



Munich Personal RePEc Archive

# Methods for regionalizing input-output tables

Szabó, Norbert

MTA-PTE Innovation and Economic Growth Research Group

July 2015

Online at <https://mpra.ub.uni-muenchen.de/73947/>

MPRA Paper No. 73947, posted 23 Sep 2016 11:27 UTC

NORBERT SZABÓ<sup>a)</sup>

## Methods for regionalizing input-output tables

### Abstract

The present paper introduces the most common methods of regionalizing national input-output tables. First we describe the different groups of methods based on our review of the international literature regarding regionalization. Then we focus on particular methods that can be applied for Hungarian counties highlighting their advantages and disadvantages and synthesize the empirical results of them again based on the literature. On the basis of these experiences we attempt to create a complex method fitted to the available Hungarian regional data. For better understanding in the end we apply our method on an illustrative example consisting of three regions with hypothetical sectors and data.

*Keywords:* input-output table, regionalization, interregional trade.

### Introduction

The aim of current study is to introduce estimation methods of the regional input-output framework. Our starting point is the I-O table itself, which is capable of depicting the interindustry relationships of the whole economy. Roots of the theory can be dated back to the 1930s, to Leontief's (1934, 1941) pioneer work. Thanks to the rapid development of information technology it became possible for his contribution (the table and other derived indicators) to be applied to the analysis of different problems worldwide. The approach quickly gained a degree of popularity among researchers; today, more and more examinations have been conducted with new extensions (energy and environmental block), at different levels (national, regional, county).

The literature of national level research is enormous but because of spreading regional science, the focus of I-O research has been shifted from the national level to the smaller local level. The greatest disadvantage of purely national analysis is that they hide certain regional differences; they consider the whole economy as an aspatial entity. However, regional differences are not negligible at all. Moreover, the interregional trade connecting the regions of the economy plays a significantly different role in periphery and central regions. The consequence of taking into account aspects like these was that the initially national methods' regionalized versions soon appeared. In principle, there were no significant obstacles but during practical applications, researchers faced many problems. Since most European countries publish only national I-O tables, there is no detailed information about the inter-industry relations of regions, which would serve as the basis of much regional research. This is the main reason several regionalization methods were

a) MTA-PTE Innovation and Economic Growth Research Group H-7622, Pécs Rákóczi út 80., Hungary. E-mail: szabon@tk.pte.hu

elaborated that were capable of generating estimates of regional I-O relations. Since their first appearances, countless examples can be found in the literature. Most of them engaged with estimating regional tables, comparing them with actual empirical data and evaluating different methods highlighting their advantages and disadvantages.

This study has three aims: 1) introduce the regionalization methods of national input-output tables, highlighting those that are not too resource intensive and can also be applied to Hungary; 2) introduce the different possibilities of estimation of interregional trade flows; 3) propose a suitable method for generating multiregional input-output table in Hungary.

### Introduction of the I-O table, the core of input-output analysis

The aim of I-O tables is to depict the interrelated system of economic processes, production, consumption and savings. The input-output table is able to “depict the relationship between different sectors of the national economy, and the structural connection of production and final demand in a consistent manner” (KSH, 2005, p. 5). A basic requirement of the table is symmetry, which means that sectoral output and use have to be equal. The fundamental component is the intermediate transaction table, which depicts the flow of output between producer and user sectors. In Figure 1, it is matrix  $Z$  where the rows indicate the distribution of output produced by different sectors among other industries. On the other hand, columns indicate the combination of inputs required by each industry to produce its output.

Besides this, one can find additional rows and columns in the table. Supply can be used not only by industries (as intermediate use) but by final users as well. These transactions are not used directly in production processes as inputs (such as final demand by households and government, investment, export). In Figure 1, it is depicted by matrix  $FD$ . The additional rows consist of different value-added factors, and the further input requirements of industries, which are not provided by other industries (such as labor, capital inputs, indirect taxes, import).

All cells in the table can be interpreted as a flow of product in forints (millions) during a single year. In principle it is possible to assemble a physical input-output table that would be a better indicator of material use; However, the data requirements of such a table are enormous and the comparison between different industries would be difficult (small city car vs. truck).

Figure 1

*The scheme of an input-output table*

	Industries (as users)	Final demand	Total output
Industries (as producers)	<b>Z</b> “inner square”	<b>FD</b> “side wing”	<b>x</b>
Value added	<b>WA</b> “lower wing”		
Total input	<b>x</b>		

Source: Zalai (2012) p. 177.

So the mentioned matrices together are called an input-output table. The national tables can be classified into groups on the basis of many criteria. One of them is the way tables

depict imports. Imports and domestic products can be represented together. Thus, all domestic supply is considered in the table (type *A* and *C*). On the other hand, the distribution of the pure domestic production can be represented. In this case, imports are depicted in a separate row in the lower part of the table. Thus the three basic types of matrices are:

- **Type A:** domestic activities are not distinguished from imports. International trade is represented as a net export.
- **Type B:** rows depict only the pure domestic supply, and imports are accounted for in a separate row (by user industries) with no further detailed information about the sectorial source of the imports.
- **Type C:** in matrix *Z*, rows of production can separately contain both imports and domestic activities.

### Regionalization of input-output tables

The aim of the application of input-output tables and models was the estimation of expected economic effects of national shocks. To date, this approach has maintained its popularity. However due to development of science in the second half of 20<sup>th</sup> century, especially from the 1960s, the intensifying interest for regional applications and for other different related regional fields of regional science has led to modifications and extensions of the original I-O framework. One of the most important points is the adjustment of the methodology to smaller territorial units (region, county) by which it became suitable for depicting local economic characteristics. In most countries, this was (is) necessary because only national input-output tables are available, and the application of national coefficients in regional studies can severely distort results. In connection with this, there are two important points to be noted:

1) The national input coefficients can be interpreted as an average of regional technical coefficients. Thus, in principle, it is possible that the input structure of a region is identical to the country's as a whole, but can also be completely different. For example, electric energy production in Pécs is based on a thermal power plant and in Paks it is based on a nuclear power plant. Obviously the two technologies are significantly different.

2) A valid thumb rule is that the smaller a territorial unit (country, region or county), the greater its dependence on external territories through trade. The self-supplying ability of regions is shrinking, which means, in input-output relations, that the intraregional coefficients are smaller because a part of the inputs are purchased from other regions via interregional imports. Thus a new element, the interregional trade (export and import) has to be accounted for in the table.

At this stage, it is important to distinguish two kinds of coefficients that can be the result of regionalization (Miller and Blair, 2009):

- 1) regional input ( $A''$ ) and
- 2) regional technical coefficients ( $A'$ ).

The input coefficient depicts the level of inputs used by industries of a region that were produced by industries in the same region. The technical coefficient represents the level of inputs used by local industries regardless of the source of the input. From now on, the current study focuses on input coefficients.

There are several well-known variations of regionalization methods in the international literature. We cannot undertake the complete introduction of all methods; however, we will endeavor to describe more thoroughly the most important and most applied techniques. During the introduction, we will proceed from higher levels toward narrower groups of approaches to finally arrive at particular techniques.

Based on the underlying methodology, techniques can be divided into three fundamental groups (Greenstreet, 1989):

- 1) Survey-,
- 2) Non-survey and
- 3) Hybrid methods.

In case of survey methods, companies in the sample provide information about their sales (towards other companies and consumers) and their purchases (from other companies) within and outside the region. Industries in different regions can produce different products or use different inputs (like the case of thermal and nuclear power plant) for their production processes. Surveys can help researchers to take into consideration such special local characteristics. Thereby, the approach can create a precise picture of the technological and trade structure of the region; although, it has its drawbacks. 1) Usually, it is a very time and resource intensive process to conduct a survey. 2) The assembling of the sample is crucial; any mistakes at this point can lead to significant distortions in the final results. 3) Often, data gathered from companies is not balanced, and the results of balancing methods are sometimes not convergent (especially in case of smaller samples).

By applying non-survey methods, researchers can save time and resources. In this case, different estimation methods can generate regional tables (using national ones as a starting point). In many cases, the precision of procedures is enhanced by using additional regional data. A clear advantage of the non-survey approach is that these techniques can be implemented relatively easily if secondary data is available. Thus they are very cost-effective. Nevertheless, there is still no consensus about the “best” procedure to date; partially as a result of the questionable performance of different approaches. Furthermore, the lack of appropriate detailed regional data restricts the variety of potential methods.

The aim of the hybrid (or partial survey) approach is to combine the advantages of the other two mentioned approaches and to moderate their potential drawbacks. These procedures consist of several steps. The first of them is usually a non-survey model, which is augmented in further steps with survey data, expert estimates or information from other databases. Consequently, the hybrid approach is less resource demanding but still preserves the credibility of results.

The focus of science in this field has changed continuously during the last 60 years. During the 50s, survey, and later non-survey methods gained remarkable popularity; then, in the 60s, tables were mostly estimated by survey techniques. After that, the 70s saw an increasing number of non-survey studies conducted, which was the consequence of intensifying regional input-output research. Later, the step-by-step emphasis moved towards a hybrid approach.

By the millennium, it became clear that mainly hybrid and ready-made methods could generate reliable regional tables for a reasonable cost. The latter will be discussed later. Although hybrid techniques are characterized by cost-efficiency, relative precision and simplicity, non-survey methods did not disappear from regional research. In most

countries, hybrid methods are not applicable simply because there is not enough regional data available to run sophisticated models, or there are barriers that prevent the conducting of representative surveys. Fortunately, several new studies are promoting the potential further development of non-survey approaches.

Furthermore, methods can be classified by their choice of local unit:

- 1) Single region
- 2) Multiregional or Interregional methods

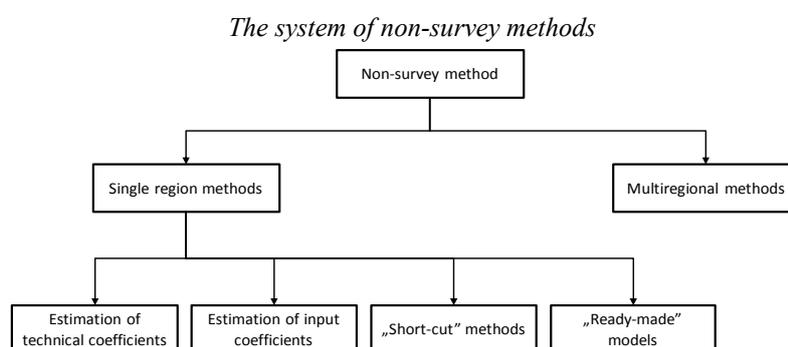
The detailed description of single region, multi- and interregional models are not part of this study; international literature has already discussed the topic in detail.

To choose the appropriate table, it is vital to match the aim of the research and model that will serve that goal. During our research, the initial intention was to estimate an interregional table that can depict how different industries are interconnected in a given region, moreover, how they are interconnected with industries in other regions. If the production of a region is affected by a policy shock, it will affect other regions through interregional linkages because regions sell to and buy from each other. Using such tables, we can take these spillover effects into account (Miller–Blair 2009).

Single region methods concentrate their attention on a specific local territory of a country. In this case, the region examined is separated from the rest of country. Suppose that the aim of our research is to estimate impacts of different policy investments in South-Transdanubia, Central Hungary and Western Transdanubia. The predicted impact of the single region models would fall short of the real effects. The reason is that regional interventions would affect the economy of other regions through interregional trade. Furthermore, the intervention would have further positive effects even in regions not affected by a policy shock (spillover effect). Thus, single region models are not capable of depicting the bilateral interconnections whilst a multiregional approach breaks down a nation into smaller regions and analyses both inter- and interregional linkages to give a more realistic picture of the economy.

During our research (for antecedents see: Varga et al., 2013) we did not consider the survey approach as a reasonable choice because of its high demand for time and resources. Based on the same reasoning, we did not choose partial non-survey methods either. In the rest of study, the focus is narrowed to purely non-survey methods. After Bonfiglio (2005), we can distinguish the following groups of non-survey methods:

Figure 2



Source: Bonfiglio (2005) p. 18

The first group consists of those methods that do not distinguish the flow of inputs between regions and within the region (technical coefficients). This group was dropped from the research because the aim was to depict interregional trade flows between regions. Since our interest is to estimate intra- and interregional flows as well, the second group of methods that seek to estimate regional input coefficients can be considered as an ideal choice. The purpose of short-cut methods is to predict the regional multipliers without estimating the actual input-output coefficients. One of the most well-known examples is the RIMS (Regional Industrial Multiplier System) model, which was elaborated by Drake (1976). Our research requires the assembling of an input-output table because our further aim is to extend an estimated table with additional blocks (e.g. environmental). Finally, ready-made models are available as an integrated part of a purchasable software package. These packages consist of valuable regional data, a regionalization method, and a ready input-output table. The software themselves can estimate regional I-O tables, multipliers, and forward and backward linkages. The most well-known examples are IMPLAN, REMI, RIMS II (Bonfiglio 2005).

In the next chapter, we will concentrate on those non-survey methods that are capable of estimating regional input coefficients. First we will describe single region methods, and then we will continue with methods of interregional trade. We will also place more emphasis on approaches that are suitable for Hungarian conditions.

### Single region non-survey methods

The basis of most non-survey methods is a national input-output table since there is no other available source of data. In the first step, we assume that technology is the same across regions and likewise in the nation as a whole. Then there is a wide variety of available methods to generate an estimate of regional coefficients. In general, these methods modify the production technology of regions (rows). There is no competition between local producers and firms in other regions. Interregional trade occurs when the region is not able to cover its needs for a given regional activity; the opposite is also true if the supply of a region can meet its needs, then any potential surplus must be an interregional export to other parts of country. Such a net approach of interregional trade can result in a possible underestimation of interregional trade coefficients and an overestimation of intraregional input coefficients (Bonfiglio–Chelli 2008).

### Regional Supply Percentage

Early works focused on using national coefficients with slight adjustments. An example of the earliest pioneer works is the regional supply percentage, which expresses the proportion available output that originated within the region and output regardless of its origin. The following equation is a general form of *RSP* index:

$$rsp_i^r = \frac{(x_i^r - e_i^r)}{(x_i^r - e_i^r + m_i^r)}$$

where the numerator depicts the difference of locally produced products ( $x_i^r$ ) and regional exports ( $e_i^r$ ), in the denominator one can find the amount of available products (including

imports( $m_i^r$ ) in region  $r$ . The intraregional coefficient matrix can be derived by the simple multiplication of the diagonal matrix of sectorial  $RSP$ s and the national coefficient matrix:

$$A^{rr} = \overline{rSP^r} A^n.$$

Modification across rows is a strict restriction, which means that we do not make a difference between national and regional technology. In other words, the proportion of input  $i$  used in production is the same at both territorial levels. A special variety of the method is the regional purchase coefficient ( $RPC$ ), which is originated from Stevens et al. (1983). The calculation is almost the same as  $RSP$  but the biggest difference is the source of data. In this case, data is generated through an econometric estimation. Again exact details are not described here due to length limitations.

Since most non-survey methods estimate regional trade as a residue after consumption and intermediate use, which can generate serious distortion, the  $RPC$  method can be considered as a promising approach. Stevens et al. (1983) defined  $RPC$ s as a function of transport cost and unit cost of products. Despite its theoretical advantages, the approach did not gain popularity due to its high data requirements, which usually cannot be satisfied at regional level.

### Methods based on location quotient

Another early group of methods is based on the well-known location quotient. The aim of the approach is to use  $LQ$  as an indicator of sectorial regional specialization and to use it to transform national technology into a regional one. There are plenty of different  $LQ$  variants in the international literature. In the next sub-chapters, we will describe the most important features of different  $LQ$  methods.

#### Simple location quotient

The first and most simple variant is the unmodified traditional location quotient. This ratio can be calculated as follows:

$$LQ_i^r = \left( \frac{x_i^r}{x_i^n} \right),$$

where  $x_i^r$  is the output of industry  $i$  in region  $r$ ,  $x_i^r$  is the total output in region  $r$ . In the denominator  $x_i^n$  and  $x^n$  stand for the same reason but for the national level. To begin with, it is worth reviewing the background of the quotient. The numerator / denominator expresses to what extent the region/nation is specialized in production of industry  $i$ . Thus the ratio can be interpreted as the relative specialization of the region compared to the nation as a whole. If  $LQ = 1.5$  it means that the share of output  $i$  in regional output in  $r$  is 1.5 times higher than at the national level. Thus the region is assumed to be more specialized in industry  $i$ . If  $LQ = 0.8$  then the opposite is true. The region is assumed to be less specialized in industry  $i$  compared to national level. In the regionalization we consider  $LQ$  as an indicator of regional self-sufficiency. As long as  $LQ > 1$  the region is relatively specialized and it can cover its own demand locally; we do not adjust the national coefficients ( $a^{ij}$ ). If  $LQ < 1$  the region needs to import; in this case the regionalization is carried out by the following equation (Miller and Blair, 2009):

$$a_{ij}^{rr} = \begin{cases} (LQ_i^r) a_{ij}^n, & \text{ha } LQ_i^r < 1 \\ a_{ij}^n, & \text{ha } LQ_i^r \geq 1 \end{cases}$$

The  $LQ$  method can also be used (in a similar manner) for estimation of final demand block. Of course, additional data is required (e.g. total regional final demand, total value added). In the absence of these, the use of simplistic assumptions is necessary to estimate these blocks too. Ultimately, if the total estimated intermediate use and final demand exceed / fall short of the regional output, further adjustments are required (see later: *RAS*). However, the method has several weaknesses. On the one hand, the quotient is asymmetric. If the value of the quotient is smaller than one, coefficients in the given row of the table should be adjusted. On the other hand, if the value is bigger than one, they are not adjusted at all. However, in reality, some regional coefficients can be bigger than their national counterparts since the national coefficients can be considered as an average of regional ones. Furthermore, the adjustment should be done across rows (producing sectors). In other words, less specialized sectors will be scaled down. Thus, the input structure will be changed due to changes in the volume of rows. However, if the input structure is different in regions, then there is a chance that the structure of sales may also be different; this aspect is neglected by the  $LQ$  approach. Further, the relative size of a region is not taken into consideration in the quotient, however, as we noted above, the relative size is in strong connection with the self-supplying ability of regions.

Since rows should be modified in the first place, the structure of inputs will also be changed. This could be a strong assumption for a lot of transnational companies like Coca-Cola, which produces the same product in all regions using the same recipe. However, it could be a reasonable assumption for companies producing different varieties in different regions (like car manufacturers).

The interregional trade can be accounted for as a net export since  $LQ$  tends to underestimate interregional trade. Thus, an industry is only able to export or import. Simultaneous export and import (cross-hauling) is not allowed in this framework.

#### ***Purchases-only location quotient (PLQ)***

In this approach, only those industries are taken into consideration that use product  $i$  as an input in their production processes.

$$PLQ_i^r = \left( \frac{x_i^r}{x_i^{*r}} \right) \left( \frac{x_i^n}{x_i^{*n}} \right)$$

This equation is almost equivalent to the previous one. The small differences can be seen in the nominator and denominator. Where  $x_i^{*r}$  and  $x_i^{*n}$  do not denote total output in region  $r$  and in the nation, they denote the total output of those industries that provide inputs for industry  $i$ . The logic behind is that if industry  $j$  does not supply inputs for another industry  $i$  then it has no influence on regional self-supplying ability of industry  $i$ . Thus it is not necessary to include it in  $PLQ$ . On the other hand if industry  $j$  does supply inputs for industry  $i$  then it can have a real effect the self-supply ability. Thus, industry  $j$  must be accounted for in the quotient. In principle it would be a superior method compared to traditional approach but empirical results (see later) did not confirm it.

**Cross-industry location quotient (CILQ)**

This measure finally relaxes the strict assumption of modifying coefficient only by rows. *CILQ* takes into account that size of both purchaser and producers industries can affect the self-supplying ability of region  $r$ . Due to cell-by-cell modification, this approach can describe I-O relations more realistically than its predecessors. The *CILQ* formula can be written as follows:

$$CILQ_{ij}^r = \left( \frac{x_i^r}{x_i^n} \frac{x_j^n}{x_j^r} \right)$$

$$a_{ij}^{rr} = \begin{cases} (CILQ_{ij}^r) a_{ij}^n, & \text{ha } CILQ_{ij}^r < 1 \\ a_{ij}^n, & \text{ha } CILQ_{ij}^r \geq 1 \end{cases}$$

It is worthwhile to look at the logic of *CILQ*. If the share of industry  $i$  (producer) in region  $r$  (compared to the national level) is higher than the share of industry  $j$  (purchaser) in the same region (again compared to the national level), then the region can satisfy industry  $j$ 's input requirements in industry  $i$  ( $CILQ_{ij}^r > 1$ ). Likewise if the share of industry  $i$  is smaller than the share of industry  $j$ , the region is not able to cover its own needs, the coefficients in the table must be adjusted according to the equations above. It should be noted that *CILQ* can be expressed as the ratio of two location quotients ( $CILQ_{ij}^r = \frac{LQ_i^r}{LQ_j^r}$ ). Furthermore it is also true that in the diagonal ( $i=j$ )  $CILQ = 1$ . So in this case *CILQ* should be replaced by the traditional *LQ*.

**Flegg location quotient (FLQ)**

*FLQ* is one of the latest modifications of the cross-industry quotient that can be attributed to Flegg et al. (1995). Factors that can affect the self-supplying ability of a region expand by the relative size of the region. Thus, *FLQ* takes into account relative specialization, the size of the purchaser and producer sectors, and also the size of the region. The Flegg formula can be expressed as follows:

$$FLQ_{ij}^r = (\lambda^r) CILQ_{ij}^r,$$

where  $\lambda$  stands for the relative size of a region and it can be calculated by using the next equation:

$$\lambda^r = \{\log_2[1 + (x_E^r/x_E^n)]\}^\delta, 0 \leq \delta < 1.$$

As one can see, the quotient is not modified directly by the relative size, but by the logarithmic value of it. Thereby, the scaling is not so intensive, and the whole expression is then raised to the power of  $\delta$ , which is a sensitivity parameter. The larger the value of  $\delta$  the stronger the adjustment in *FLQ*. The determination of  $\delta$  is crucial to generate realistic results, but the way how it can be carried is not so evident. There can be found several attempts to estimate  $\delta$  ( $0.15 < \delta < 0.30$ ) in international literature but there still no absolute consensus. Finally, the quotient can be applied like its predecessor:

$$a_{ij}^{rr} = \begin{cases} (FLQ_{ij}^r) a_{ij}^n, & \text{ha } FLQ_{ij}^r < 1 \\ a_{ij}^n, & \text{ha } FLQ_{ij}^r \geq 1 \end{cases}$$

In case of diagonal ( $i=j$ ) elements, traditional *LQ* can substitute *FLQ* again.

### Other LQ variations

The literature is remarkably rich in further modifications of the original method (semi-logarithmic  $LQ$ , semi-logarithmic  $CILQ$ , augmented  $FLQ$ ). However, these variants could not gain significant popularity in comparison with previous examples. Further details can be found in Miller and Blair (2009).

Nevertheless, the work of McCann and Dewhurst (1998) drew attention to an important aspect of regionalization. The authors emphasized that the values of regional coefficients could exceed their national counterparts because they can be considered as an average of regional ones. This aspect was not considered in any of the previous examples. Flegg and Webber (1998) published a modification of  $FLQ$  to overcome this shortcoming. The augmented quotient ( $AFLQ$ ) improved by an additional logarithmic factor, which stands for the relative specialization of a region:

$$AFLQ_{ij}^r = FLQ_{ij}^r \log_2(1 + LQ_j) = (\lambda^r) CIQ_{ij}^r \log_2(1 + LQ_j).$$

Consequently if region  $r$  can be considered as specialized ( $LQ > 1$ ), it will use its inputs more intensively, thus the intraregional coefficients can be higher than the national ones. It should be noted that this approach simply neglects the possibility of increasing returns to scale that will occur with increasing specialization and will lead to decreasing intensity of input use. Besides, empirical results did not prove it to be more realistic.

### Commodity balance method

The commodity balance (or supply-demand pool /SDP/) is simply the difference between regional supply and demand (Bonfiglio 2005):

$$\tilde{x}_i^r = \sum_j a_{ij}^n x_j^r + \sum_f c_{if}^n f_f^r.$$

In the first step, regional demand  $\tilde{x}_i^r$  is estimated using national input coefficients ( $a_{ij}^n$ ), regional sectorial output ( $x_j^r$ ), national final demand coefficients ( $c_{if}^n$ ) and total regional final demand ( $f_f^r$ ). In the second step the balance ( $b_i^r$ ) can be calculated:

$$b_i^r = x_i^r - \tilde{x}_i^r.$$

If  $b$  is greater than 0 then region  $r$  is capable of supplying its need, thus coefficients remain unadjusted. Of course, if supply is exceeded by demand then region  $r$  must rely on imports, and coefficients need to be scaled down. The correction is carried out in a similar manner as before in case of  $LQ$  methods:

$$a_{ij}^{rr} = \begin{cases} (x_i^r / \tilde{x}_i^r) a_{ij}^n, & \text{if } b_i^r < 0 \\ a_{ij}^n, & \text{if } b_i^r \geq 0 \end{cases}$$

Round (1972) showed that both  $CILQ$  and  $CB$  methods share common advantages and shortcomings.

### The role of simultaneous import and export (cross-hauling)

Obviously, the net approach of interregional trade is too strict an assumption. Still, most non-survey methods cannot overcome this problem. However, several attempts were made to shed light on this matter; there are only a few recommendations on how to assemble

input-output framework with cross-hauling ((first Jackson (1998) then Kroenberg (2007)).

The introduction of cross-hauling is not straightforward due to data unavailability. The first potential ad hoc estimation procedure was suggested by Jackson (1998). He introduced  $k_i^r$  coefficient as the share of cross-hauling as a percentage of total sectoral output in region  $r$ . If there is available data on interregional transportation or trade, one can derive  $k$ , thus the calculation of regional cross-hauling can be completed ( $CH_i^r = k_i^r x_i^r$ ). At this stage using net interregional trade and cross-hauling, gross interregional trade volumes can easily be derived.

Later, Kroenberg (2007) reformed Jackson's idea to a more concrete empirical example. The procedure is known as the Cross-Hauling Adjusted Regionalization Method (CHARM), which is a non-survey method using international trade data to estimate product heterogeneity and re-export. First, it is assumed that cross-hauling is a function of product heterogeneity because interregional trade is motivated by heterogeneous products. In a fictional world where all products are homogenous, there would be no intention for simultaneous export and import. Furthermore, Kroenberg also noted that re-export would also be affected by the size of region; his formulation can be expressed as follows:

$$CH = f(\varepsilon, X, Z, D), \text{ in simpler form: } CH = \varepsilon(X + Z + D),$$

where cross-hauling ( $CH$ ) is a function of product heterogeneity ( $\varepsilon$ ), total output ( $X$ ), total intermediate use ( $Z$ ) and total demand ( $D$ ).  $\varepsilon$  can be expressed from the equation on the right-hand side:

$$\varepsilon = \frac{V - |B|}{2(X + Z + D)},$$

where  $V$  is the total gross trade (import and export) and  $B$  is trade balance. The expression in the nominator stands for the difference between gross trade volume and the absolute value of the trade balance, which is in fact twice the cross-hauling (that is the reason for the multiplication by 1/2).  $X$ ,  $Z$  and  $D$  can be calculated easily for region  $r$  using any non-survey regionalization method. According to Kroenberg,  $\varepsilon$  should be estimated using national data, thus  $\varepsilon$  is only national and not region specific. Of course, the problem is more complex, but it is necessary to make such assumptions to keep mathematical and statistical calculations simple.

### **Bi-proportional methods (RAS, entropy and mathematical programming models)**

The second group of non-survey methods relies on national input-output tables and supposes that only the framework of a regional table is available; thus all cells within the table have to be estimated. For example, in case of intermediate consumption, usually the national table, the national total intermediate use and sales (as a frame) is known. Furthermore, total regional intermediate use and sales (as a regional frame) is also available. The aim is to generate a regional table that will preserve the national structure as much as possible and at the same time will satisfy the available regional frames as constraints (Lahr–Mesnard 2004). These procedures can be categorized, at the least, into two groups:

- 1) Bi-proportional methods – RAS and alternative techniques,
- 2) Mathematical programming (optimization) models.

*RAS* is the most well-known bi-proportional method. Stone (1961) first recognized the potential application of *RAS* in regionalizing problems. *RAS* is considered as a partial or a non-survey method due to its low data requirements. Unfortunately, I-O tables are not published in each year, therefore, for years between the publications of two tables, researchers have to generate estimated tables. Initially, *RAS* was considered as such an updating technique for national I-O tables (especially for intermediate use matrix). Later (as a consequence of greater interest for regional sciences), more attention was drawn to *RAS* as a “spatial updating” (regionalizing) method (Pigozzi and Hinojosa, 1985). In case of regionalization, the application is quite straightforward (just like in the case of updating). The initial matrix is always the national table (or sometimes the I-O table of another region’s table /that is similar in an economic sense/). In the first step, the regional table is assumed to be identical to the national one ( $\mathbf{Z}^0 = \mathbf{Z}^n$ ), which obviously cannot satisfy the equality criteria between the total of rows, columns and regional frames. Thus further adjustment is needed, which will be carried out in two steps. First, rows need to be scaled down by a certain ratio to satisfy equality of the regional frame and total supply in  $\mathbf{Z}$ . Second, the same procedure has to be run for columns. The row scaling ratio (column vector) is  $r_i^1 = \frac{z_i^r}{\sum_j z_{ij}^r}$ , where the numerator is actual regional data and the denominator is the sum of the estimated table by  $j$ . If  $r_i^1 < 1$ , then elements in row  $i$  of the estimated table are higher than they should be and vice versa. Thus the rows of the estimated table will satisfy constraints by multiplying the table by this vector (diagonal matrix). But, at this stage, column totals will differ from the regional column frame (column constraint). Thus, the same procedure also has to be applied for columns. The column scaling ratio (row vector) is  $s_j^1 = \frac{z_j^c}{\sum_i z_{ij}^c}$ . If  $s_j^1 < 1$ , then elements in the estimated table are higher than they should be, thus they need to be scaled down by  $s_j^1$  to achieve consistency by columns ( $\mathbf{Z}^2 = \hat{\mathbf{r}}^1 \mathbf{Z}^0 \hat{\mathbf{s}}^1$ ). At this stage, it is likely that the rows will not satisfy regional constraints again, thus the procedure has to be started again. The sequential repeat of step 1 and step 2 will adjust the initial table to be constrained by regional frames. Usually the procedure is convergent and after a few iterations, estimated values will be very close to regional frames.

Although *RAS* is the most accepted and applied method, it is not the only bi-proportional technique. The literature is rich in other similar alternative scaling methods (e.g. *DSS*) and improved *RAS* methods (*CRAS*, *GRAS*, additive *RAS* – for additive *RAS* see Révész 2011).

Where *GRAS* is a specialized procedure that can adjust both negative and positive elements by separating them in two different matrices (one with non-negative elements and one with absolute values), then the adjustment of rows and columns will be carried out for the sum of two matrices, and in the final step, the results will be adjusted by negative elements (Junius and Oosterhaven 2002).

A further evolution of *RAS* is the two-staged *CRAS* (cell-corrected *RAS*). In the first stage, traditional *RAS* is run; thus the matrix will be fitted to regional constraints. In the second stage, the matrix is modified by using distributions calculated from tables of other regions. The modification is carried out by a constrained optimization problem, which can take spatial variations into consideration (Mínguez et al. 2009).

The last group of non-survey methods consists of entropy methods that are close to RAS techniques (Jaynes 1957, Wilson 1970). These models are based on information theory, and their aim is to preserve as much information from the initial table as possible by maximizing entropy (a well-known concept in thermodynamics). The following model (*Minimum Sum of Weighted Cross-Entropy*) illustrates the basic concept of the approach. Suppose that  $c_{ij}$  is the initial coefficient matrix, and regional frame ( $x_i^r$  and  $x_j^r$ ) is also available:

$$c_{ij} = \begin{bmatrix} 0,714 & 0,2 & 0,060 \\ 0,143 & 0,6 & 0,176 \\ 0,143 & 0,2 & 0,765 \end{bmatrix}.$$

In this approach, entropy is defined by the following equation:

$$H = \sum_i \frac{x_j^r}{\sum_j x_j^r} \sum_j b_{ij} \log(b_{ij}/c_{ij}),$$

where  $b_{ij}$  is objective variable. If the structure of this estimated matrix is precisely identical to the initial matrix  $c_{ij}$  then entropy is zero ( $H_{MSCE}=0$ ), which means that the preservation of the initial matrix was a 100 % success. If there is a slight change in  $b_{ij}$  ( $b_{11} = 0.8$ ) then  $H_{MSCE} = 0.039$ , which means preservation of information is not complete. The final model searches for the maximum of entropy (minimal  $H$ ) that will satisfy additional row and column constraints:

$$\sum_j b_{ij} x_{rj} = x_{ri} \text{ and } \sum_i b_{ij} = 1.$$

At this stage, one can show (by deriving Lagrangian of the problem and the first order conditions (McDougall, 1999)) that both entropy and RAS methods will generate the same result.

Besides these, there are several other mathematical optimization methods in the literature that cannot be simplified to bi-proportional techniques. These models use different special objective functions to find the optimal values of regional I-O tables. The objective function can be formulated in many different ways (for example, the sum of traditional, weighted, normalized absolute or squared differences). It should be noted that not all methods are capable of preserving the sign of coefficients, consequently, negative values can appear in estimated I-O tables even if all elements of the initial matrix were nonnegative. Based on literature review, we can state that methods based on entropy and squared differences are the most widely used approaches.

### Empirical results of non-survey methods

Fortunately, the literature is rich in empirical studies. Most of them focus on evaluating the precision and reliability of different regionalization methods. They can be carried out only in regions where regional survey generated tables are available as a benchmark. Since tables consist of hundreds of cells, it is difficult to compare initial and estimated matrices. Therefore several error formulas (such as squared differences) that can be applied to compress the information regarding the difference between coefficients (Lahr 2001).

Non-survey methods are still recommended, especially where there is a lack of appropriate regional data. Usually, these studies compare the performance of different LQ methods. Based on findings of early studies, it can be stated that traditional LQ generated the most precise results (Morrison–Smith 1974) by exceeding other developed LQ indices

(for example *PLQ*, *CILQ*, *RLQ*). However, the recent notion of Flegg et al. (1995) has gained remarkable popularity in the last two decades. Furthermore, it can also be seen that *FLQ* generally performed better than its predecessors and proved to be one of the best non-survey alternatives to estimate regional I-O relationships (Miller and Blair 2009, Bonfiglio és Chelli 2008, Kowalewski 2012, Lindberg 2011, Riddington et al. 2006, Swaminathan 2008, Tohmo 2004, Bonfiglio 2005). The only shortcoming of *FLQ* is the estimation of  $\delta$ , the sensitivity parameter, which, at best, should be carried out by econometric estimation, although, the lack of data makes it difficult. Thus, the optimal range of  $\delta$  can be determined by reviewing the literature. Based on studies of Flegg and Tohmo (2013) in Finland, Kowalewski (2012) in Germany, Flegg and Webber (2000) in Scotland and Bonfiglio and Chelli (2008), we can state that  $\delta$  is between 0.1 and 0.3.

Furthermore, as we have seen, *RAS* is the alternative (augmentation) to quotient-based regionalization. *RAS* can be applied alone, but most hybrid methods use *RAS* as an intermediate step in the procedure. In general, *RAS* produced satisfactory results (Miller and Blair 2009, Morrison and Smith 1974, Jalili 2005, Riddington et al. 2006, Round 1983, Harris and Liu 1997). Flegg and Tohmo (2013) suggested the combined use of *FLQ* and *RAS* to improve the accuracy of estimation. In the first step, national coefficients are modified by standard *FLQ*; then the table is constrained to be given a regional frame by *RAS*. The resulting regional tables are considered as a reasonably good representation of true regional I-O relations (compared to other non-survey methods).

This combination was chosen in our previous research in which we examined the possible effects of introducing a Blue Economy innovation in the Transdanubian region. First we carried out the two-stage regionalization (*FLQ+RAS*), then the table was integrated into the multisectoral block of a spatial computable general equilibrium model that was capable of describing the expected effects of different policy scenarios. The complete impact analysis was carried out by running off the full GMR model. For additional information, please see our published study in Regional Statistic (Varga et al. 2013).

The other part of the literature completely refuses non-survey methods because, despite all the advantages, they cannot produce reliable results that would justify their application. Nevertheless, non-survey methods are widely applied, and their significance should not be underestimated. Lahr (1993) emphasized that most hybrid methods (in their initial step) are based on non-survey techniques, which are then later enhanced by additional regional data. One of the earliest methods (Morrison and Smith 1974) comparing different methods suggests the use of hybrid techniques, although, since then a lot of other studies have also reinforced these findings: (Brand (2012) in Finland, Kowalewski (2012) and Lindberg (2011) in Sweden, Patriquin et al. (2002) in Western-Middle Alberta, Ralston and Hastings (1986) in Delaware, Harris and Liu (1997) in Porthmouth, Bonfiglio (2005) in Marche region, Oosterhaven et al. (2003) in the Netherland, Jiang et al. (2012) in China and Round (1983)). By examining several methods (*LQ*, *PLQ*, *RLQ*, *FLQ*, *AFLQ*, *SDP*, *RSP*, *RAS*), the authors concluded that even if improved methods (like *FLQ*) can generate relatively accurate results, the assumptions of non-survey methods are generally too strict to reflect true input-output relations.

Others came to even stricter results; McMenamin and Haring (1974) and Kipnis (1984) found that even hybrid input-output tables are not accurate enough. Thus, the only reliable

way of estimating regional tables is the survey method. Later, Harris and Liu (1997) found that hybrid methods can excel non-survey techniques; however, they are still too resource and input demanding. The improvement in accuracy of results is not in line with the increased cost of the procedure, thus, it is worthwhile investing a bit more resources and generating pure survey based tables. On the other hand, Riddington et al. (2006) revealed some interesting aspects; they found that hybrid methods were able to generate misleading, distorted results in the case of relatively small regions. Nevertheless, in bigger regions their reliability is improved greatly.

Based on the above-mentioned studies, we can conclude that there is no consensus in the literature. It did not become clear whether the pure non-survey method can provide reliable results or if they should only be used as a step in hybrid methods. Despite this, most studies agreed that non-survey or hybrid methods are clearly efficient ways of estimating regional tables, and they can generate a good approximation of sectorial relationships.

### Estimation of interregional trade

In case of multiple regions, it is crucial to estimate transactions between regions; thus both intra- and interregional data are required. In practice, usually there is no appropriate available data to assemble a table of interregional trade. Furthermore, by increasing the number of regions, the number of tables will increase faster. In case of two regions, four tables need to be estimated, in case of three regions nine, in case of four regions sixteen tables. In the next sections, different interregional methods will be presented.

#### *Extended LQ-method*

One of the simplest approaches is the well-known *LQ* approach for multiple regions. The basic concept is the same as before:

$$a_{ij}^{rr} = \begin{cases} (LQ_i^r)a_{ij}^n, & \text{if } LQ_i^r < 1 \\ a_{ij}^n, & \text{if } LQ_i^r \geq 1 \end{cases}$$

For transactions from other regions the following is true:

$$a_{ij}^{sr} = \begin{cases} (1 - LQ_i^r)a_{ij}^n, & \text{if } LQ_i^r < 1 \\ 0, & \text{if } LQ_i^r \geq 1 \end{cases}$$

The same can be carried out in case of region  $s$ . These trade flows have to be balanced by each other, which means that imports from region  $s$  in region  $r$  should be equal to exports from  $s$  to  $r$ . As mentioned previously, the approach can calculate only net exports and imports, which means that if region  $r$  exports product  $I$ , then  $s$  can only import it.

Next, consider an example consisting of three regions. Our aim is to estimate an interregional input-output table for the three regions system. First, the usual *LQ* method is used for all regions to generate an intraregional input-output table, with net exports to and imports from the rest of country, thus we can calculate the input-output table for rest of country ( $Z^{NN}$ ) as well. For region 1 it can be represented by:

$$\begin{bmatrix} Z^{11} & Z^{1N} \\ Z^{N1} & Z^{NN} \end{bmatrix},$$

where  $Z^{11}$  is the intraregional matrix,  $Z^{1N}$  is the total outflow from region (to region 2 and 3),  $Z^{N1}$  is total inflow to region from rest of country, and finally  $Z^{NN}$  is the total intraregional transactions outside the region. The same procedure can be carried out for the other regions, but it is worth noting that the meaning of  $Z^{rN}$ ,  $Z^{Nr}$  and  $Z^{NN}$  will be different because different regions will form the rest of country. In the end the following matrices will be calculated:

$$\begin{bmatrix} Z^{11} & \blacksquare & \blacksquare & Z^{1N} \\ \blacksquare & Z^{22} & \blacksquare & Z^{2N} \\ \blacksquare & \blacksquare & Z^{33} & Z^{3N} \\ Z^{N1} & Z^{N2} & Z^{N3} & 0 \end{bmatrix},$$

where diagonal elements are intraregional input-output tables. Black squares stand for unknown interregional transactions we would like to estimate. Red vectors represent the estimated total outflow in each region. In the last step, these values are distributed amongst black matrices. It can be performed in many different ways (for simplicity, assume equal distribution). At this stage, the complete interregional system takes the following form (without intraregional transactions):

$$\begin{bmatrix} 0 & Z^{12} & Z^{13} & Z^{1N} \\ Z^{21} & 0 & Z^{23} & Z^{2N} \\ Z^{31} & Z^{32} & 0 & Z^{3N} \\ Z^{N1} & Z^{N2} & Z^{N3} & 0 \end{bmatrix}.$$

The last task is to achieve consistency in the system. The estimated interregional trade flow data is not guaranteed to satisfy constraints in the frame, but *RAS* provides a simple way of ensuring full consistency within the system.

### Gravity method

The concept of the gravity model is based on Newton’s law in physics to estimate interregional trade flows using available regional data. In this approach, if two regions are close to each other in space and large in an economic sense, then it is more likely that they will trade with each other. The basic generalized equation can be expressed as follow:

$$Z_i^{rs} = \frac{(c_i^r x_i^r)(b_i^s x_i^s)}{(d^{rs})^{e_i}} = (k_i^{rs}) \frac{x_i^r x_i^s}{(d^{rs})^{e_i}}.$$

It can be seen that the magnitude of interregional trade ( $Z_i^{rs}$ ) is affected by two factors: the size of output in regions ( $x_i^r, x_i^s$ ) and distance ( $d^{rs}$ ). Of course the latter has a negative effect on trade. The other output-sensitivity parameters ( $c_i^r, b_i^s, k_i^{rs}$ ) need to be estimated in order to generate reliable results. Finally  $e_i$  is the sensitivity of distance, which also needs to be estimated outside the model.

The most obvious advantage of the gravity approach is its ability to estimate interregional trade on a gross basis compared to *LQ*. On the other hand, further data requirements make it difficult to calculate all parameters of the model (Black 1972). Furthermore, it can be seen that the gravity method is only capable of generating trade between regions in commodity  $i$ . It cannot show which industry is going to use this product in its production procedure. For this purpose, the extended *LQ* method offers a potential alternative.

### *Mathematical programming models*

Besides, other methods approach the estimation of interregional trade and regional input-output table as a constrained mathematical optimization problem. In this case, different parts of the estimated regional table are restricted by given constraints (regional data as frames). Clearly, a great advantage of mathematical models is that the convergence of the procedure is guaranteed while for example in case of *RAS*, it always depends on the initial table and regional data (especially in case of interregional trade).

Based on Canning and Wang (2005), we can reformulate the problem as follows:

$$\sum_{s=1}^g \sum_{j=1}^n z_{ij}^{rs} + \sum_{s=1}^g \sum_{k=1}^h y_{ik}^{rs} + e_i^r = x_i^r.$$

In this sense, total output in region  $r$  ( $x_i^r$ ) is allocated between intraregional ( $z_{ij}^{rr}$ ), interregional intermediate use ( $z_{ij}^{rs}$ ) and intra- ( $y_{ik}^{rr}$ ), interregional final consumption ( $y_{ik}^{rs}$ ) and international export ( $e_i^r$ ). On the expenditure side, the following is true:

$$\sum_{r=1}^g \sum_{i=1}^n z_{ij}^{rs} + \sum_{i=1}^n m_{ij}^s + v_j^s = x_j^s.$$

Output of region  $s$  is produced using local ( $z_{ij}^{ss}$ ) and interregional intermediate use ( $z_{ij}^{rs}$ ), international import ( $m_{ij}^s$ ) and value added ( $v_j^s$ ). For final demand the constraint is the following:

$$\sum_{h=1}^k \sum_{s=1}^g y_{ih}^{sr} + \sum_{h=1}^k m_{ih}^r = y_i^r.$$

In this case regional final demand ( $y_i^r$ ) can be satisfied by local production ( $y_{ih}^{rr}$ ), interregional trade ( $y_{ih}^{sr}$ ) and international imports ( $m_{ih}^r$ ). Finally intermediate use is described by the last equation:

$$\sum_{r=1}^g (\sum_{s=1}^g z_{ij}^{sr} + m_{ij}^r) = z_{ij}.$$

This is an important constraint because it requires the sum of intraregional intermediate use ( $z_{ij}^{rr}$ ), interregional trade ( $z_{ij}^{sr}$ ) and imports ( $m_{ij}^r$ ) to be perfectly fitted to national intermediate use. After that, the remaining constraints will ensure equality between the sum of regional variables and their national counterparts (e.g.  $\sum_{r=1}^g m_{ij}^r = m_{ij}$ ). Then a complex objective function that measures the difference between estimated regional and national data will be minimized. For European (including Hungarian) applications, please see Thissen et al. (2010).

### **The proposed method and its introduction through a numerical example**

The estimation of a full interregional table on the basis of pure non-survey methods requires the combination of several techniques. First, based on its empirical results, *FLQ* can modify the national table to incorporate special regional characteristics. By applying *FLQ*, it is possible to estimate the total sectorial interregional in- and outflow of the region, although, it can be applied only in single region cases. Thus, it is necessary to extend the method to estimate region-to-region economic relationships, which can be accomplished by the idea of the standard extended *LQ* method. The next step is to achieve full consistency of the table. By applying *FLQ*, the rows and columns of the initial table are modified so there is no guarantee that it will be constrained by actual regional data. These inconsistencies can be eliminated by bi-proportional matrix updating methods (like *RAS*). For ease of understanding, in the following, we can see a numerical example of the method

where our starting point is the national intermediate transaction table ( $Z_n$ , and its coefficient table  $A_n$ ):

Table 1

*An example for national intermediate use matrix (absolute values and coefficients)*

$Z_n$	Agr	Ind	Ser	$A_n$	Agr	Ind	Ser
Agr	50	20	10	Agr	0.56	0.17	0.05
Ind	10	60	30	Ind	0.11	0.50	0.16
Ser	10	20	130	Ser	0.11	0.17	0.68

Source: author's own editing.

Furthermore, total regional sectorial output is also available; thus it is possible to calculate  $FLQ$ s:

Table 2

*An example for  $FLQ$  in three regions*

FLQ1	Agr	Ind	Ser	Region1	Output
Agr	1.48	1.78	1.69	Agr	40
Ind	0.56	0.83	0.95	Ind	30
Ser	0.59	1.05	0.88	Ser	50
				Total	120

FLQ2	Agr	Ind	Ser	Region2	Output
Agr	0.95	1.00	0.90	Agr	30
Ind	1.00	0.95	0.90	Ind	40
Ser	1.11	1.11	1.05	Ser	70
				Total	140

FLQ3	Agr	Ind	Ser	Region3	Output
Agr	0.63	0.53	0.60	Agr	20
Ind	1.88	1.19	1.13	Ind	50
Ser	1.66	0.88	1.05	Ser	70
				Total	140

Source: author's own editing.

Considering the calculated values of the quotient, we can see that certain cells are below 1, which means that for example in region 1, industry and service cannot satisfy the intermediate demand of agriculture, thus the region needs to import. On the other hand, agriculture is capable of supplying both industry and service and even exporting to other regions ( $FLQ > 1$ ).

Using traditional rules of the LQ approach, we can now adjust cells of the national table using  $FLQ$ . If  $FLQ < 1$ , cells should be scaled down or otherwise unadjusted. The result can be seen below in the diagonal of the matrices:

Table 3

*An example for an interregional intermediate transaction matrix*

Interregional intermediate use		Region1			Region2			Region3			Total
		Agr	Ind	Ser	Agr	Ind	Ser	Agr	Ind	Ser	
Region1	Agr	22.2	5.00	2.63	0.40	0.61	0.48	0.48	2.35	0.94	41,59
	Ind	2.01	12.50	6.04	0.31	0.48	1.45	0.00	0.00	0.43	22,55
	Ser	2.12	4.24	30.01	0.16	0.33	0.00	0.00	1.16	0.00	37,99
Region2	Agr	0.00	0.00	0.00	15.87	5.44	2.72	2.03	2.35	0.94	25,93
	Ind	1.22	1.25	0.93	2.72	19.05	8.16	0.00	0.00	0.43	32,45
	Ser	1.16	0.38	1.25	3.01	6.01	47.89	0.00	1.16	0.00	60,80
Region3	Agr	0.00	0.00	0.00	0.40	0.61	0.48	7.05	3.63	1.81	13,75
	Ind	1.22	1.25	0.93	0.31	0.48	1.45	2.22	25.00	10.20	44,14
	Ser	1.16	0.38	2.10	0.16	0.33	0.00	2.22	6.01	47.89	60,80
<i>Total</i>		<i>31,11</i>	<i>25,00</i>	<i>44,74</i>	<i>23,33</i>	<i>33,33</i>	<i>62,63</i>	<i>15,56</i>	<i>41,67</i>	<i>62,63</i>	<i>340,00</i>

Source: author's own editing.

The next step is generating an estimate of interregional trade (off-diagonal matrices). On a net basis,  $FLQ$  can be used to calculate trade flows between regions using the following equation:

$$T_{ij}^{r,roc,r} = (1 - FLQ_{ij}^r) \cdot a_{ij}^n \cdot x_j^r,$$

where  $T_{ij}^{r,roc}$  is net trade flow entering the region (from rest of country). In this hypothetical example, we distributed these imports equally between regions (for example in case of interregional industrial imports from regions 2 and 3 to agriculture in region 1 /1.22-1.22/). But again, this should be carried out by estimation based on further regional data. At this stage the problem is that the estimated table is not consistent. The columns satisfy regional intermediate use constraints (row total) (because of the import side distribution of  $FLQ$  generated imports), but in the rows, intermediate supply is not guaranteed to be consistent. Using export-oriented distribution of trade ( $T_{ij}^{r,roc} = (FLQ_{ij}^r - 1) \cdot a_{ij}^n \cdot x_j^r$ ), the desired sum of rows can be generated that will be not in line with the actual sum of the rows (e.g. in first row:  $35.71 < 41.59$ ). Otherwise, total intermediate supply can also be approximated by subtracting (estimated) final demand from regional industrial output. Thus in the final step, the matrix needs to be balanced by a matrix updating technique such as *RAS*.

## Summary

The current study first attempted to introduce the most important methods of estimating symmetric regional input-output tables in Hungary. However, even this concise review indicates how enormous the international literature is and how widely spread the applications are.

In the absence of clear consensus in regionalization, considering the advantages and shortcoming of different approaches, we can conclude that the hybrid method is still the most suitable choice. Hybrid results are relatively accurate and are not as time and resource

intensive, although, it is essential to acquire further regional data to enhance accuracy (i.e. surveys or experts)

In case of scarce regional data, non-survey methods offer alternatives for describing interregional interindustry relationships. Although, without a regional reference table, results cannot be evaluated, there are still numerous examples in international literature with reliable results confirming the possibilities of non-survey approaches.

Based on these experiences, we recommend the application of a combined method, which is attributable to several reasons. First, these methods pursue simplicity, second, they do not differ from the mainstream approach of international literature and third, they can be considered as reliable and efficient methods (based on international experiences due to a lack of domestic references). Our starting point is *FLQ* because its theoretical ground is more solid than its predecessors, and its empirical results are satisfactory. The missing  $\delta$  parameter can be estimated based on literature, or the average of previous deltas can be used as a rule of thumb. After the estimation of interregional trade and the intraregional table, there are multiple further options. First, simultaneous exports and imports (cross-hauling) can be incorporated based on Jackson (1998) and Kroenenberg (2007). Then using these trade data (either net or gross), the gravity method can be applied to estimate interregional flows. If results of this approach are not detailed enough, the extended LQ method can generate a four-dimensional (region-to-region, industry-to-industry) interregional table. In the final step, we need to ensure consistency between and within tables by using all available regional data. By preserving the basic structure of tables, *RAS* can modify them to achieve full consistency between tables and constraints.

\* \* \*

#### Acknowledgements

The current study was supported by MTA-PTE Innovation and Economic Growth Research Group. I would like to express thanks to Professor Randall Jackson and Professor Attila Varga for their valuable comments and advice.

#### REFERENCES

- Black, William R. (1972): Interregional commodity flows: Some experiments with the gravity model *Journal of Regional Science* 12 (1): 107–118.
- Bonfiglio, A. (2005): *A Sensitivity Analysis of the Impact of CAP Reform. Alternative Methods of Constructing Regional I-O Tables* PhD Dissertation. Polytechnic University of Marche Ancona, Italy.
- Bonfiglio, A.–Chelli, F. (2008): Assessing the Behaviour of Non-Survey Methods for Constructing Regional Input–Output Tables through a Monte Carlo Simulation *Economic Systems Research* 20 (3): 243–258.
- Brand, S. (2012): *A Note on Methods of Estimating Regional Input-Output Tables: Can the FLQ Improve the RAS Algorithm?* The Business School with Plymouth University, Working Paper., Plymouth.
- Canning, P.–Wang, Z. (2005): A flexible mathematical programming model to estimate interregional input-output accounts *Journal of Regional Science* 45 (3): 539–563.
- Európai Bizottság (2010): A forrás-felhasználás táblák és az input-output keretrendszer In: *Javaslat az Európai Parlament és a Tanács Rendelete az Európai Unióban alkalmazandó nemzeti és regionális számlák európai rendszeréről* 2010.12.20. COM(2010) 774 végleges. A. melléklet/9. fejezet, Brüsszel.
- Flegg, A. T.–Webber, C. D. (2000): Regional Size, Regional Specialization and the FLQ Formula *Regional Studies* 34 (6): 563–569.

- Flegg, A. T.–Tohmo, T. (2013): Regional Input–Output Tables and the FLQ Formula: A Case Study of Finland *Regional Studies* 47 (5): 713–721.
- Flegg, A. T.–Webber, C. D.–Elliott, M. V. (1995): On the Appropriate Use of Location Quotients in Generating Regional Input–Output Tables *Regional Studies* 29 (6): 547–561.
- Greenstreet, D. (1989): A Conceptual Framework for Construction of Hybrid Regional Input–Output Models *Socio-Economic Planning Sciences* 23 (5): 283–289.
- Harris, R. I. D.–Liu, A. (1997): Input–Output Modelling of the Urban and Regional Economy the Importance of External Trade *Regional Studies* 32 (9): 851–862.
- Jackson, R. W. (1998): Regionalizing National Commodity-by-Industry Accounts *Economic Systems Research* 10 (3): 223–238.
- Jiang, X.–Dietzenbacher, E.–Los, B. (2012): Improved Estimation of Regional Input–Output Tables Using Cross-regional Methods *Regional Studies* 46 (5): 621–637.
- Junius, T.–Oosterhaven, J. (2003): The Solution of Updating or Regionalizing a Matrix with both Positive and Negative Entries *Economic System Research* 15 (1): 87–97.
- Kipnis, B. A. (1984): Input–Output Tables for Medium-Sized Cities: Survey Coefficients or Short-cut Methods? A Case Study in Brazil *Journal of Regional Science* 24 (3): 443–450.
- Kowalewski, J. (2012): *Regionalization of national input-output tables: empirical evidence on the use of the FLQ formula* Hamburg Institute of Regional Economics, HWWI Research Paper No. 126., Hamburg.
- Központi Statisztikai Hivatal (2005): *Az Ágazati Kapcsolatok Mérlegének matematikai feldolgozása, 2000* KSH, Budapest.
- Kronenberg, T. (2007): *How Can Regionalization Methods Deal With Cross-hauling?* Institut für Energieforschung (IEF), Systemforschung und Technologische Entwicklung (STE), Working Paper 2007/14., Jülich.
- Lahr, M. L. (1993): A Review of the Literature Supporting the Hybrid Approach to Constructing Regional Input–Output Models *Economic Systems Research* 5 (3): 277–293.
- Lahr, M. L. (2001): A strategy for producing hybrid regional input-output tables In: Lahr M. L.–Dietzenbacher, E. *Input–Output Analysis: Frontiers and Extensions* pp. 1–31., Palgrave, London.
- Lahr, M. L.–de Mesnard, L. (2004): Biproportional Techniques in Input–Output Analysis: Table Updating and Structural Analysis *Economic System Research* 16 (2): 115–134.
- Leontief, W. (1936): Quantitative Input–Output Relations in the Economic System of the United States *Review of Economics and Statistics* 18: 105–125.
- Leontief, W. (1941): *The Structure of American Economy 1919–1939* Oxford University Press, New York.
- Lindberg, G. (2011): *On the appropriate use of (input-output) coefficients to generate non-survey regional input-output tables: Implications for the determination of output multipliers* ERSA Conference Paper, No. ersa10p800.
- Mínguez, R.–Oosterhaven, J.–Escobedo, F. (2009): Cell-corrected RAS method (CRAS) for updating or regionalizing an input-output matrix *Journal of Regional Science* 49 (2): 329–348.
- McDougall, R. A. (1999): *Entropy Theory and RAS are Friends* GTAP Working Paper No. 06., Purdue University West Lafayette, USA.
- McCann, P.–Dewhurst, J. HH (1998): Regional Size, Industrial Location and Input–Output Expenditure Coefficients *Regional Studies* 32 (5): 435–444.
- McMenamin, D. G.–Haring, J. E. (1974): An Appraisal of Nonsurvey Techniques for Estimating Regional Input–Output Models *Journal of Regional Science* 14 (2): 191–205.
- Miller, R. E.–Blair, P. D. (2009): *Input–Output Analysis: Foundations and Extensions* Cambridge University Press, New York.
- Morrison, W. I.–Smith, P. (1974): Nonsurvey Input–Output Techniques at the Small Area Level: An Evaluation *Journal of Regional Science* 14 (1): 1–14.
- Oosterhaven, J.–van der Knijff, E. C.–Eding, G. J. (2003): Estimating interregional economic impacts: an evaluation of nonsurvey, semisurvey, and full-survey methods *Environment and Planning A* 35 (1): 5–18.
- Patriquin, M. N.–Alavalapati, J. R.R.–Wellstead, A. M.–White, W. A. (2002): A comparison of impact measures from hybrid and synthetic techniques: A case study of the Foothills Model Forest *The Annals of Regional Science* 36 (2): 265–278.

- Pigozzi, B. WM.–Hinojosa, R. C. (1985): Regional Input-Output Inverse Coefficients Adjusted from National Tables *Growth and Change* 16 (1): 8–12.
- Ralston, S. N.–Hastings, S. E. (1986): Improving Regional I-O Models: Evidence Against Uniform Purchase Coefficients Across Rows *The Annals of Regional Science* 20 (1) 65–80.
- Révész, T. (2011): A Magyar Gazdaság 2010. Évi Ágazati Kapcsolatok Mérlegeinek Becslése. ENERGIACLUB Szakpolitikai Intézet és Módszertani Központ, Budapest.
- Riddington, G.–Gibson, H.–Anderson, J. (2006): Comparison of Gravity Model, Survey and Location Quotient-based Local Area Tables and Multipliers *Regional Studies* 40 (9): 1069–1081.
- Round, J. I. (1972): Regional Input-Output Models in the UK: A Reappraisal of Some Techniques *Regional Studies* 6 (1): 1–9.
- Round, J. I. (1983): Nonsurvey Techniques: A Critical Review of the Theory and the Evidence *International Regional Science Review* 8 (3): 189–212.
- Stevens B. H.–Treyz, G. I.–Ehrlich, D. J.–Bower, J. R. (1983): A New Technique for the Construction of Non-Survey Regional Input-Output Models and Comparison with Two Survey-Based Models *International Regional Science Review* 8: 271–286.
- Stone, R. (1961): *Input-Output and National Accounts* Organization for Economic Cooperation and Development, Paris.
- Swaminathan, A. M. (2008): *Methods for generation of a regional input-output table for the state of Maharashtra: A comparative analysis* Working Paper No.29, UDE Dr. Vibhooti Shukla Unit in Urban Economics & Regional Development, University of Mumbai, Mumbai.
- Thissen, M.–van Oort, F.–Diodato, D.–Ruijs, A (2010): *Regional Competitiveness and Smart Specialization in Europe: Place-based Development in International Economic Networks* Edward Elgar Publishing, Cheltenham.
- Tohmo, T. (2004): New Developments in the Use of Location Quotients to Estimate Regional Input-Output Coefficients and Multipliers *Regional Studies* 38 (1): 43–54.
- Varga, A.–Hau-Horváth, O.–Szabó, N.–Járosi, P. (2013): A GMR–Európa-modell alkalmazása két gazdaság-típusú innovációk hatásvizsgálatára. *Területi Statisztika* 53 (5): 411–434.
- Zalai, E. (2012): *Matematikai Közgazdaságtan II.: Többszektoros modellek és makrogazdasági elemzések* Akadémia Kiadó, Budapest.