Working title: The use of the Midelfart-Knarvik model to study historical case studies

Authors:
Stijn Ronsse (University of Ghent)
Glenn Rayp (University of Ghent)

Abstract

The last couple of decades, economic historians increasingly became more interested in the use of an empirical framework to study industrialization patterns in the 19th and 20th century and to test the relative importance of two main location theories. The first one, the Hecksher-Ohlin theory, basically holds that under perfect competition, non-increasing returns to scale, the absence of trading costs and for homogeneous products, industries tend to settle near regions that offer demanded endowments. The New Economic Geography in its turn implies that under intermediate transportation costs, economic activity concentrates near regions with a high market potential. As both are not mutually exclusive and considering the main role of transportation costs, combining them in a nested model offers a suited thinking scheme to analyse the geographical distribution of economic activity throughout the 19th and 20th century. As this is exactly what the intuitive Midelfart-Knarvik model (2000) does, the last two decades it became paradigmatic to study historical industrial locations. From this point of view, in this chapter, we explore the potential of the model, the included mechanisms, its use and its pitfalls.

1. Introduction

The last couple of decades, both economists and historians started to show more interest in the study of industrial locations. From an economic point of view, this renewed interest already started in the 90's of the past century, after Paul Krugman theorized his ideas on the importance of regional market potential, known as New Economic Geography. Although early economists already showed interest in location theories in the 19th and 20th century, they rather soon settled with the classical Heckscher-Ohlin (HO) idea of the importance of natural endowments in explaining industrial locations. Throughout the rest of the 20th century, spatial issues were basically regarded as minor issues.

However, as stated, the last three decades, economists again started to show interest in location theory. This was not only the result of the upcoming field of the New Economic Geography, McCann (2003) also highlights the importance of new international institutional arrangements, the renewed interest within the private and the governmental sector, the growth in information, data availability and changing academic fashion and trends. In terms of the latter, the field became more interdisciplinary oriented and started to span several disciplines.

In this manner, historians also joined the debate. For them however, interest in the field was not new at all, curiosity in locations of industry never decreased since the beginning of the 20th century. Under the impact of the success of business and entrepreneurial history, the study of locations has always occupied many historians. Nevertheless, the last couple of decades, the field also started to change. Not necessarily in terms of interest, but more in terms of methodology. More specifically, attention increasingly shifted from analyzing case studies towards quantification and testing of theories. As the use of empirical models became more popular, the distinction between purely economic research and business history became less clear.

Retrospectively, we believe that the publication of MidelFart-Knarvik in 2000 was an important milestone for the interdisciplinary study of the distribution of economic activity. Although the authors studied contemporary location patterns of European industries, the used model became increasingly popular among both economists and historians. Throughout the past two decades, it became one of the main frameworks to study historical industrialization patterns. From this point of view, we want to contribute to this book by discussing its potential and its limitations in more detail. In what follows, we will first discuss location theory in general, to continue with the discussion of the model itself. In a next step, we will highlight the models limitations when it is used for the study of historical case studies. More specifically, in the third section of this article, we will discuss the potential econometric errors, the process of variable selection and the interpretation of the outcomes of the model. Lastly, we will also suggest some ideas how future researchers can try to deal with these.

2. Location of industries: the conceptual and empirical framework

2.1. Theoretical background

Despite some early 19^{th} century contributions, the relation of economic thinking with respect to space, was essentially one of neglect (Combes et al., 2009). In traditional economic theory, the only relevant distance was that between nation-states, i.e. the border that demarcates countries and then mostly in terms of differences of internationally immobile labor and capital or of technological knowledge. *Within* nation-states, production factors were traditionally assumed to be fully mobile, as if distance had vanished with the abolition of local tolls and the improvement of the transportation network.

The explanation why might be theoretical: including space in a modelling framework implies that the assumptions of perfect competition and constant returns to scale may have to be abandoned, something economists were reluctant to do, as they were the cornerstone of the major formal theoretical achievements in economic science (like e.g. the Arrow-Debreu model).

The incompatibility of space with perfect competition and constant return to scale is the essence of the *Starret Impossiblity Theorem* (Starrett, 1978). Space implies indivisibilities (a finite number of locations) and transportation costs (between locations). Starrett showed that a perfectly competitive equilibrium with trade between locations (transportation of goods) is then impossible. Suppose there is one consumer and one firm, each occupying a (*homogeneous*) space of one unit, too small to accommodate both the firm and the consumer. Suppose in addition that moving people and goods from the one unit location to the other is costly. The firm produces one unit of a good, using labour supplied by the consumer. The wage earned by the consumer is used to buy the good, the revenue of which allows the firm to earn a profit. In addition to the transportation costs of the good and the worker/consumer, the firm and the consumer pay a rent for their location that the consumer collects. Consumer and firm are respectively utility and profit maximizing agents. Then, they will always have an incentive to move closer to each other (i.e. to share the same location) in order to avoid the transportation costs, which is however impossible as firm and consumer cannot use less than one unit of space. Hence, there is no equilibrium: the agents have a permanent incentive to move wherever they are located.

The question then is, is there a conceptual ways out of the Starrett Impossibility Theorem? To make this possible, either perfect competition and constant returns to scale are need to be maintained by assuming *exogeneous differences* in space such that all locations are not equivalent anymore. This is precisely what the HO-model of international trade does by assuming given differences in factor endowments between countries. In relation to industrial locations, the theory (published in 1933) then holds that these are determined by the natural endowments and production factors of a region. Under perfect competition, non-increasing returns to scale, the absence of trading costs and for homogeneous products, underlying factors determine the industrial geographical pattern (Brülhart, 1998). For example, historically, the model would expect that industries in high demand of coal would settle in regions endowed with mines.

A second alternative, still consistent with perfect competition, is the assumption of Marshallian externalities, from which endogeneous differences in space emerge, yet without specifying the explanatory mechanism. The central idea is that the industrial concentration in one region generates scale effects that are beneficial for the whole industrial agglomeration. The main mechanisms of this process are labor pooling, knowledge spill-over and close proximity to other industrial actors, allowing efficient forward and backward linkages (Lynn and Rodriguez-Clare, 2011). Both of these effects contribute in their turn to the self-enforcing effect of industrial agglomerations.

The third option is to abandon a competitive equilibrium and to allow for some form of imperfect competition (predominantly monopolistic competition). The latter is the approach of the so-called New Economic Geography (NEG). The NEG explains geographical patterns of economic activity from the interaction between increasing returns (actually pecuniary externalities) and transportation costs, summarized in what Krugman (e.g. 1991) defined as the *home market effect* (HME): conditional upon the transportation costs, firms (economic activity) will concentrate more than proportionately in the largest market, which is (*ceteris paribus*) the most profitable location (as it allows to sell a higher share of output without paying transportation costs and hence is the most attractive location for export as well). In this idea, location of economic activity is determined by centripetal and centrifugal forces. The centrifugal (or dispersion) forces refer to the higher competition that agglomeration of firms and economic activity imply (and therefore induce firms to move to avoid competition and earn higher margins). The centripetal forces, allowed in agglomerations, refer to the higher sales exempt of transportation costs to consumers and downward firms (forward spillovers), as well as the cheaper intermediate goods (by avoiding transport) that a firm can buy (backward spillover effects). At very high transportation costs (e.g. prohibitively high), the centrifugal forces prevail and an equilibrium with dispersion is expected. If transportation costs are reduced to insignificance, then location becomes irrelevant (therefore no bias against dispersion). Typically for *intermediate levels* of the transportation costs, agglomeration of economic activity (e.g. core-periphery patterns) will prevail. Hence, NEG-models expect that under the constant decrease of transportation costs, a bell shaped agglomeration pattern will emerge. Specifically, to explain the reallocation of industries, the theory holds that under decreasing transport costs, industries get incentives to locate near strong markets, close to customers and other industries. Hence, regions with a strong market potential attract industries and economic activity. In this perspective, the benefits of the agglomeration effects are larger than the cost of transporting resources.

In terms of the study of historical case studies, all three theories look appealing as both historical and contemporary examples indicate the importance of both natural endowments and agglomeration effects. For traditional industrial regions, it is expected that the initial kick-off can be determined by endowments, leading to concentration of industry and hence allowing for agglomeration effects. However, when locational patterns of the last two centuries are studied, the impact of transportation costs also needs to be considered. The industrial revolution was characterized by a spectacular fall in transportation costs (due to innovations in shipping and the introduction of the railway). This evolution continued throughout the 20th century. Infrastructure all over Europe and in the US multiplied and technological progress intensified, resulting in a further decrease in transportation costs and a dramatic change in economic activity. As a result, not only endowments and agglomeration, but also the NEG-framework were taken up by historians to understand location of economic activity.

2.2. *Empirical Implementation*

As stated above, the home market effect was a crucial element in this last framework. The first to empirically examine this effect and hence the relevance of the NEG, was Krugman (1980). From his theoretical model, it follows that a regression of share of producers on the share of demand identifies the home market effect. To do so, he had to made some restrictive assumptions (in particular factor price equalization between regions and identical fixed costs), which can be supported for within country location.

Later onwards, Davis and Weinstein (2003) argued that they could identify the home market effect from a regression with production of a good k in a country r as dependent variable. Two explaining variables are included in the model. Firstly, the production of good k in country r considered that the share in production of good k in country r would be the same as in the rest of the world. Secondly, a variable with *idiodem* as acronym, defined as the deviation of good k in the expenditures of country r with respect to the share of expenditures in the rest of the world. The coefficient of the latter variable is considered to be informative about the home market effect. As external control variables Davis and Weinstein added the endowments in country r of labour, capital and land. Hence, the model allowed to empirically estimate the impact of both NEG- and HO-factors. The test on the coefficient of *ideodem* is exclusive: either they conclude from a significant home market effect (i.e. a parameter estimate significant higher than 1) to a NEG model or, to a constant returns to scale Heckscher Ohlin model (when the estimated parameter of *ideodem* is significantly smaller than 1).

However, although the above mentioned study also combined both the NEG- and HO-theory, this is not the model that would get a more or less paradigmatic position in the historical analysis of location. In their analysis, economic historians tend to prefer a model of location in which the two main mechanisms, exogenous differences in space (endowments, the Heckscher-Ohlin determinants) as well as transportation costs and agglomeration effects (the NEG determinants) can be integrated, for example to evaluate their respective impact on location. To intuitively to do, the Midelfart-Kvarnik et al. (2000) model suited perfectly. (henceforth MK). As a result, since it's publication, it has been repeatedly been used in country specific studies[1]. This not only allows to compare the relative impact of both mentioned theories on industrialization patterns, but also to compare the impact on different case studies.

3. The MK-model

3.1. Specification

MK assume perfect competition and constant returns to scale production. Preferences are characterized by love-of-variety, represented by a CES utility function (hence product differentiation) and each industry produces an exogenously fixed number of varieties[2]. Goods are mobile between regions but subject to an industry and region specific iceberg type transportation cost.

From these assumptions, it follows that the value of production in each industry and region is determined by factor supply prices, intermediate goods prices and the geographical distribution of demand. The number of varieties a region can produce in an industry is assumed to be proportional to size of the region and the industry, more specifically the region's and industry's share In total production. For the empirical model, the value of production in region and industry (as the industry-region production share) is twice normalized by the region's and industry's share in total production.

By assuming that the input factor prices depend on endowments, the model captures the Heckscher-Ohlin determinants of location. The presence of transportation costs (and an exogenous assumed number of varieties) implies that geography and location of demand are included as determinants of regional industrial activity as well. The *market potential* of each region will therefore matter for the spatial distribution of economic activity. The market potential of a region consists of two components: final goods demand and intermediate goods

[1] At present for the UK, Spain, Poland, Belgium and Italy.

[2] In view of the Starrett impossibility theorem, the assumption of perfect competition and constant returns to scale in a framework highlighting the location choice, may be surprising. One may ask how with these assumptions, regional autarky is avoided. Assuming constant returns to scale implies that there is no scale penalty and the goods can be equally efficient produced at any arbitrary scale. Perfect divisibility then allows to produce all varieties locally, thus avoiding transportation costs and resulting in autarkic regional equilibria. However, this is excluded in MK by assuming an exogenously limited number of varieties that can be produced in each region and industry. Love of variety then implies that there will be trade between regions (there will always be a positive demand for a variety, whatever the price) and, therefore, that location matters.

demand by other producers. The latter constitute the *forward linkages* (downstream demand of the goods a firm produces) between industries. Finally, given that in production intermediate goods are used as an input as well and constitute an argument of the unit cost function, *backward linkages* between industries are included as determinant of location as well. In regions of bigger size, firms are able to buy more intermediate goods exempt from transportation costs.

After linearizing the equation of (normalized) value of production around a reference level in unit costs and market potential, MK obtain the following empirical model, for industry *k* in region *i*:

$$ln\left(r_i^k\right) = \sum_j \beta\,[j](x_i[j] - \bar{x}[j])(y^k[j] - \bar{y}[j]) + \varepsilon_i^k$$

After working out the products in the summation, the following specification is obtained:

$$ln\left(r_i^k\right) = \xi + \sum_j \langle \beta[j]x_i[j]y^k[j] - \beta[j]\bar{x}[j]y^k[j] - \beta[j]\bar{y}[j]x_i[j]\rangle + \varepsilon_i^k,$$

where $\xi = \sum_j \beta[j]\bar{x}[j]\,\bar{y}[j]$ does not vary with *I* or *k* and therefore represents the constant of the model. In economic historical literature, the original model is adapted and the following specification is obtained:

$$Ln(s_i^k) = \alpha \ln(pop_i) + \beta \ln(man_i) + \sum_j (\beta[j]y[j]_i z[j]^k - \beta[j]\gamma[j]z[j]^k - \beta[j]K\,[j]y[j]_i) + \varepsilon_t;$$

On the left side, s_i^k is the share of industry k in region i at a certain time t, because of data availability most often estimated in terms of employment. pop_i and man_i stand for the size of the population and the number of employment in manufacturing in each region and are included to control for size effect. More specifically, these variables capture the probability that larger regions have a larger industrial base. $y[j]_i$ is the strength of a regional characteristic j in region i, $z[j]^k$ is the industrial characteristic. Hence, the model estimates the importance of interactions between several regional and industrial characteristics. An interaction for example can be the presence of a skilled labour force in one region interacted with the demand for skilled employers in a certain industry. $\beta[j]y[j]_i z[j]^k$ is the interaction between industrial and regional characteristics. $\beta[j]\gamma[j]z[j]^k$ and $\beta[j]K[j]y[j]_i$ are included to respectively check the impact of industrial and regional characteristics. For example, $\beta[j]\gamma[j]z[j]^k$ can be the demand of a certain industry for endowments out of mines, $\beta[j]K[j]y[j]_i$ is then the number of mines in a region and $\beta[j]y[j]_i z[j]^k$ represents the interaction between both. Essentially, all studies using this model are interested in the impact of these interaction terms. Most often, the following variables are included in the model, although it is open to other factors. Wolf (2007) for example also included innovative factors as explaining variables.

Table 1. Industrial and regional characteristics

Regional characteristics	**Industrial characteristics**
HO-factors	
Agricultural surface	Agricultural input
Mines	Input of mines
Active population	Labour intensity
Educated population	White collar intensity
NEG-factors	
Market potential	Use of intermediate goods in % of output
	Sales to industry in % of output
	Size

Source: Ronsse and Rayp (2014)

In the left column, we find for each region the agricultural surface, the number of mines, the size of the active population, the size of the educated population and the market potential. This last variable was first proposed by Harris (1954) and is used to date to include the accessibility of the regional markets. In this theoretical framework, it is an estimate of closeness to consumers and other industries. In the column on the right, we find the industrial demand for agricultural input, inputs out of mines, labour intensity, white collar intensity, the demand for intermediate goods, the dependence of sales to other industries and the size of an industry. Based on these variables, the following interactions are included in the model.

Interactions included in the model
HO-factor
Agricultural surface x Agricultural input
Mines x Input of mines
Active population x Labour intensity
Educated population x White collar intensity
NEG-factors
Market potential x Intermediate goods
Market potential x Sales to industry
Market potential x Size

3.2. Case studies

The past two decades, the Midelfart-Knarvik model was used by several researchers to test the relative importance of HO- and NEG-determinants in explaining the pattern of industrial locations. Up to date, patterns cannot be generalized and the results are determined by the national context and the studied time frame. However, most often it is concluded that both theories are not mutually exclusive and both have an impact on the industrial pattern of the analysed case study, a result that was also found in the original article of Midelfart-Knarvik (2000).

Examples are the studies of Wolf (2007), Martinez-Galaraga (2012), Crafts and Mulatu (2005) and Klein and Crafts (2012), respectively on industrialisation patterns in Poland, Spain, the UK and the US. More recently, we also contributed by using the model to study the shifting industrialisation pattern in Belgium.

Wolf (2007) analysed how the unification of the internal market combined with investments in infrastructure and changing endowment (relocation of human capital) influenced industrial relocations in Poland. He concluded for the period 1902-1937 the importance of both skilled labour and inter-industry-linkages. Crafts and Mulatu included the end of the 18th century in their study (1871-1931) and concluded the importance of the proximity to mines and human capital. However, they argue that their results probably indicate a pattern of path dependence, in which the initial location decision was based on endowment, but that these regions also became industrial centres and hence offered scale benefits. They stated… *the implication is that 'factor endowment' variables do not always equate to a pure HO story but sometimes have NEG overtones.* (Crafts and Mulatu, 2005).

Klein and Crafts (2012) studies the persistence of the manufacturing belt in the US for a similar period (1880-1920) and concluded a stronger impact of the NEG- compared to the HO-determinants. Of the former, both scale effects and forward and backward linkages were significant, while the regional abundance of land only had a small impact on the location decision of a subset of manufacturing. Martinez-Galarraga (2012) studied a similar time frame as Crafts and Mulatu) and questioned the relative importance of both the HO- and NEG-theory in explaining the Spanish industrial concentration pattern. In terms of comparative advantage, the abundance of land, measured in terms of regional agricultural production, had the biggest impact. However, the increasing centralisation of Spanish industry was explained by the increasing importance of NEG-factors. An important finding in this study is that the found equilibria were determined by the studied year and shifted constantly throughout time. In terms of interpretation of the MK-model, this conclusion warns for an overestimation of the stability of the found equilibrium.

More recently, we contributed to the debate by studying the shifting industrial pattern in Belgium during the 20th century. We concluded that NEG-factors determined this shifting pattern. At the end of the 19th century, endowments were still crucial as industries could be found near the mines in the southern part of the country. However, as a result of the internationalisation of the port of Antwerp, the market potential of northern regions increased, leading to the formation of agglomerations throughout the 20th century (Ronsse and Rayp, 2015).

Concluding this section, most studies found the increasing importance of NEG-factors in determining industrial locations. The above literature overview however indicates that two nuances should be considered. First of all, the result heavily depends on the studied time frame and context and it seems that a constantly shifting equilibrium should be considered. This is for example interesting when the question is raised whether these case studies can contribute to the understanding of modern industrialisation patterns. Secondly, as indicated in Crafts and Mulatu (2012), the interaction between HO- and NEG-factors and path dependence deserves more attention. Although most of researchers are limited by data availability, it could be interesting to study the importance of comparative advantage at the initial start-up of agglomerations and regions rich in market potential.

4. Limitations

4.1. In terms of data

Although Midelfart-Knarvik offers an interesting model to test for the relative importance of both the NEG- and HO-theory, researchers should be aware that there is a plethora of limitations. These are not only related to the limited availability of data, but there are also specific econometric pit-falls. Dealing with these is crucial to determine the relative impact of the mentioned theories.

First, the model is demanding in the amount of qualitative data on the most detailed geographical level possible. As most articles study 19th and 20th century evolutions, the most often used sources are: population, agricultural, socio-economic censuses and input-output tables –at max performed every ten years. This results in two types of problems. First of all, a broad range of assumptions and extrapolations is demanded to have data on all desired data. some variables can readily be found, for others, proxies are a necessity. The easiest variables are these straight from the sources: the number of mines and agricultural surface. Also the dependant variable is often found in these sources, in the form of employment in industrial sectors. For educated population, correlation tests between number of literates and school for the beginning and the end of the period are necessary.

Secondly, the measurement of market potential can be discussed. As stated by Crafts (2005) there are several options to estimate the openness and distance to consumers and other industries, i.e. Market Potential. The first and still widely used specification is that of Harris (1954). As the formulae shows, it is very intuitive as it is the sum of the sizes of the region itself and all regions nearby, divided by the distance to the base region. The formula is specified as:

$$MP_i = \sum_j (\mathrm{M_j}/\mathrm{D_{ij}});$$

With M as the size of the region and D as the distance between the regions.

As indicator of size , most studies use GDP, which implies that data are required of all regions and even, in order to include export market potential of at least the most important trading partners. If historical regional GDP data are not available as is most often the case, a good alternative seems the Geary and Stark (2002) approximation (see e.g. Crafts and Mulatu,2004), who estimate the GDP of region *I*, defined as:

$$Y_i = \sum_k y_{ik} L_{ik}$$

where y_{ik} is the average output per worker in industry *k* and region *i* and L_{ik} employment in industry *k* and region *i*, by using:

- the national average of output per worker in industry *k* y_k;
- assuming that $\frac{y_{ik}}{y_k}=\beta_k \frac{w_{ik}}{w_k}$, i.e. average output per worker in industry *k* and region *i* in term of the national average in industry *k* is proportional to the ratio of the wage in industry *k* and region *i* and the nation-wide wage in industry *k*. Here is the factor of proportion as such that the GDP of industry *k* at the country level (Y_k)equals the sum of the regional estimations, i.e.:

$$\beta_k = \frac{Y_k}{\sum_i y_k \frac{w_{ik}}{w_k} L_{ik}}$$

Then :

$$Y_i = \sum_k y_k \frac{w_{ik}}{w_k} L_{ik}$$

Provided that the data required are available, the Geary and Stark (2002) approximation may be useful to estimate the GDP's at the regional level of the neighbouring countries of continental countries and this way, to proxy export market potential better than just taking the national GDP (or income), weighted by distance.

An alternative for the Geary and Stark (2002) method to estimate market potential is estimating a gravity equation of bilateral trade, from which it can be derived. For two regions (countries) r and s, the exports from r to s are modelled as (e.g. Combes et al., 2009):

$$\ln X_{rs} = FX_r + \ln \phi_{rs} + FM_s + \varepsilon_{rs}$$

where ϕ_{rs} represents the distance between regions (countries) r and s. FX_r and FM_s are respectively exporter and importer fixed effects. From the Dixit-Stiglitz-Krugman (DSK) NEG model, the latter is an estimate of the market potential of region s. Summing FM_sover all regions s then gives the market potential of region r. Wolf (2007) in his study of location of economic activity in interwar Poland is the only example where a gravity equation is used to estimate market potential (although not completely along DSK lines). Intraregional trade flows as data requirement is probably even more demanding than regional GDP data (Wolf, 2007 apparently had both). The advantage of the gravity approach is though that one obtains a proxy closer to the theoretical defined concept of market potential than from the more ad hoc Harris (1954) empirical approach.

In the distance discounting of regional GDP, Crafts and Mulatu (2004), like Martinez-Gallaraga (2012) correct for transportation modes by converting sea shipping in miles in a land based equivalent based on data of transportation costs. None of the existing studies, except Wolf (2005) seem to be able to take real market potential into account by deflating the regional GDP estimations in nominal terms[3]. One may argue that, to study location in a national context, deflating does matter that much insofar wages and prices within a country do not show too much interregional variance. However, it may matter to include the market potential of foreign regions or countries, because the assumption of international factor price equalisation and the law of one price for commodities is much harder to make. If we may consider that the error made by not deflating is identical for all regions considered, one could accept that this doesn't influence the relative market potential between regions, which is the one that matters. Pooling data for different years to analyse location of economic activity is of course disputable if market potential cannot be deflated.

[3] The theoretical defined market potential concept in the DSK-model of a region r is:$\mu_s Y_s P_s^{\sigma-1}$, where μ_s represents the share of non-traditional goods in total expenditure and P_s is associated the (ideal) price index, such that the market potential is defined in real, i.e. utility terms.

4.2. Econometric issues

As described in Ronsse and Rayp (2015), a first econometric pitfall is a result of the use of employment in an industrial sector as dependent variable, making it necessary to control for different levels of production across regions. As in Klein and Crafts (2012) and suggested by Wolf (2007) this can be dealt with by including regional and industrial dummies. As users of the Midelfart-Knarvik model are mainly interested in the impact of the interaction terms, this is an acceptable approach. Hence, the specification becomes:

$$\ln(s_{i,t}^{k}) = c + \sum_{ik,t}\beta_i^k\left([j]_{i,t}\ [j]_t^k\right) + \sum_i\gamma_i D_{i,t} + \sum_k\theta_k D_{k,t} + \varepsilon_t;$$

With $[j]_{i,t}\ [j]_t^k$ as the terms that we are interested in and $D_{i,t}$ and $D_{k,t}$ respectively as regional and industrial dummies. β_i^k are the coefficients that we are interested in, that is the impact of the seven interaction terms.

However, most of the applications of the Midelfart-Knarvik-model to historical location of economic activity does not quite apply the double normalisation of the dependent variable that follows from the theoretical model (see Midelfart-Kvarnik et al., 2000). Wolf (2005), Crafts and Mulatu (2004) and Ronsse and Rayp (2016) use regional employment data by industry as indicator of the location of economic activity, whereas Martinez-Gallaraga (2011) uses tax data. Only Wolf (2005) standardises the industry's share (s_{ik}) in the region's economic activity (employment) : $s_{ik} = {x_{ik}}/{\sum_i x_{ik}}$ with the industry's share in total economic activity: ${\sum_i x_{ik}}/{\sum_i\sum_k x_{ik}}$, as in Midelfart-Kvarnik et al. (2000), where the industry-region production share in economic activity is normalised by the region's and the industry's share in total activity. In the other studies, the share of regional population and the share of regional manufacturing are included as control variables in the empirical model, which goes part ways the MK-normalisation.

Though intuitive, to some extent the Midelfart-Kvarnik et al. (2000) normalisation is arbitrary, as it follows from an ad-hoc specification of the number of firms (n_i^k) in the model: $n_i^k = s_i s^k\, e^{\varepsilon_i^k}$, as a function of the share of region *i* and industry *k* in economic activity and a stochastic component ε_i^k. The assumption of a fixed number of firms as such is very specific to Midelfart-Kvarnik et al. (2000) and is a consequence of the perfect competition and constant to returns to scale assumption, which is not very common in the literature. In most NEG models, the number of firms (varieties) is endogenous and follows from the Chamberlin long-run equilibrium condition (zero net profits). However, no normalisation implies that the number of firms is implicitly considered as purely stochastic, i.e. $n_i^k = e^{\varepsilon_i^k}$, which is even more difficult to defend. In addition, if the ambition is to use the Midelfart-Kvarnik to compare the analysis of historical location of economic activity between countries, then some convergence in normalisation is indicated.

As regards the estimation of the empirical model, firstly, we notice that a restriction on the constant in the model does not seem to be taken into account (seemingly neither in Midelfart-Kvarnik et al., 2000). From the theoretical model, the constant is equal to $\xi = \sum_j \beta[j]\bar{x}[j]\,\bar{y}[j]$. The empirical model gives an independent estimation of $\beta[j]$, as well as $\beta[j]\bar{x}[j]$ and $\beta[j]\bar{y}[j]$, from which $\beta[j]$, $\bar{x}[j]$ and $\bar{y}[j]$ can be independenty identified. However, the restriction on the estimation of the constant term is not included. In absence of other methodological issues, this will affect the efficiency of the estimations (i.e. the standard error of the estimates) and bias downward statistical significance, but not their consistency.

Secondly, the industry characteristics included in the model are taken from available input-output tables that register the use of intermediate goods, sales to other industries and the use of factor inputs from which indicators of input intensity can be constructed. However, as input-output tables are a tool developed in the 20th century, the first editions in most cases go only back the earliest to the 1930s. For location of economic activity in the UK, Crafts and Mulatu (2004) were able to use input-output sources that give estimates for the 19th century (1871-1911 and 1851), but the other studies extrapolate input-output data from the 1930s of 1950s to the previous century. Obviously, this causes an error in variables, whatever the direction of the error (either an under- or overestimation of the factor and intermediate good intensities), such that the OLS estimate of $\beta[j]$ is inconsistent and biased towards zero (the *attenuation bias*), see e.g. Johnston and Di Nardo (1997). Even if the measurement error is uncorrelated with the stochastic error term of the empirical model, this implies an underestimation of the true significance of the variables in the model. In particular regarding the significance of the interaction terms, the attenuation bias can be very different between variables, given that the industry characteristics (derived from the input-output tables) are multiplied with those of the region. Given that in particular the interaction effects capture Heckscher-Ohlin as well as NEG determinants of location, the constraint of using ex-post input-output tables may bias the conclusions, first as regards the significance of the respective determinants of location (a higher type II error: erroneously not rejecting the zero hypothesis of insignificance) as well as an appreciation of their respective quantitative impact on location.

The Kim model (1995) can be used as a robustness test. Although it does not offer the same benefits as the MK-model –no interaction terms are included- it can be used as a robustness test. If the estimations indicate a similar evolution as the MK-model, one can be more certain that extrapolating the input-output tables is not problematic. The Kim-model (1995) is specified as:

$$ln(s_{i,t}^{k}) = \; + \beta_1 Agricultural\ Areal + \beta_2\ Mines + \beta_3\ Active\ Population$$

$$+\beta_4\ Educated\ Population + \beta_5\ Market\ Potential + \textstyle\sum_i \gamma_i D_{i,t} + \ \sum_k \theta_k D_{k,t} + \ \varepsilon_t;$$

In addition, scholars should be encouraged to build, if allowed by available data, input-output tables that go further back in time, e.g. along the lines of the sources cited in Crafts and Mulatu (2004). To the extent that the latter sources are available and provided that factor and intermediate good prices do not differ too much from the British in the second half of the 19th century, one could estimate the model using their factor and intermediate good intensities for the UK, as we may assume that industrial technology in European countries was comparable. If data on input intensities for the 19th century are however unavailable or impossible to build and conscious that other valid instruments for the factor intensities (i.e. uncorrelated with the error term, correlated with the input intensities and only influencing the dependent variable through their correlation with the input intensities) are unlikely, then one may try an econometric solution that we detail below.

Finally, including regional market potential as an explanatory variable of the location of economic activity, there is a clear case of simultaneity bias (the larger the region's share in industrial activity, the higher the region's market potential). In the different country estimations of the MK-model of location, it is the methodological problem to which most attention is paid to. Martinez-Galarraga (2012) and Ronsse and Rayp (2016) use an IV-estimation of the model (using respectively lagged variables of market potential and an indicator of centrality as instruments), and Wolf (2007) uses an exogenous political shock (the reconstitution of Poland from the three parts it was divided before the first World War). The instrumental variables estimation do not seem to give substantially different results compared to the OLS estimates, which may be considered as reassuring. However, to the extent that the exclusion restriction (i.e. the instruments affect only the dependent variable through their correlation with the potentially endogenous variable) is not valid, this may still be unsatisfactory.

To the extent that exogenous shocks, like the one considered in Wolf (2007), are scarce and valid instruments difficult to find, alternative approaches could be tried. Particularly interesting for quantitative historical analysis in our view is the method Lewbel (2012) proposes to cope with mismeasurement of variables or a simultaneity bias. Lewbel (2012) shows that if the estimated model shows heteroscedascticity in the error terms and, provided that restrictions can be imposed on the structure of this heteroscedasticy, then the correlation of (a subset of) the exogenous variables in the model with the second moment of the error terms provides valid instruments for the potentially endogenous or mismeasured variables. Besides of its ease of application, the main advantage of this approach is the redundancy of external variables as instruments, which can be important when data are not abundantly available like in a historical framework. More specifically, assume the following (triangular) model :

$$Y_1 = \beta_{01} + \beta_{11}X + \beta_{21}Y_2 + \epsilon_1, \qquad \epsilon_1 = \alpha_1 U + V_1$$
$$Y_2 = \beta_{02} + \beta_{12}X + \epsilon_2, \qquad \epsilon_2 = \alpha_2 U + V_2$$

In which Y_1 and Y_2 are endogenous (because of the common factor U in the error term) but where the set of exogenous variables, X cannot provide an instrument for Y_2, given that it is included both equations. Assume a subset Z of X exists, then the following conditions are sufficient for identification of β_{21} :

$$E(X\epsilon_1) = 0; \qquad E(X\epsilon_2) = 0; \qquad cov(Z, \epsilon_1\epsilon_2) = 0; \qquad cov(Z, \varepsilon_2^2) \neq 0$$

Which implies that the subset Z of variables in the model are valid instruments for the endogenous variables. Hence, theoretically, there is an econometric approach that could deal with one of the main estimation problems when the MK-model is used to analyse the distribution of historical economic activity.

5. Conclusion

Since the 1990s, under the impact of the increasing popularity of the New Economic Geography, economists started to show more interest in location theory again. One of the articles that was published in this slipstream, the Midelfart-Knarvik publication (2000) would have a major impact on the study of the historical distribution of economic activity. More specifically, it nested two main models that both tried to explain industrial locations. At the one hand, the Hecksher-Ohlin theory, holding the importance of natural endowments. On the other hand, New Economic Geography itself, highlighting the importance of closeness to market potential. Throughout the years following the publication, the MK-model would become paradigmatic in the empirical analysis of historical industrial locations.

The main advantage of the model is that it includes interaction terms that measure on the one hand whether industries that demand endowments tend to locate near regions that are rich in for example human capital, mines or agricultural land. On the other hand, interaction variables are included to control for the impact of market potential through scale effects and backward and forward linkages.

However, although the model is very intuitive, the empirical estimation has certain pitfalls that users should be aware of. In this chapter, we highlighted some of these and paid much attention to the use of input-output tables and a potential simultaneous bias. For both, it was suggested that additional and alternative efforts are advisable to obtain more robust estimations.