \cdot w_j)} \right)}{\sum_i \left((W_s - w_i)/(W - w_i) \right)}$$

where $M_s(r)$ is the M function for sector s and a given radius r, $c_s(i,j,r)$ and $c(i,j,r)$ are dummy functions which equal 1 when plant j is within radius r of plant i, and 0 when not, and w_j is the employment in a given pixel. W_s and W are the total weights of respectively sector s and all sectors together within the entire space of reference.
I rearranged the function in line with similar rearrangement of the location quotient to estimate exports (Isserman, 1980) and applied the resulting function to all pixels of 1km^2 where production sites are located, summing the employment of all sites within the same pixel. Contrary to Marcon and Puech, I did not substract local employment from the total reference economy employment. Thus, the equation becomes:

$$X_{sr} = (L_{sr}/L_{sn} - I_i \cdot (1 - e_s)) \cdot L_{sn}$$

where L_{sr} equals $\sum_{j \neq i} (c_s(i,j,r) \cdot w_j)$ with w_i being employment and L_{sn} equals W_s. Any value of $X_{sr} > 0$ is taken as being exported.
To estimate the rate of exports RX_{sr} within the radius r and sector s, I divided the result by total employment within the same radius:

$$RX_{sr} = X_{sr}/L_r$$

This value can then be applied to the central pixel and allows to calculate the total export-oriented employment in that pixel. Summing this result by spatial unit does not, however, have the same meaning as for the huff-like model as the sum estimates exports beyond a certain radius, not beyond any given spatial unit.

Data and tools

Both models were tested with 2010 data on production units from the Belgian firm register (DGSIE, 2010). Each unit was geocoded to the closest point on a 100mx100m grid by the data provider using parcel data and includes a 5-digit NACE code and total employment. For the analyses I used a 1km resolution. I conducted all analyses in continuous space using GRASS GIS (Neteler *et al.*, 2012) and calculations for aggregated spatial units in SQL in the PostgreSQL relational database management system (PostgreSQL Global Development Group, 2014). All source code of the analyses is available on request from the author.

Results

Place is too limited, here, to present extenisve results, but two illustrations should allow to get a vision of possibly results. Figure 1 illustrates the M-function based model results for a 45km radius.

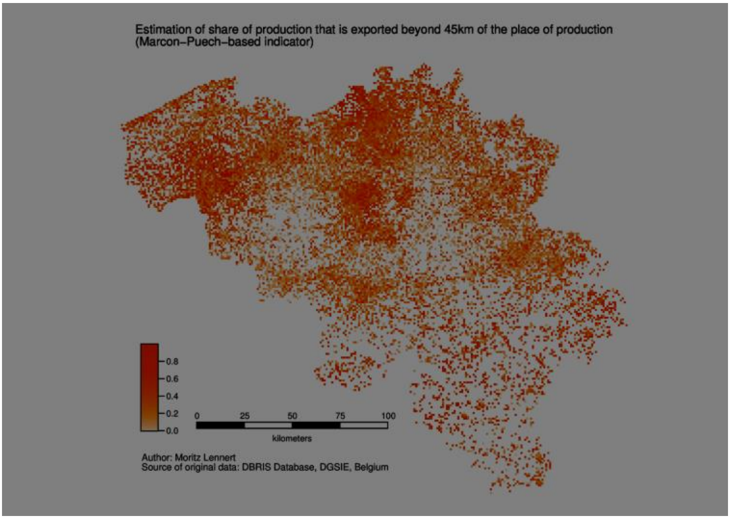

Figure 1 - Example map of Marcon-Puech-based indicator

Figure 2 shows a comparison between the results of a Huff-like model and a simple location quotient in large functional urban areas in Belgium. Results are similar (Spearman correlation coefficient of 0.76), but show a slightly more centralized geography for the Huff-like model.

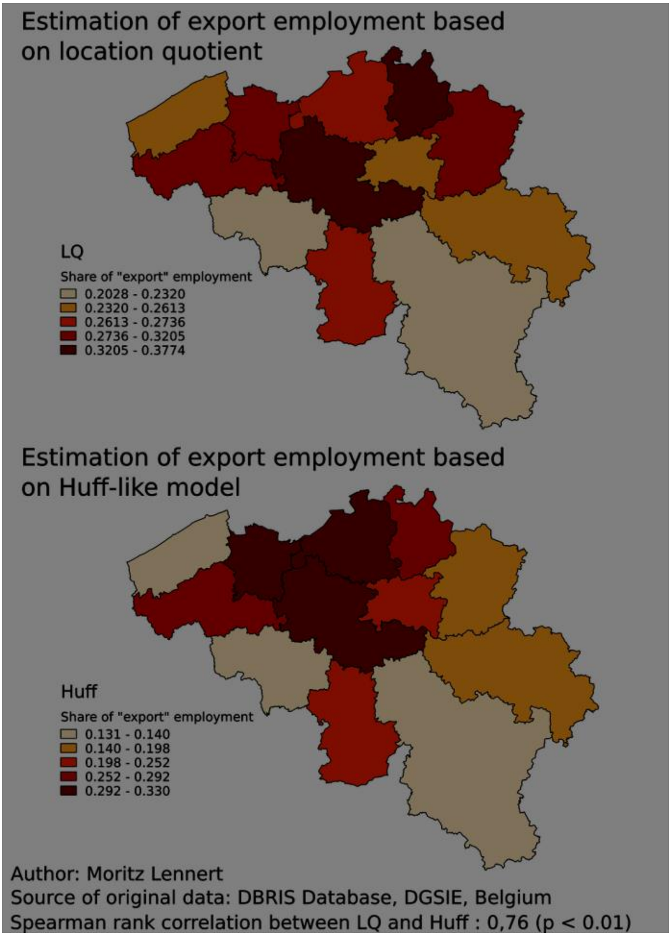

Figure 2 - Comparison of Huff-like model results with results of simple location quotient

These maps illustrate the variety of solutions possible with such models, including reaggregation in spatial units.

Conclusions

In this short paper I presented the basis of two new models for estimating exports in continuous space. A lot of work remains to be done testing these models and improving their results and usefulness. However, they are a promising start on the road to shedding the restrictive MAUP-causing spatial delineations. Further work includes assessing the robustness of the models in face of changing input data, and in other geographical configurations, as well as better understanding their general behaviour and meaning.

References

Arbia G. (2001a), The role of spatial effects in the empirical analysis of regional concentration, *Journal of Geographical Systems*, 3, pp. 271–281

Arbia G. (2001b), Modelling the geography of economic activities on a continuous space, *Papers in Regional Science*, 80, pp. 411–424

Billings S.B., Johnson E.B. (2012), A non-parametric test for industrial specialization, *Journal of Urban Economics*, 71, pp. 312–331

Crawley A., Beynon M., Munday M. (2013), Making location quotients more relevant as a policy aid in regional spatial analysis, *Urban Studies*, 50, pp. 1854–1869.

DGSIE (2010), *Répertoire d'entreprises (DBRIS)*, Bruxelles, Belgique.

Duranton G., Overman H.G. (2005), Testing for localization using micro-geographic data, *Review of Economic Studies*, 72, pp. 1077–1106

Grasland C., Madelin M. (2006), The modifiable area unit problem, *Final Report No. 3.4.3*, ESPON.

Grasland C., Mathian H., Vincent J.M. (2000), Multiscalar analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences, *Statistical Journal of the United Nations Economic Commission for Europe*, 17, pp. 157–188.

Huff D.L. (1964), Defining and estimating a trading area. *Journal of Marketing*, 28, pp. 34–38.

Isserman A.M. (1980), Estimating export activity in a regional economy: A theoretical and empirical analysis of alternative methods, *International Regional Science Review*, 5, pp. 155–184

Marcon E., Puech F. (2012), A typology of distance-based measures of spatial concentration, *HAL Archives Ouvertes*, halshs-00679993v2.

Marcon E., Puech F. (2010), Measures of the geographic concentration of industries: Improving distance-based methods, Journal of Economic Geography, 10, pp. 745–762

Neteler M., Bowman M.H., Landa M., Metz M. (2012), GRASS GIS: A multi-purpose open source GIS, *Environmental Modelling & Software* 31, pp. 124–130.

Openshaw S. (1983), *The Modifiable Areal Unit Problem*, Norwick, Geo Books.

PostgreSQL Global Development Group (2014), *PostgreSQL*.

Richardson H.W. (1985), Input-output and economic base multipliers: Looking backward and forward, Journl of Regional Science, 25, pp. 607–661

Ripley B.D. (1976), The second-order analysis of stationary point processes, *Journal of Applied Probability*, 13, pp. 255–266

Strotebeck F. (2010), *The Location Quotient – Assembly and Application of Methodological Enhancements* [http://mpra.ub.uni-muenchen.de/47988/].

Tian Z. (2013), Measuring agglomeration using the standardized location quotient with a bootstrap method, Journal of Regional Analysis & Policy, 43, pp. 186–197.

Tiebout C. (1956), Exports and regional economic growth, Journal of Political Economy, 64, pp. 160–164.

Tobler W.R. (1989), Frame independent spatial analysis, in Goodchild M.F., Gopal S. (eds.), *The Accuracy of Spatial Databases*, London, CRC Press, pp. 115–122.