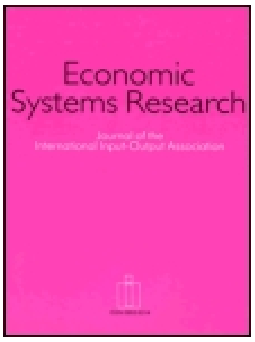
Economic Systems Research

ISSN: 0953-5314 (Print) 1469-5758 (Online) Journal homepage: https://www.tandfonline.com/loi/cesr20



INTER-REGIONAL TRADE FLOW ESTIMATION THROUGH NON-SURVEY MODELS: AN EMPIRICAL ASSESSMENT

Ana L.M. Sargento , Pedro Nogueira Ramos & Geoffrey J.D. Hewings

Economic Systems Research, 2012, Vol. 24(2), June, pp. 173–193

INTER-REGIONAL TRADE FLOW ESTIMATION THROUGH NON-SURVEY MODELS: AN EMPIRICAL ASSESSMENT

ANA L.M. SARGENTO[a*], PEDRO NOGUEIRA RAMOS[b] and GEOFFREY J.D. HEWINGS[c]

[a]*Polytechnic Institute of Leiria, School of Technology and Management, Leiria, Portugal;* [b]*University of Coimbra, Faculty of Economics, Coimbra, Portugal;* [c]*University of Illinois, Urbana-Champaign, USA*

(*Received 29 July 2010; In final form 21 March 2011*)

Inter-regional trade estimation has been pointed out as a crucial problem when constructing a multiregional input–output system. Knowledge of inter-regional trade flows, at least of the pooled volume of exports and imports by commodity, is critical in accounting for important spillover and feedback effects deriving from inter-regional linkages. However, in most countries, there are no completely reliable survey-based statistics on inter-regional trade. Thus, this paper intends to evaluate the reasonability of using indirect inter-regional trade estimates, comparing different estimating methods and assessing the sensitivity of the model results. Based on our empirical comparisons we conclude that input–output models are not greatly affected by the insertion of different trade values. Thus, our results support the use of indirect estimates for inter-regional trade, whenever survey-based data are unavailable.

Keywords: Trade; Input–Output; Methods

1. INTRODUCTION

Inter-regional trade estimation has been extensively pointed out as a crucial problem to be overcome when constructing any multi-region or multi-country input–output system. In fact, knowledge of inter-regional trade flows, or at least of the pooled volume of exports and imports by commodity, is essential in the consideration of the important spillover and feedback effects deriving from inter-regional linkages (Miller and Blair, 2009, Ch. 3).

In spite of the recognized importance of inter-regional trade, in most countries, there are no completely reliable survey-based trade statistics. This has encouraged an increasing amount of research within the inter-regional input–output field devoted to the application of non-survey methods to inter-regional trade estimation. In spite of the important role played by transport statistics as a proxy for inter-regional flows, such statistics must be used cautiously, taking into consideration and overcoming the following limitations: (1) they do not cover service trading; (2) transport flows are expressed in physical units, requiring access to some value/volume relation; (3) regions with transport platforms

*Corresponding author. E-mail: ana.sargento@ipleiria.pt

ISSN 0953-5314 print; ISSN 1469-5758 online
http://dx.doi.org/10.1080/09535314.2011.574609

show an over-estimation of trade flows (Ramos, 2001); (4) flows shipped by manufacturers are not distinguished from flows shipped by resellers, leading to problems of double-counting (Alward et al., 1998), and (5) detailed region-to-region origin-destination matrices often have non-numerical entries such as "unreliable value" or other annotations (Schwarm et al., 2006), indicating a low degree of confidence in the estimate. Some authors have chosen to develop methods designed to deal with some of the above-mentioned problems (see, for example, Llano, 2004), while others have used transport statistics only as an indirect source of data in inter-regional trade estimation (as, for example, in Schwarm et al., 2006; and in Ferreira, 2008). Hence, it can be understood that non-survey methods do represent an alternative approach to the generation of undisclosed values of inter-regional trade or, at least, a means of complementing partial information concerning those flows.

The problem of inter-regional trade estimation using non-survey methods can be studied within the context of spatial interaction models: the aim is to estimate a set of flows among several origins and several destinations, separated in space. In the family of spatial interaction models, there are different methodologies that vary not only in their theoretical foundations, but also in their practical applicability, determined especially by data demand issues. It is, therefore, important to conduct an in-depth analysis of each of the alternatives, keeping in mind the objective of the model and its adequacy for the problem under study: trade flow estimation. The use of spatial interaction models in the context of inter-regional trade estimation is not new. Examples of such applications can be found in the literature (Wilson 1970; Kim et al. 1983; Alward et al. 1998; Schwarm et al. 2006). However, it is usual for researchers to opt for some specific type of model, without making a comparison between the results provided by that model and the existing alternatives. This is due to the fact that model results evaluation requires access to some benchmark values, which are typically unavailable (Hewings and Jensen, 1986; Liu and Vilain, 2004; Canning and Wang, 2006); and it is the non-existence of these data which serves as precisely the motivation for the application of the model in the first place. Nonetheless, we consider it paramount to investigate the relative accuracy of the various proposals, in order to evaluate the reasonability of using those non-survey methods as viable alternatives to survey methods in input–output table construction. Hence, in the present paper, the empirical comparison between different inter-regional trade estimation methods is made using European countries instead of regions, officially known inter-country trade flows, as the benchmark for model accuracy evaluation.

In this work, we assume the context of very limited information as the data scenario faced by the researcher: it is assumed that the only previously known information concerning trade flows consists of the total value, by product, that enters into each destination region and the total value, also by product, that is shipped from each origin region. The motivation for such an assumption derives from the fact that the conclusions of this study are intended to be used by regional input–output assemblers, who in most cases face the obstacle of a complete lack of data on inter-regional trade flows, except from those values of total exports and imports – usually obtained previously, along with the remaining components of the regional input–output table. The main objective of this paper is to make an empirical comparison between different methodologies of inter-regional trade estimation. This comparison is guided towards the following research questions:

- What is the degree of closeness of each estimated matrix to the real matrix of flows?
- Which method generates the most accurate estimated matrix?
- How sensitive are the values obtained in the final trade matrix to different estimating methods?
- How sensitive is the solution of the input–output model to the insertion of different inter-regional trade values? In other words, how important is the choice of inter-regional trade estimation method to the solution of the input–output model?

The answers to these questions are of extreme importance to any researcher who intends to use a non-survey method to estimate inter-regional trade. For example, if the solution of the input–output model is found to have a low sensitivity to the choice of trade-estimating method, then this can be used as an argument to opt for a simple non-survey method of trade estimation. Conversely, if the choice of method reflects greatly on the trade values and also on the results of the input–output model, then that choice should be made carefully and a pure non-survey method may not constitute a viable alternative.

The paper is organized in four sections, including this Introduction. The second section comprises a specification of the problem under study, as well as a brief review of the gravity model since this is the foundation of two of the proposals tested further on. Section 3 contains a description of the empirical application, in which a double-type comparison (in absolute and in analytical terms) is made between different methods of inter-regional trade estimation. Finally, Section 4 presents a summary of the main conclusions.

2. INTER-REGIONAL TRADE FLOW ESTIMATION THROUGH NON-SURVEY MODELS

As stated in Hewings et al. (2004, p. 13), "inter-regional interaction is not only inevitable, it is often the key to enhancing the success of a region". Among the different types of interaction that may exist between different regions, our interest is focused on inter-regional trade. This paper is concerned with the specific problem of estimating the values to be inserted into an Origin-Destination (O-D) matrix with the frame shown in Figure 1, depicting – for each product – inter-regional trade flows from each region of origin to each region of destination.

It is assumed that the column and row totals are previously known ('*roc*' stands for 'rest of the country'). Moreover, since only inter-regional trade flows are being estimated (and not intra-regional ones), the main diagonal values are previously set equal to zero.[1] Assuming a system with k regions of origin (denoted by a superscript r) and k regions of destination (denoted by a superscript s), then the problem consists of estimating the inter-regional shipments of j, x_j^{rs}, $r,\ s = 1, \ldots, k$, $(r \neq s)$, as illustrated by Figure 1.

[1] The estimation of these total gross inflows and outflows, as well as the assessment of intra-regional flows, by commodity and for each region, is another important problem to overcome before estimating net inter-regional trade flows. Such a problem is usually addressed at the single-region input–output table construction phase, for which there is also a vast stream of literature (see for example: Isserman, 1980; Eding et al., 1997; Lahr, 1998; Jackson, 1998; Piispala, 2000; Jensen-Butler and Madsen, 2003, or Sargento, 2009b, for a literature review on this theme).

FIGURE 1. Inter-regional trade flows of commodity j from region r to region s: x_j^{rs}.

Destination / Origin	Region 1	Region 2	...	Region k	Sum
Region 1	0	x_j^{12}	...	x_j^{1k}	$x_j^{1\,roc}$
Region 2	x_j^{21}	0	...	x_j^{2k}	$x_j^{2\,roc}$
...	...	...	0	...	...
Region k	x_j^{k1}	x_j^{k2}	...	0	$x_j^{k\,roc}$
Sum	$x_j^{roc\,1}$	$x_j^{roc\,2}$	...	$x_j^{roc\,k}$	x_j

Besides the best-known gravity-based models, several other approaches have been proposed by the literature in order to solve problems of inter-regional trade estimation such as the one just described. These range from the probabilistic models, established on the concepts of entropy and information theory (see, for example, Wilson, 1970; Snickars and Weibull, 1977; Roy and Thill, 2004; Batten, 1982; Batten and Boyce, 1986; Harrigan, 1990; Batten and Martellato, 1985), to neuronal networks models (Nijkamp et al., 2004), passing through behavior-based models (Isard, 1998; Cesario and Smith, 1975; Sheppard, 1978). In spite of the diverse theoretic foundations of such approaches, it has been shown that some of the different theoretical models lead to conclusions that "are more notable for their similarities than for their differences" (Batten and Boyce, 1986, p. 357). However, these models are quite different in their applicability to a context of very limited a priori information.

In the present paper, the methodologies proposed and tested rely heavily on the gravity-based equation, given its suitability for application in the absence of a prior matrix of flows, as well as the proven reasonability of the results produced, especially in empirical applications to trade (Sen and Smith, 1995; Llano et al., 2010).

Analytically, the basic equation that is used to express the gravity hypothesis on trade flows between origin r and destination s is:

$$x^{rs} = G\frac{(P^r)^{\alpha_1}(P^s)^{\alpha_2}}{(d^{rs})^{\alpha_3}} \tag{1}$$

in which: x^{rs} represents exports from origin r to destination s, G is a constant of proportionality, P^r and P^s express the sizes of origin r and destination s, with weights α_1 and α_2, respectively, d^{rs} represents spatial separation between each origin r and each destination s and α_3 is the so-called distance decay parameter, measuring the flow sensitivity to spatial separation.[2] When the column and row totals of Figure 1 are previously known, the estimated values must verify the following sum constraints: $\sum_s \tilde{x}^{rs} = x^r$ and $\sum_r \tilde{x}^{rs} = x^s$, and the model is doubly-constrained (Isard, 1998). The substitution of the known values of variables P^r, P^s and d^{rs} will most certainly produce a matrix of flows in which the row and column totals do not match with the ex ante values. The agreement with the sum constraints can, then, be assured through an iterative procedure of bi-proportional adjustment, akin to the RAS technique (Batten and Boyce, 1986; McDougall, 1999).

Obviously, the precise results obtained by the RAS-type iterative procedure are determined by the values considered in the starting matrix. In fact, taking the general case of k origins (and destinations), the problem corresponds to a system with $2k$ sum restrictions and k^2 elements x^{rs} to be determined.[3] Whenever $k > 2$, this system admits several solutions (Lahr and de Mesnard, 2004). Yet, the solution provided by the RAS iterative procedure tends to preserve, as much as possible, the structure of the initial matrix, changing it only by the minimum amount necessary to respect the row and column sum constraints (Jackson and Murray, 2004). This demonstrates that the starting matrix is a determinant of the final solution, leading us back to the fundamental issue: finding an adequate model to accurately generate the first estimate of inter-regional flows, i.e. finding the initial matrix.

3. ABSOLUTE AND ANALYTICAL COMPARISON BETWEEN INTER-REGIONAL TRADE ESTIMATION METHODS

This section presents a comparison between distinct inter-regional trade estimation methods, both in absolute and in analytical terms. First of all, it is necessary to clarify what we mean by 'absolute' and 'analytical' comparison. The absolute comparison relies on the differences observed among Origin-Destination matrices of trade flows generated by different inter-regional trade estimation methods. The analytical comparison goes further and involves the assessment of the impact on the multipliers obtained from the model, created by the insertion of different inter-regional trade values. In order to allow for this sort of comparison, it is necessary to have access to a complete multi-regional system upon which a multi-regional input–output model can be developed. Thus, the comparison was conducted in four stages, as listed below:

(1) Multi-regional input–output table assembly and development of the corresponding multi-regional input–output model.[4]

---

[2] This equation comprises some quite vague concepts, such as size and spatial separation (Sen and Smith, 1995), which allow for different interpretations. Spatial separation, for instance, can be expressed by physical distance or other concepts of separation, such as political or cultural distance. Also, the whole set of specific formulations that are consistent with the gravity hypothesis is vast, Equation (2) being a particular case. The debate over the different variables to express each of the above mentioned concepts and the different formulations of the gravity model is beyond the scope of the present paper, but is minutely discussed in Isard (1998).

[3] If the main diagonal elements are a priori considered to be null, then there will only be $k^2 - k$ elements to be determined.

[4] Given that our objective is to use the conclusions from this work in a multi-regional table assembly, we opt for the designations 'region' and 'multi-regional' in the context of this empirical application, even though we are actually dealing with a system of countries.

(2) Estimation of O/D trade matrices, for each product group, on the basis of three different methodologies.
(3) Comparison between the three different matrices obtained before (absolute comparison).
(4) Consecutive insertion of the different O/D matrices into the multi-regional input–output system and model simulation in a context of final demand change (analytical comparison).

3.1. Multi-regional Input–Output System Development

As mentioned before, the multi-regional input–output system was developed considering a set of 14 countries belonging to the European Union before enlargement (15 minus one, since Belgium and Luxembourg are considered jointly). The input–output data came from the ESA 95 Input–Output Table database (taking 2002 as the reference year), provided by EUROSTAT, which is available on-line at the EUROSTAT webpage. The original EUROSTAT tables are published in rectangular format (thus including make and use tables) and the Use tables are composed of intermediate and final use flows, which include not only domestically produced products, but also imported ones. Thus, we label them as Total Use tables. Besides, Supply tables are evaluated at basic prices (*bp*) including a transformation into purchasers' prices (*pp*) and Use tables are at purchasers' prices. The bilateral trade data came from the OECD Bilateral Trade Database, Edition 2002. This trade database is used to estimate the distribution of intra-regional trade among the 14 countries included in the system. Given the discrepancies that occurred between export-based data and import-based data (commonly known as the 'mirror statistics puzzle', meaning that the total exports reported by country A differ from the sum of total imports coming from country A as reported by the other countries), a short-cut method was adopted, that being to consider the mean value between the export-based and the import-based OECD data. Additionally, the totals of intra-regional imports and intra-regional exports were imposed by the values indicated by the EUROSTAT input–output tables. In order to avoid an extremely heavy multi-regional table, the original classification embodied in the EUROSTAT input–output tables was aggregated into six categories of products and industries, namely: A+B – 'Products of agriculture, hunting, forestry and fishing', C – 'Mining and quarrying', D – 'Industry', E – 'Electricity, gas and water supply', F – 'Construction', and G to P – 'Services'.

Before explaining how the multi-regional input–output system was developed, an elucidation must be made. The assembly of the multi-regional system upon which the model was based was not the ultimate objective of this empirical study. Instead, it was merely an instrument to be used in the comparison between the different inter-regional trade estimation methods. Thus, the high level of aggregation considered, as well as the several simplifying hypotheses adopted, should not be over-emphasized. In fact, all the subsequent experiments will be made using the assembled input–output system as the common starting point. Hence, the conclusions obtained by comparing those different experiments to one another should not be affected to a great extent by the hypotheses assumed in the model construction stage.

Making use of the above described database, the assembly process involved the assumption of certain simplifying hypotheses, in order to obtain the final multi-regional make and use system, at basic prices, with import content expurgated from intermediate and final use flows and aggregated in six product categories.[5] The basic structure of the final table can be observed in Figure 2, which illustrates the case for three regions (j refers to the products and i to the industries).

Having such a multi-regional system as a starting point (and using the example of three regions), the multi-regional input–output model can be developed as follows. Let the bold notation designate the column vectors and the matrices, corresponding to the variables depicted in Figure 2. $\mathbf{i}$ is a 1-filled column-vector. For example, $\mathbf{U^{AB}}$ stands for the matrix of intermediate consumption composed by flows u_{ji}^{AB}, and $\mathbf{y^{AB}}$ represents the vector composed by flows y_j^{AB}. Hence, we can write the system:

$$
\begin{aligned}
\mathbf{v^A} &= \mathbf{U^{AA}i} + \mathbf{y^{AA}} + \mathbf{U^{AB}i} + \mathbf{y^{AB}} + \mathbf{U^{AC}i} + \mathbf{y^{AC}} + \mathbf{y^{A\,ROW}} \\
\mathbf{v^B} &= \mathbf{U^{BA}i} + \mathbf{y^{BA}} + \mathbf{U^{BB}i} + \mathbf{y^{BB}} + \mathbf{U^{BC}i} + \mathbf{y^{BC}} + \mathbf{y^{B\,ROW}} \\
\mathbf{v^C} &= \mathbf{U^{CA}i} + \mathbf{y^{BA}} + \mathbf{U^{CB}i} + \mathbf{y^{CB}} + \mathbf{U^{CC}i} + \mathbf{y^{CC}} + \mathbf{y^{C\,ROW}}
\end{aligned}
\tag{2}
$$

Additionally, making use of the intra-regional input coefficients and of the inter-regional trade coefficients, defined as:

$$
\begin{aligned}
q_{ji}^{AA} &= \frac{u_{ji}^{AA}}{g_i^A} \Rightarrow u_{ji}^{AA} = q_{ji}^{AA} g_i^A \\
q_{ji}^{AB} &= \frac{u_{ji}^{AB}}{g_i^B} \Rightarrow u_{ji}^{AB} = q_{ji}^{AB} g_i^B
\end{aligned}
\tag{3}
$$

for the example of regions A and B, and the equivalent assumptions for the other cases, the

[5] The conversion of purchasers' prices into basic prices as well as the elimination of the import content from intermediate and final uses were operated applying the commonly used proportionality hypotheses. Concerning imports, for example, a constant average import propensity was assumed, meaning that, for each product, the same rate of imports (coming from the rest of the world and from the rest of the regions involved in the system) is embodied in intermediate and final use of that product. Oosterhaven and Stelder (2007) have tested the magnitude of errors created by such an assumption, in their comparison of four alternative non-survey inter-country input–output table construction methods, for nine Asian countries and the USA. In one of the non-survey input–output tables, they assume that there is no imports matrix and use the imports proportionality assumption to indirectly estimate it. The comparison between this table and the benchmark (which is a semi-survey-based inter-country table) allows the authors to conclude that in general, "The tests show that the impact of using self-sufficiency ratios to estimate the domestic flows is small [. . .]" (Oosterhaven and Stelder, 2007, p. 258). A complete explanation of the remaining hypotheses adopted is available at Sargento (2009a).

FIGURE 2. Basic structure of the multi-regional make and use system.

		Region A			Region B			Region C			ROW	SUM
		Products	Industries	Final demand	Products	Industries	Final demand	Products	Industries	Final demand		
Region A	Products	---	$\left[u_{ji}^{AA}\right]$	y_j^{AA}	---	$\left[u_{ji}^{AB}\right]$	y_j^{AB}	---	$\left[u_{ji}^{AC}\right]$	y_j^{AC}	$y_j^{A\,ROW}$	v_j^A b.p.
	Industries	$\left[v_{ij}^{A}\right]$	---	---	---	---	---	---	---	---	---	g_i^A b.p.
Region B	Products	---	$\left[u_{ji}^{BA}\right]$	y_j^{BA}	---	$\left[u_{ji}^{BB}\right]$	y_j^{BB}	---	$\left[u_{ji}^{BC}\right]$	y_j^{BC}	$y_j^{B\,ROW}$	v_j^B b.p.
	Industries	---	---	---	$\left[v_{ij}^{B}\right]$	---	---	---	---	---	---	g_i^B b.p.
Region C	Products	---	$\left[u_{ji}^{CA}\right]$	y_j^{CA}	---	$\left[u_{ji}^{CB}\right]$	y_j^{CB}	---	$\left[u_{ji}^{CC}\right]$	y_j^{CC}	$y_j^{C\,ROW}$	v_j^C b.p
	Industries	---	---	---	---	---	---	$\left[v_{ij}^{C}\right]$	---	---	---	g_i^C b.p.
ROW		---	$\sum_j u_{ji}^{ROW\,A}$	---	---	$\sum_j u_{ji}^{ROW\,B}$	---	---	$\sum_j u_{ji}^{ROW\,C}$	---		
Trade and transp marg		---	d_i^A	---	---	d_i^B	---	---	d_i^C	---		
Taxes less sub on prod		---	l_i^A	---	---	l_i^B	---	---	l_i^C	---		
Total Int. Consumption		---	IC_i^A p.p.	---	---	IC_i^B p.p.	---	---	IC_i^C p.p.	---		
Value Added		---	VA_i^A	---	---	VA_i^B	---	---	VA_i^C	---		
SUM		v_j^A b.p.	g_i^A b.p.	---	v_j^B bp.	g_i^B b.p.	---	v_j^C bp.	g_i^C b.p.	---		

Notation:

- u_{ji}^{AA} – generic element of the regional production Use matrix $\mathbf{U^{AA}}$, which indicates the amount of product *j* produced in region A that is used by industry *i* in region A.
- u_{ji}^{AB} – generic element of the Use matrix $\mathbf{U^{AB}}$, which indicates the amount of product *j* produced in region A that is used by industry *i* in region B.
- $u_{ji}^{ROW\,B}$ – amount of product *j* produced in the rest of the world that is used by industry *i* in region B.
- y_j^{AA} – amount of product *j* produced in region A which is used for final demand in the region.
- y_j^{AB} – amount of product *j* produced in region A which is used for final demand in region B.
- $y_j^{A\,ROW}$ – amount of product *j* produced in region A which is exported to the rest of the world.
- v_{ij}^{A} – generic element of the Make matrix in region A. It represents the amount of product *j* produced by industry *i* in region A.
- d_i^A (l_i^A) – total amount of margins (taxes, less subsidies) embodied in intermediate consumption of industry *i*; it corresponds to the column sum of the matrix of margins (net taxes), for industry *i*; the matrix of margins is computed under certain hypotheses, exposed further on.
- IC_i^A – total intermediate consumption of industry *i* in region A.
- VA_i^A – value added of industry *i* in region A.
- v_j^A – total production of product *j* in region A.
- g_i^A – total production of industry *i* in region A.

system becomes:

$$\begin{aligned}
\mathbf{v^A} &= \mathbf{Q^{AA}g^A} + \mathbf{y^{AA}} + \mathbf{Q^{AB}g^B} + \mathbf{y^{AB}} + \mathbf{Q^{AC}g^C} + \mathbf{y^{AC}} + \mathbf{y^{A\,ROW}} \\
\mathbf{v^B} &= \mathbf{Q^{BA}g^A} + \mathbf{y^{BA}} + \mathbf{Q^{BB}g^B} + \mathbf{y^{BB}} + \mathbf{Q^{BC}g^C} + \mathbf{y^{BC}} + \mathbf{y^{B\,ROW}} \\
\mathbf{v^C} &= \mathbf{Q^{CA}g^A} + \mathbf{y^{CA}} + \mathbf{Q^{CB}g^B} + \mathbf{y^{CB}} + \mathbf{Q^{CC}g^C} + \mathbf{y^{CC}} + \mathbf{y^{C\,ROW}}
\end{aligned} \tag{4}$$

Besides the common fixed input coefficients hypothesis, another assumption must be made – a proposition that relates industry's output with commodity's output. To do so, we assume that each product is produced in fixed proportions by several industries, implying that the structure inherent in each column of the make matrix is assumed invariant:[6]

$$s_{ij}^A = \frac{v_{ij}^A}{v_j^A} \Rightarrow v_{ij}^A = s_{ij}^A v_j^A \tag{5}$$

In matrix terms, this corresponds to:

$$\mathbf{V^A i} = \mathbf{S^A v^A} \Leftrightarrow \mathbf{g^A} = \mathbf{S^A v^A} \tag{6}$$

Obviously, the same applies to regions B and C. Introducing Equation 6 into Equation 4, we get:

$$\begin{aligned}
\mathbf{v^A} &= \mathbf{Q^{AA}S^A v^A} + \mathbf{Q^{AB}S^B v^B} + \mathbf{Q^{AC}S^C v^C} + \mathbf{y^{AA}} + \mathbf{y^{AB}} + \mathbf{y^{AC}} + \mathbf{y^{A\,ROW}} \\
\mathbf{v^B} &= \mathbf{Q^{BA}S^A v^A} + \mathbf{Q^{BB}S^B v^B} + \mathbf{Q^{BC}S^C v^C} + \mathbf{y^{BA}} + \mathbf{y^{BB}} + \mathbf{y^{BC}} + \mathbf{y^{B\,ROW}} \\
\mathbf{v^C} &= \mathbf{Q^{CA}S^A v^A} + \mathbf{Q^{CB}S^B v^B} + \mathbf{Q^{CC}S^C v^C} + \mathbf{y^{CA}} + \mathbf{y^{CB}} + \mathbf{y^{CC}} + \mathbf{y^{C\,ROW}}
\end{aligned} \tag{7}$$

which can still be represented by:

$$\begin{aligned}
\mathbf{v^A} &= \mathbf{Q^{AA}S^A v^A} + \mathbf{Q^{AB}S^B v^B} + \mathbf{Q^{AC}S^C v^C} + \mathbf{y^{A\bullet}} \\
\mathbf{v^B} &= \mathbf{Q^{BA}S^A v^A} + \mathbf{Q^{BB}S^B v^B} + \mathbf{Q^{BC}S^C v^C} + \mathbf{y^{B\bullet}} \\
\mathbf{v^C} &= \mathbf{Q^{CA}S^A v^A} + \mathbf{Q^{CB}S^B v^B} + \mathbf{Q^{CC}S^C v^C} + \mathbf{y^{C\bullet}}
\end{aligned} \tag{8}$$

in which $\mathbf{y}^{A\bullet}$ represents final demand for region A's production.

---

[6] This corresponds to the hypothesis commonly known as the Industry Technology-based Assumption (ITA), and it implies that all goods produced by an industry are produced with the same input structure, meaning that there is one technology assigned to each industry. The discussion of the reasonability of this hypothesis, as well as the analysis of alternative assumptions goes beyond the objectives of this paper.

If we take the following block matrices and vectors:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v^A} \\ \mathbf{v^B} \\ \mathbf{v^C} \end{bmatrix}; \mathbf{y} = \begin{bmatrix} \mathbf{y^{A\bullet}} \\ \mathbf{y^{B\bullet}} \\ \mathbf{y^{C\bullet}} \end{bmatrix}; \mathbf{g} = \begin{bmatrix} \mathbf{g^A} \\ \mathbf{g^B} \\ \mathbf{g^C} \end{bmatrix}; \mathbf{Q} = \begin{bmatrix} \mathbf{Q^{AA}} & \mathbf{Q^{AB}} & \mathbf{Q^{AC}} \\ \mathbf{Q^{BA}} & \mathbf{Q^{BB}} & \mathbf{Q^{BC}} \\ \mathbf{Q^{CA}} & \mathbf{Q^{CB}} & \mathbf{Q^{CC}} \end{bmatrix} \text{and}$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{S^A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S^B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S^C} \end{bmatrix}$$

we may write:

$$\begin{aligned} &\mathbf{v} = \mathbf{QSv} + \mathbf{y} \\ &(\mathbf{I} - \mathbf{QS})\mathbf{v} = \mathbf{y} \\ &\mathbf{v} = (\mathbf{I} - \mathbf{QS})^{-1}\mathbf{yy} \\ &\mathbf{g} = \mathbf{S}(\mathbf{I} - \mathbf{QS})^{-1}\mathbf{y} \end{aligned} \tag{9}$$

These equations allow the assessment of the impacts on the production of the several regions (concerning the effect on regional product supply as well as on regional industry production) caused by changes in the vector of final demand for regional production. Such impact analysis implies that the elements of the inverse matrix $(\mathbf{I} - \mathbf{QS})^{-1}$ remain unaltered in the face of exogenous shocks. Thus, this involves not only the assumption of constant input coefficients q_{ji}, but also the assumption of constant market shares s_{ij}.

3.2. Alternative Methodologies to Estimate Inter-regional Trade

In this empirical exercise, we are trying to find the input data for a constrained matrix balancing problem, which involves two stages:

(1) generating data priors for the O/D matrix;
(2) updating the prior O/D matrix in order to meet the row and column totals' restrictions.

It should be noted that we know in advance the true value of the whole trade O/D matrix (and not only the column and row totals) in respect of the 14 European countries in our database, although we assume that the inner part of the matrix is not known, which corresponds to the usual information context when we are dealing with inter-regional rather than international trade. Actual international trade is used as a benchmark for testing the non-survey estimating methods.

The purpose of this Section is to describe the alternative methodologies applied to the estimation of the core part of the O/D matrices. All the methodologies share a common point: they depart from the same information on the row and column totals for the O/D matrices. However, as noted by Hulu and Hewings (1993), the fact that all the estimates observe the same sum conditions does not guarantee that the final estimates are the same;

in the authors' words "the bi-proportional adjustment process only guarantees accuracy at the margins [...]" (Hulu and Hewings, 1993, p. 142). Thus, it is expected that different initial matrices in stage (1) also originate different final matrices, after the adjustment procedure of step (2). The analysis of the results provided by the several methodologies described below will allow the sensitivity of the final estimates to the different priors to be inferred.

The different methodologies are named as Experiments, ranging from Experiment 1 to Experiment 3. As mentioned already, in practical applications, the gravity model continues to be one of the most frequently used among spatial interaction models. This empirical exercise is not an exception: in two of the three methodologies suggested to generate the priors for the O/D matrix, the formulation is based on the gravitational formula. The remaining method consists of a straightforward way of generating the initial matrix – a data pooling structure, employed with the aim of understanding the sensitivity of the final matrix to completely different initial estimates.

Experiment 1. Simple Gravity Model

For each of the six categories of products, the initial stage (1) O/D matrix (with flows represented by $\tilde{x}_0^{rs}$) was estimated through the application of a simplified gravity-based model, in which almost all the unknown parameters were arbitrarily set equal to one. More precisely, we have made:

$$\begin{aligned} \tilde{x}_0^{rs} &= G^r \frac{P^r P^s DS^r}{d^{rs}} \\ G^r &= \tilde{x}_0^r \left(\sum_s \frac{P^r P^s DS^r}{d^{rs}} \right)^{-1} \end{aligned} \tag{10}$$

The constant of proportionality G^r is the one that guarantees the exact observance of the rth row summing up constraint: $\sum_s \tilde{x}_0^{rs} = x^r$. It is also considered as an additional variable, DS^r, standing for the Degree of Specialization of the origin region in exporting the product under study; such a variable differs by region and by product.[7] Afterwards, the initial values were adjusted to the additivity restrictions, using the RAS method.

[7] In Sargento (2009a) it was demonstrated that the inclusion of this additional variable improves the performance of the gravity model. For each product j, The Degree of Specialization is computed as:

$$DS_j^r = \frac{x_j^r \Big/ \sum_{j=1}^{6} x_j^r}{\sum_{r=1}^{14} x_j^r \Big/ \sum_{r=1}^{14} \sum_{j=1}^{6} x_j^r}$$

The numerator of this index represents the weight of product j on origin r's total exports; the denominator indicates the weight of product j on all origins' exports. Actually, this is no more than a Location Quotient, computed using exports (x) as the variable of specialization.

In this exercise, we used per capita GDP for the origin and destination variables (the source data were: (1) Population, for year 2001, in thousands (OECD, 2005a) and (2) Gross Domestic Product, year 2001, in USD and current prices (OECD, 2005b)). As for distances, we applied a measure that is slightly different from the traditional one. Usually, distance between countries used in this kind of study consists of the straight line distance between capital cities. However, two problems are associated with this measure: first, trade flows occur between each and every locale of each economic area and not only between capital cities; secondly, capital cities are not always the central focus of a country, in geographic terms. Thus, we opted for using a weighted distance between each country, which takes into account the several distances between all regions of country A and all regions of country B, instead of considering merely the distance between two specific points of the country. Considering two countries, A and B, with two regions each, 1 and 2, the distance between A and B was computed as:[8]

$$d_{AB}=\frac{GDP_{A1}GDP_{B1}d_{A1B1}+GDP_{A1}GDP_{B2}d_{A1B2}+GDP_{A2}GDP_{B1}d_{A2B1}+GDP_{A2}GDP_{B2}d_{A2B2}}{GDP_{A1}GDP_{B1}+GDP_{A1}GDP_{B2}+GDP_{A2}GDP_{B1}+GDP_{A2}GDP_{B2}} \tag{11}$$

Experiment 2. Data Pooling Structure

Similarly to what was done by Hulu and Hewings (1993) in their attempt to assemble an inter-regional input–output table for Indonesia, we began by assuming that, in each country, the total amount of imports coming from the remaining 13 countries was equally divided by each of those 13 supplying countries:

$$\tilde{x}_0^{rs}=\frac{x^s}{13}, \text{ in which } \sum_r \tilde{x}_0^{rs}=x^s \tag{12}$$

So, the column sums of this initial matrix were necessarily equal to the reference values. However, given that the row sums did not verify the correspondent additivity constraints, the RAS method was adopted, as in Experiment 1.

Experiment 3. Enhanced Gravity Model

In Experiment 3, the initial values of trade flows were determined by an equation similar to Equation 10, except for the fact that the distance decay parameter was estimated by an alternative method, rather than being merely set equal to one. More precisely, this parameter, which we name β, is determined in order to minimize the following indicator of error: $I=\sum_s\left|\sum_r x^{rs}-\sum_r \tilde{x}_0^{rs}\right|\Big/\sum_r\sum_s x^{rs}$. Such minimization leads to product-specific distance decay parameters, since row and column totals are also different

[8] The distances between NUTs II were calculated using the polygon centroid's (x,y) coordinates for each one.

for each group of products.[9] As stated in Schwarm et al. (2006 p. 87), the ideal procedure to estimate the values of the gravity parameters should be to “minimize the absolute differences between estimated and observed flows”. Given that observed flows are supposedly unknown in our case, we have opted for constructing an indicator of error using only the available information on row totals (of course, an equally valid alternative would be to consider the information on column totals). This methodology had already been applied in previous empirical exercises, such as in Ramos and Sargento (2003) and in Sargento (2007), and it was demonstrated that, on average, the distance between the final matrix and the real one was a bit smaller than when all the parameters were set equal to one. Again, RAS adjustment was applied in order to verify the row and column sum constraints.

3.3. Comparison between the Results of the Different Experiments

Before proceeding to presentation of the results, it is necessary to note that the estimation of origin-destination matrices through the experiments described above was made only in respect of four of the six product categories used in the multi-regional input–output system: ‘A+B – Products of agriculture, hunting, forestry and fishing’, ‘C – Mining and quarrying’, ‘D – Industry’ and ‘E – Electricity, gas and water supply’. This option was justified by the fact that, for the remaining categories ‘F – Construction’ and ‘G to P – Services’, we do not have ‘real’ trade data from the OECD Bilateral Trade Database, which only covers the trade of goods, and excludes the other sectors. The several O/D matrices obtained through the different methodologies were compared with each other and also to the corresponding real matrices, using the Standard Total Percentage Error as a measure of distance between matrices:[10]

$$STPE = 100 \cdot \frac{\sum_{s}\sum_{r}\left|x^{rs} - \tilde{x}_1^{rs}\right|}{\sum_{s}\sum_{r}x^{rs}} \tag{13}$$

The results are presented in Table 1. In the first column of the table, the performance of each of the different experiments can be evaluated: each estimated matrix is compared with the corresponding real matrix. In the remaining columns, a comparison is made between the results provided by the different estimates, meaning that an evaluation is made of the sensitivity of the final results to each methodology. A weighted average was computed for each experiment, which expresses the error associated with the corresponding methodology considering all product categories, weighting by their relative importance in aggregate trade. The errors concerning category ‘D – Industry’ and ‘A+B – Products of agriculture, hunting, forestry and fishing’ are the most relevant, given the higher relative weights these products have in international trade.

---

[9] In Sargento (2007), a spatial econometric application of the gravity model was applied to study the behaviour of inter-country trade flows concerning ten different manufactured products. In such a study it was demonstrated, as expected, that specific products exhibit different distance decay parameters.

[10] For a discussion of the alternative measures for matrix comparison, please refer to Jackson and Murray (2004) and Lahr (1998), for example.

TABLE 1. Summary results from the comparison between the three different inter-regional trade estimation methods.

STPE versus…		REAL	EXP 2	EXP 3
EXP 1	A + B	31.9%	44.6%	19.9%
	C	43.0%	34.6%	15.3%
	D	28.9%	43.8%	2.0%
	E	72.7%	37.4%	45.7%
	aggregate	**28.4%**	**43.5%**	**2.1%**
EXP 2	A + B	44.9%		24.9%
	C	35.7%		48.9%
	D	28.1%		41.9%
	E	67.4%		55.8%
	aggregate	**28.2%**		**41.5%**
EXP 3	A + B	31.4%		
	C	50.9%		
	D	27.6%		
	E	78.0%		
	aggregate	**27.2%**		

From the results, three main comments can be made:

(1) Taking the first column and the aggregate error for each experiment as a reference, we conclude that Experiment 3 (gravitational model in which the distance parameter is computed through the minimization of indicator *I*) is the one that originates the smaller aggregate error against the real values. The simple gravitational model (Experiment 1) seems to generate the highest aggregate error. This result confirms what had been verified in the ten-product application used in Sargento (2007), implying that the introduction of an independent estimate for the distance-decay parameter, instead of considering it to equal 1, may represent an improvement in the results. Nevertheless, the differences among aggregate errors are small when the comparison is made against the real values.
(2) The errors generated by the three experiments are quite high for some products (achieving 78% in product 'E', Experiment 3), but they are lower in the most representative products. If we limit the analysis to the 'Industry' case, the most relevant in international trade, we may state that the non-survey methods proposed here produce quite reasonable results.
(3) Observing the mutual differences between the three experiments, we conclude that these are larger between Experiments 1 and 2 and between Experiments 2 and 3, than between Experiments 1 and 3. In other words, the only case in which we do not use a gravitational formula as a starting point – Experiment 2 – generates more outlying results, demonstrating that the way by which initial estimates are obtained is not innocuous.

In the particular case of these data and the set of methods used, the results allow us to conclude that a gravity-based model to generate the initial values jointly with the RAS adjusting procedure originates the best results (Experiment 3 provides the closer matrices). Still, some large errors observed in the first column, as well as the considerable differences existing among the three experiments, constitute an impetus to perform an analytical comparison between the different methodologies. As explained before, the ultimate aim is to assess the extent to which these different estimates are reflected in different results in practical applications of the input–output model. In other words, how important is the accuracy of the inter-regional trade estimates to the model accuracy?

3.4. Input–Output Model Sensitivity to the Alternative Methodologies of Inter-regional Trade Estimation

The analytical comparison between alternative inter-regional trade estimation methods consists of a sort of exercise that is not usually performed.[11] With such an exercise, we aim to assess the sensitivity of the input–output model solution to the insertion of the O/D matrices derived from each of the three experiments previously described. In other words, the division of the intra-EU imports between the 13 remaining supplying countries is made using the percentages given by the O/D matrices derived from each experiment, instead of the percentages derived from the OECD-based matrices. Hence, we obtain a set of three different multi-regional systems (besides the reference one), each one corresponding to a different inter-regional trade estimation methodology. The only difference among the obtained multi-regional tables consists precisely in the inter-regional Use tables, caused by the different inter-regional trade data considered (in terms of Figure 2, the difference relies on the off-diagonal components of the multi-regional table, given that the intra-EU imports distribution varies from one table to another). This implies that the inter-regional trade coefficients embodied in the off-diagonal components of the block matrix $\mathbf{Q}$ are being substituted by different values in each different experiment. Observing Equation 9, it becomes clear that this has an impact on the model solution. The empirical exercise reported in this section aims to investigate the extent to which that solution is affected.

Let us take the following simulation exercise: considering the real growth rate observed along the elements of final demand vector ($\mathbf{y}$) from year 2000 to 2001, what is the model estimate for growth in aggregate Gross Value Added (GVA) resulting from the use of the different multi-regional tables (the reference one and the other three)? In order to answer such question, the following equations were applied, computing the impact of final demand change on product and industry output vectors:[12]

$$\begin{aligned} \Delta\mathbf{v} &= (\mathbf{I} - \mathbf{QS})^{-1}\Delta\mathbf{y} \\ \Delta\mathbf{g} &= \mathbf{S}(\mathbf{I} - \mathbf{QS})^{-1}\Delta\mathbf{y} \end{aligned} \tag{14}$$

[11] Exceptions may be found in Oosterhaven and Stelder (2007) and also in Robinson and Liu (2006), although each has different research aims.

[12] It must be noted that the final demand vectors for each country, recorded in the EUROSTAT Use tables for year 2001, were previously converted from current prices to constant prices and also from *pp* total use flows into partial flow vectors, valuated at *bp*. In the latter conversion, the previously referred proportionality hypotheses were applied.

Finally, a Value Added coefficient was considered to compute the vector of changes in Value Added. Taking $va_i^A = \frac{VA_i^A}{g_i^A}$ as the proportion of Value Added included in the output of industry I in region A, we may write: $VA_i^A = va_i^A g_i^A$. If we assume constant value added coefficients, we have, in matrix terms:

$$\begin{aligned} \mathbf{\Delta VA} &= \mathbf{v\hat{a}} \cdot \mathbf{\Delta g} \Leftrightarrow \\ \mathbf{\Delta VA} &= \mathbf{v\hat{a}} \cdot \mathbf{S}(\mathbf{I} - \mathbf{QS})^{-1}\mathbf{\Delta y} \end{aligned} \tag{15}$$

in which $\mathbf{v\hat{a}}$ represents a diagonal block matrix with the country-specific diagonal matrices of value added coefficients in the main diagonal. This input–output model estimate for real GVA growth rate was first computed using the real data on inter-regional trade and subsequently calculated using the three experiments' estimate for trade flows. Each of the three estimated GVA growth rates derived from the inclusion of the trade flow estimates is compared with the input–output model estimate for GVA growth rate, when real trade data are considered – which we designate by 'reference GVA growth rate'. It must be noted, however, that this reference GVA growth rate already embodies some error, given the hypotheses implicit in the impact analysis represented by Equations 15, namely, the fact that input coefficients, industry market shares, and value added proportions are all assumed invariant. Nevertheless, the objective of this application is not to accurately calculate growth rates through the model, but rather to provide a reference to evaluate the deviations generated by the three inter-regional trade estimation methods.

An extraction of the full results of this sensitivity analysis, taking the French economy as a representative example, is displayed in Tables 2 and 3.[13] In Table 2, we may observe GVA growth rate estimated by the model, for each industry, as well as for the aggregate economy, using the four versions of the multi-regional table (the reference one and the other three). For each experiment, there is also a column reporting the difference (in percentage points) between the GVA growth rate obtained in that experiment and the one obtained using real inter-regional trade data. All the differences are computed in absolute percentage points. The 'average' column is calculated taking the mean of all the differences (labeled as 'diff.' in the table) obtained from the three experiments. This reflects the mean error resulting from using non-survey trade values instead of using the real values. Looking at the Industry sector – D, for example, we may conclude that: calculating inter-regional trade through a simple gravity model (Exp. 1) makes the predicted growth rate for Value Added 0.69% instead of 0.68%.

In Table 3 the same indicators are used, but the comparison is made among the three experiments and not with the reference table.

In order to classify the sensitivity of the input–output model solution to the insertion of inter-regional trade estimates, we considered that absolute differences against the real values to be moderately high if they exceeded the third quartile by more than 1.5 times the inter-quartile range of the whole distribution of differences (using the rule of thumb

[13] Only an extraction of the results is presented in the paper. However, a deep analysis was made of the full tables of results, including all the 14 countries and the six product groups being considered. These full tables are available, upon request, from the corresponding author.

TABLE 2. Comparison between the GVA growth rate using trade estimates from each individual experiment and the reference GVA growth rate.

			Exp 1		Exp 2		Exp 3		
		Reference GVA growth %	growth in GVA %	diff.	growth in GVA %	diff.	growth in GVA %	diff.	average diff.
F R A	A + B	0.67%	0.67%	0.00%	0.65%	0.02%	0.67%	0.01%	0.01%
	C	1.15%	1.14%	0.00%	1.12%	0.02%	1.14%	0.01%	0.01%
	D	0.68%	0.69%	0.01%	0.66%	0.02%	0.69%	0.01%	0.01%
	E	3.96%	3.96%	0.00%	3.95%	0.02%	3.96%	0.00%	0.00%
	F	2.20%	2.20%	0.00%	2.20%	0.00%	2.20%	0.00%	0.00%
	G to P	1.66%	1.66%	0.00%	1.65%	0.00%	1.66%	0.00%	0.00%
	AGGR.	1.52%	1.52%	0.00%	1.51%	0.01%	1.52%	0.00%	0.00%

TABLE 3. Comparison of the GVA growth rates generated by the different experiments.

		Exp 2 against Exp 1	Exp 3 against Exp 1	Exp 3 against Exp 2	average diff.
F R A	A + B	0.022%	0.003%	0.019%	0.015%
	C	0.020%	0.002%	0.018%	0.014%
	D	0.032%	0.001%	0.031%	0.021%
	E	0.009%	0.001%	0.008%	0.006%
	F	0.001%	0.000%	0.001%	0.001%
	G to P	0.004%	0.000%	0.004%	0.003%
	AGGREGATE	0.010%	0.000%	0.009%	0.007%

suggested by Tukey (1977) to define outlier thresholds). Using such boundaries, and taking the full set of results from which Table 2 was extracted, we found that a small fraction of the computed differences (only 9.5%) was beyond the limit mentioned above. The three experiments are responsible for a roughly identical fraction of these outliers (35.2% for Experiment 1, 31.5% for Experiment 2 and 33.3% for Experiment 3).

Concerning the complete comparison between experiments (the full version of Table 3), the percentage of moderately high differences (applying the same thresholds) ascends to 20% (of these, 37.1% of the cases come from Experiment 1, 25.7% from Experiment 2, and 37.1% from Experiment 3). We may also observe that the mean difference between Experiments 3 and 1, which are gravity-based experiments, is non-significant, and smaller than between each of those experiments and Experiment 2.

Given these observations, we may conclude that, in general, the multi-regional input–output model shows a moderate sensitivity to the insertion of different estimates for inter-regional trade. The results do not reject the reasonability of using indirect estimates for inter-regional trade, given that large deviations in the results of the model are an exception. Nevertheless, it cannot be stated that the choice of one specific method among several alternatives is completely innocuous. Even recognizing that the 'choice' is most frequently constrained by the availability of information, the researcher must be aware that the results of the model will necessarily be affected, although only in a moderate manner.

4. CONCLUSIONS AND FUTURE DEVELOPMENTS

The main objective of this paper was to study different inter-regional trade estimation methodologies and make an empirical comparison between them.

A practical exercise was conducted, consisting of an examination of two distinct formulations of the gravity model as a generator of undisclosed values of inter-regional trade vis-à-vis a data pooling structure. The latter, consisting of evenly distributing the amount of imports by each supplier country, was applied with the aim of assessing the impact of a straightforward method for the prior matrix, rather than using a widely investigated model such as the gravity model. Each of the three different methods generated the corresponding prior matrices of flows, which were afterwards adjusted in order to comply with the previously known row and column totals, by means of the RAS method.

The comparison between the results provided by these different methodologies allowed us to infer three main conclusions.

First, among the three experiments, the gravity-based model with an independent estimate of the distance decay parameter seems to originate the most accurate matrix (although the differences between aggregate errors obtained from Experiments 1 to 3 are small).

Secondly, the initial matrix seems to have an influence on the final results. In fact, when comparing the different experiments with each other, we have concluded that the only case that is not gravity-based – Experiment 2 – generates more outlying results, demonstrating that the way in which initial estimates are obtained is not innocuous.

Finally, the mean errors generated by the three experiments are not very high in the products that are most representative in international trade (around 28%). Thus, we may state that the non-survey methods proposed here produce quite reasonable results.

In order to assess the sensitivity of the input–output model to the insertion in the input–output system of the estimated Origin-Destination matrices, an analytical comparison of the different methodologies was also made. This implied the construction of a simplified multi-regional input–output system, involving the 14 European countries of the sample, using a dataset composed of the individual make and use tables, and bilateral trade data. The impact on the model results was measured through the different GVA growth rates estimated as a consequence of an exogenous change in final demand. We have concluded that the results of the input–output model were not greatly affected by the consideration of different trade flow values, since large deviations between the obtained growth rates were the exception and not the rule.

The main practical contribution of this paper consists precisely of the conclusions that can be drawn from the absolute and analytical comparison between the different trade estimation methods, namely: (1) among the several experiments performed, the one that generated the most accurate matrix corresponded to a gravity-based model, with independent estimation of the distance decay parameter and using RAS as the adjusting procedure; (2) the impacts on the input–output model of using differently estimated trade flows are only moderate – thus, the results do not reject the reasonability of using indirect estimates for inter-regional trade.

Although it is not advisable to generalize these results, given that they were obtained from a particular set of data and using a specific set of hypotheses, we consider that these practical contributions are most relevant to regional input–output researchers. In fact, these results provide actual estimates (based on an actual multi-regional input–output data system and model) for input–output model sensitivity to the inclusion of different inter-regional trade values. This is especially significant for those who intend to assemble an input–output model in a context of absent information on inter-regional trade flows (the most frequent situation at the sub-national level).

In spite of the above mentioned results, three main limitations must be recognized in respect of the research work described here. First, the high aggregation used in the product classification (namely, the fact that all industry products are compacted under a single product group) may be partially responsible for some of the similarity found between different estimation techniques and the reference values. A more disaggregated analysis consists of one of the future developments of the present work, especially concerning manufactured products (using data available at the OECD trade database). Secondly, the simplifying hypotheses used in the MRIO table gathering may also have a consequence for the results, given that such a table is used as a point of reference to test the performance of different estimation techniques. Thus, another development that can improve our research involves the refinement of the MRIO system (namely substituting the constant import propensity with the values provided by import matrices for all countries). Thirdly, regions are obviously more open than countries; thus, it is expected that different trade estimates within an inter-regional model will have more impact on the model solution than the one calculated before. However, the magnitude of such an impact is much more difficult to confirm, given that it would require access to a complete multi-regional input–output system (with official statistics for inter-regional trade flows), to serve as a reference benchmark.

Acknowledgements

The authors gratefully acknowledge the referees and the editor for their valuable suggestions and comments.

References

Alward, G., D. Olson and S. Lindall (1998) Using a Double-constrained Gravity Model to Derive Regional Purchase Coefficients. Paper presented at the 45th Annual Meeting of the Regional Science Association International, Santa Fe, New Mexico. November.

Batten, D.F. (1982) The Inter-regional Linkages between National and Regional Input–Output Models. *International Regional Science Review*, 7, 53–68.

Batten, D. and D. Boyce (1986) Spatial Interaction, Transportation and Inter-regional Commodity Flow Models. In:: P. Nijkamp (ed.) *Handbook of Regional and Urban Economics*. Vol. 1. Amsterdam, Free University, 357–406.

Batten, D. and D. Martellato (1985) Classical Versus Modern Approaches to Inter-regional Input–Output Analysis. *Annals of Regional Science*, 19, 1–15.

Canning, P. and Z. Wang (2006) A Flexible Mathematical Programming Model to Estimate Inter-regional Input–Output Accounts. *Journal of Regional Science*, 45, 539–563.

Cesario, F. and T. Smith (1975) Directions for Future Research in Spatial Interaction Modeling. *Papers of the Regional Science Association*, 35, 57–72.

Eding, G., J. Oosterhaven, B. de Vet and H. Nijmeijer (1997) Constructing Regional Supply and Use Tables: Dutch Experiences. In: G.D.J. Hewings, M. Sonis, M. Madden and Y. Kimura (eds.) *Understanding and Interpreting Economic Structure*. Berlin, Springer Verlag, 237–263.

Ferreira, R. (2008) *Comércio Inter-regional Ibérico*. Ed. Associação de Desenvolvimento Regional do Instituto Politécnico de Portalegre, Portalegre.

Harrigan, J.F. (1990) The Reconciliation of Inconsistent Economic Data: The Information Gain. *Economic System Research*, 2, 17–25.

Hewings, G. and R.C. Jensen (1986) Regional, Inter-regional and Multiregional Analysis. In: P. Nijkamp (ed.) *Handbook of Regional and Urban Economics*. Vol. 1. Amsterdam, Free University, 295–355.

Hewings, G., S. Nazara and C. Dridi (2004) Channels of Synthesis Forty Years On: Integrated Analysis of Spatial Economic Systems. *Journal of Geographical Systems*, 6, 7–25.

Hulu, E. and G.J.D. Hewings (1993) The Development and Use of Inter-Regional Input–Output Models for Indonesia under Conditions of Limited Information. *Review of Regional Development Studies*, 5, 135–153.

Isard, W. (1998) Gravity and Spatial Interaction Models. In: W. Isard et al. (eds.) *Methods of Inter-regional and Regional Analysis.* Brookfield, VT, Ashgate Publishing, 243–280.

Isserman, A. (1980) Estimating Export Activity in a Regional Economy: A Theoretical and Empirical Analysis of Alternative Methods. *International Regional Science Review*, 5, 155–184.

Jackson, R. (1998) Regionalizing National Commodity-by-industry Accounts. *Economic Systems Research*, 10, 223–238.

Jackson, R. and A. Murray (2004) Alternate Input–Output Matrix Updating Formulations. *Economic Systems Research*, 16, 135–148.

Jensen-Butler, C. and B. Madsen (2003) Intraregional and Inter-Regional Trade in Regional Commodity Balances: Estimation and Results for Denmark. Paper presented at the 50th Annual North American Meeting of the Regional Science Association International, Philadelphia, 20–22 November 2003.

Kim, T.J., D.E. Boyce and G.J.D. Hewings (1983) Combined Input–Output and Commodity Flow Models for Inter-regional Development-Planning – Insights from a Korean Application. *Geographical Analysis*, 15, 330–342.

Lahr, M. (1998) A Strategy for Producing Hybrid Regional Input–Output Tables. Paper presented at the 12th International Conference on Input–Output Techniques, New York City, 21 May1998.

Lahr, M.L. and L. de Mesnard (2004) Biproportional Techniques in Input–Output Analysis: Table Updating and Structural Analysis. *Economic Systems Research*, 16, 115–134.

Liu, L. and P. Vilain (2004) Estimating Commodity Inflows to a Substate Region Using Input–Output Data: Commodity Flow Survey Accuracy Tests. *Journal of Transportation Statistics*, 7, 23–37.

Llano, C. (2004) The Inter-regional Trade in the Context of a Multirregional Input–output Model for Spain. *Estudios de Economía Aplicada*, 22, 1–34.

Llano, C., A. Esteban, J. Pérez and A. Pulido (2010) Opening the Inter-regional Trade Black Box: The C-Intereg Database for the Spanish Economy (1995–2005). *International Regional Science Review*, 33, 302–337.

McDougall, R. (1999) Entropy Theory and RAS are Friends. GTAP Working Papers 300, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Miller, R. and P. Blair (2009) *Input–Output Analysis – Foundations and Extensions,* 2nd edn. Cambridge, UK, Cambridge University Press.

Nijkamp, P., A. Reggiani and W. Tsang (2004) Comparative Modelling of Interregional Transport Flows: Applications to Multimodal European Freight Transport. *European Journal of Operational Research*, 155, 584–602.

OECD (2002) Bilateral Trade Database (edition 2002). OECD.

OECD (2005a) OECD Member Countries' Population 1981-2004 (Thousands and Indices: 2000=100). Labour Force Statistics, OECD, edition 2005.

OECD (2005b) Gross Domestic Product for OECD Countries (National Accounts of OECD Countries, Vol. 1, edition 2005).

Oosterhaven, J. and D. Stelder (2007) Evaluation of Non-survey International Input–Output Construction Methods with the Asian-Pacific Input–Output Table. Papers and Proceedings of the International Workshop Emergence of Chinese Economy and Re-organization of Asian Industrial Structure, 14–15 December 2006.

Piispala, J. (2000) On Regionalising Input–Output Tables – Experiences from Compiling Regional Supply and Use Tables in Finland. Paper presented at the XIII International Conference on Input–Output Techniques at University of Macerata, Italy, 21–25 August 2000.

Ramos, P. (2001) Comércio Intra e Inter-Regional no Continente Português: a Perspectiva das Estatísticas dos Transportes. *Cadernos Regionais, Região Centro*, 12.

Ramos, P. and A. Sargento (2003) Estimating Trade Flows between Portuguese Regions using an Input–Output Approach. Paper presented at the 43rd Congress of the European Regional Science Association, Jyväskylä, Finland, 27–30 August 2003.

Robinson, D. and Z. Liu (2006) The Effects of Interregional Trade Flow Estimation Procedures on Multiregional Social Accounting Matrix Multipliers. *The Journal of Regional Analysis and Policy*, 36, 94–114.

Roy, J.R. and J. Thill (2004) Spatial Interaction Modeling. *Papers in Regional Science*, 83, 339–362.

Sargento, A. (2007) Empirical Examination of the Gravity Model in Two Different Contexts: Estimation and Explanation. *Jahrbuch für Regionalwissenschaft*, 27, 103–127.

Sargento, A. (2009a) Regional Input–Output Tables and Models: Inter-Regional Trade Estimation and Input–Output Modelling based on Total Use Rectangular Tables. PhD dissertation, University of Coimbra, Portugal.

Sargento, A. (2009b) Introducing Input–output Analysis at the Regional Level: Basic Notions and Specific Issues. Regional Economics Application Laboratory (REAL), Discussion Paper n. 09-T-04.

Schwarm, W., R. Jackson and Y. Okuyama (2006) An Evaluation of Method for Constructing Commodity by Industry Flow Matrices. *The Journal of Regional Analysis and Policy*, 36, 84–93.

Sen, A. and T. Smith (1995) *Gravity Models of Spatial Interaction Behavior*. Berlin, Springer-Verlag.

Sheppard, E. (1978) Theoretical Underpinnings of the Gravity Hypothesis. *Geographical Analysis*, 10, 386–402.

Snickars, F. and J. Weibull (1977) A Minimum Information Principle. *Regional Science and Urban Economics*, 7, 137–168.

Tukey, J. (1977) *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Wilson, A.G. (1970) Inter-regional Commodity Flows: Entropy Maximizing Approaches. *Geographical Analysis*, 2, 255–282.