I nternational
I nput
O utput
M eeting on
M anaging the
E nvironment

Input - Output & Environment

SEVILLE (SPAIN)
July 9 - 11, 2008
http://www.upo.es/econ/IIOMME08

Hybrid Interregional Input-Output Construction Methods. Phase One: Estimation of RPCs

Escobedo, Fernando[a]*; Oosterhaven, Jan[b]

[a] *University of Castilla - La Mancha, School of Civil Engineering and Planning*
Avda. de Camilo José Cela, 2; 13071 Ciudad Real, Spain
Phone + 34 926 295 300. Fax + 34 926 295 391. E-mail: fernando.escobedo@uclm.es

[b] *University of Groningen, Faculty of Economics and Business,*
PO Box 800. 9700 AV Groningen, The Netherlands
Phone + 31 50 363 3728. Fax + 31 50 363 3720. E-mail: j.oosterhaven@rug.nl

*Corresponding author

Abstract

This paper searches for an optimal combination of non-survey methods when constructing a Spanish interregional input-output table for the region of Madrid and the five provinces of the region of Castilla-La Mancha (CLM), given the available data.

The Spanish case is particular in that some nine Spanish regions have survey type regional, domestic imports and foreign imports input-output tables with many sectors, mostly for around 2000, but not for CLM. Hence, we develop an optimal mixture of the full information intra-regional input-output table for Madrid, the limited information intra-regional input-output table for CLM and the input-output regression analysis for the intra-regional trade matrices for the provinces of CLM.

This regression analysis is based on statistical data on the transport of goods, especially as regards road transport, and uses a new method to add cell-specific information to the aggregate regressions. Moreover, we develop new methods to use with the richness of the Spanish survey data about regional purchase coefficients, based on input-output interpolation and extrapolation techniques.

Keywords: Interregional input-output analysis, Non-survey methods, Regression analysis, Commodity trade flows, Spanish regions.

1. Introduction

The debate about the pros and cons of different methods to construct regional and interregional input-output tables (IOTs) is predominantly phrased in terms of “either/or”. It evolves around such research questions as whether non-survey methods are acceptable or not and even whether using non-survey sector multipliers without IOTs are acceptable. At a more complex level questions arise about which type of coefficients, such as location quotients (LQs) or cross industry quotients, are best when estimating a non-survey IOT (Schaffer and Chu, 1969; Round, 1978), and whether using national coefficients and adapting them with RAS to the totals of the region at hand is acceptable or not (Hewings, 1969). Even more subtle issues are the choice of survey strategies: Is it better to ask firms for their sales/exports behaviour or for their purchase/import behaviour, and how useful is it to over-sample wholesale and transport sectors in order to get information on the export or the import coefficients of the products traded or shipped? (Boomsma and Oosterhaven, 1992). Also there is the issue whether the accuracy of non-survey methods should be evaluated at the level of the cells of the IO table (partitive accuracy) or better at the level of IO sector multipliers (holistic accuracy) (Jensen, 1980). Finally, specifically at the interregional level, at the edge between IO table construction and IO model building, there is the choice between interregional IO versus multiregional IO versus gravity IO models (Isard, 1951; Chenery, 1953; Moses, 1955; Leontief and Strout, 1963).

In many practical cases, however, the issue is not so much about which method is better or worse, but which method is more appropriate in which data situations or in which model applications, and in which it less. Thus, it may occur that both Beemiller (1990) and Bourque (1990) are right. Bourque in claiming that the RIMS non-survey IOTs based on LQs produces too large errors to be acceptable, and Beemiller in claiming that combining survey information for the exogenous impulse with a non-survey IOT produces acceptable impact estimates. The same conclusion was drawn from a tourism impact study in the Netherlands. Multipliers for the tourism-related sectors that were survey-based were very close to those of a comparable survey IOT, whereas the non-survey multipliers for the other sectors were quite off the mark, notwithstanding that they did not influence the impact variables much (Spijker, 1985). Consequently, West (1990) and Lahr (1993) conclude that the future is for hybrid IO tables and models. When constructing a series of intercountry IOT for the EU Van der Linden and Oosterhaven (1995) use a – given the available data – optimal combination of the interregional IO model, the multiregional IO model, and RAS to re-price the import matrices from ex customs prices to producers’ prices.

In this paper, we present the first construction phase of an interregional IOT for the Spanish regions of Madrid and Castilla-La Mancha (CLM), given the available data. This first phase has three main steps. In the first step we obtain the total output and the value added in the five provinces of CLM and the totals of the final demand. The second step has as its main goal to approximate the regional technical coefficients (RTCs). Finally in the third and last step of this first phase the regional purchase coefficients (RPCs) are estimated, and in this way we are able to calculate the intra-regional transactions matrix for the five CLM provinces.

For this purpose we have the national survey IO table and nine survey type regional,[1] domestic imports and foreign imports input-output tables with between 60 and 120 sectors, mostly for around 2000. Besides, we develop new methods to cope with the richness of survey data about RPCs, based on the IO interpolation and extrapolation techniques from Oosterhaven (2005).

2. Construction problem and solution strategy

The accounting structure of the ideal interregional input-output table (IOT) for a certain nation is given in Table 1. Each of the larger rectangular submatrices has the size I*(I+Q), where I = number of industries and Q = number of domestic final demand categories. The number of these interregional submatrices equals (R+1)*R, where R = number of regions and 1 = foreign imports. The other submatrices relate to remaining final demand **F** (foreign exports and changes in stocks) and gross value added at market prices **V** (product taxes minus subsidies, labour cost and operating surplus). Note that their overall totals equal total imports, exports and value added.[2] Thus, Table 1 depicts the sectoral disaggregation of both the regional and the national macro economic accounting identities, (Y = C + I + G + E – M).

Z^{11} Y^{11}	. . .	Z^{1R} Y^{1R}	F^1	x^1
:	:	:	:	:
Z^{R1} Y^{R1}	. . .	Z^{RR} Y^{RR}	F^R	x^R
M^1 $M_y^{\,1}$	. . .	M^R $M_y^{\,R}$	0	m^E
V^1 $V_y^{\,1}$	. . .	V^R $V_y^{\,R}$	0	v^E
$(x^1\ y^1)'$	. . .	$(x^R\ y^R)'$	$(f^E)'$	

Table 1. Layout of an ideal interregional input-output table.

[1] Spain is a country constituted by seventeen regions, that is, Andalucía, Aragón, Asturias, Baleares, Canarias, Cantabria, Castilla La Mancha, Castilla León, Cataluña, Extremadura, Galicia, Madrid, Murcia, Navarra, Rioja, Comunidad Valenciana and País Vasco, and two autonomous cities, Ceuta and Melilla.
[2] The zeros indicate that transit trade is assumed to be equal to zero.

Both for Spain (E) and for Madrid (M) there are detailed industry-by-industry IOTs with 73-74 sectors for 2000. In the case of Spain a full matrix with imports from the Rest of the World (RoW) is available, whereas for exports only columns for the Rest of the EU (RoEU) and the Remainder of the RoW (RRoW) are available. Additional to that information, the comparable IOT for Madrid also has a full import matrix for the Rest of Spain (RoS), separate import matrices for the RoEU and the RRoW, and an export column for the RoS. Four Spanish regional IOTs (Andalucía, Asturias, Baleares, and Galicia) basically have the same information as the Madrid IOT. Four other Spanish regions (Canarias, Catalunya, Aragón and País Vasco) have fewer import matrices, have the same export data, also have numbers of sectors above 60, and also have IOTs for 2000 or years close to that. Finally other three regions, Castilla León, Navarra and Comunidad Valenciana, although they have survey IOTs, do not have import matrices. The region of Castilla-La Mancha, however, only has a 30 sector non-survey IOT for 1995, which for these three characteristics is unusable for our present purpose. Other four regions, Cantabria, Extremadura, Murcia and Rioja, either they do not have IOT or they have non-survey IOT.

The only detailed information on the size of the sectors in de five Provinces of CLM and in the Region of CLM for 2000 relates to the employment[3] of 60 sectors[4], the two digit sectors NACE Classification[5]. Besides, at the level of 6 sectors for the provinces and 29 sectors for the regions, aggregate data about sectoral gross values added is available. Furthermore, detailed foreign export data are available for all provinces and detailed road transport data for all regions. Thus, hardly more information is available for the Region of CLM than for the Provinces of CLM. Consequently, we have chosen to construct a seven region IOT for Spain, with Madrid, the five CLM provinces and the RoS instead of a three region table with Madrid, CLM and the RoS. In this way, we are able to reach the NUTS 3 level[6] in the regional statistical classification of the European Union[7]. This seven-region IOT will be estimated such that it is consistent with the national IOT for Spain, which will thus function as the double-entry control total for the

[3] Another source of information is the Spanish census of houses and people, "Censo de población y viviendas", of 2001. There, we can find basic demographic data (sex, birth year, civil state, age, nationality and place of birth), place of residence, academic profile (sex, birth year, civil state, age, nationality and place of birth), employment data (professional position and situation, socioeconomic condition, hours worked, economic sector), and commuting data (daily trips, trip time, job place and transport used).

[4] These data come from the "Tesorería General de la Seguridad Social", social security general office of treasurer, an agency of the Spanish Ministry of Labour and Inmigration, which controls the employees affiliated to the national social security system. It has to be taken into account too the public servants affiliated to the specific security system "Muface", the soldiers affiliated to another specific security system "Isfas" and other professionals affiliated to private security systems.

[5] "Nomenclature generale des Activites economiques dans les Communautes Europeennes" (NACE) refers to the industrial classification as defined in Revision 1 which is used by Eurostat. It has 17 Sections (letters A to Q), 31 Subsections (2-character alphabetical codes), 60 Divisions (2-digit codes), 222 Groups (3-digit codes) and 503 Classes (4-digit codes).

[6] The five provinces of CLM are regions NUTS 3 as it is stated in the Regulation No 1059/2003 on the establishment of a common classification of territorial units for statistics (NUTS).

[7] The Nomenclature of Territorial Units for Statistics (NUTS) was established by Eurostat more than 30 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union.

interregional IOT (see Boomsma & Oosterhaven, 1992, for the optimal use of double-entry bi-regional book-keeping).

The location of the two regions and the constituent five Provinces of CLM are given in Figure 1. Their relative locations underscore the economic logic of combining the region of Madrid (M) with that of Castilla-La Mancha (C) into one IOT. The five CLM-provinces are: Albacete (A), Ciudad Real (D), Cuenca (U), Guadalajara (G) and Toledo (T). The region Rest of Spain (R) is Spain less Madrid and Castilla La Mancha, that is, Andalucía, Aragón, Asturias, Islas Baleares, Canarias, Cantabria, Castilla León, Cataluña, Extremadura, Galicia, La Rioja, Murcia, Navarra, País Vasco, Valencia and the autonomous cities of Ceuta and Melilla. Finally the region Rest of the World (RoW) is the World less Spain.

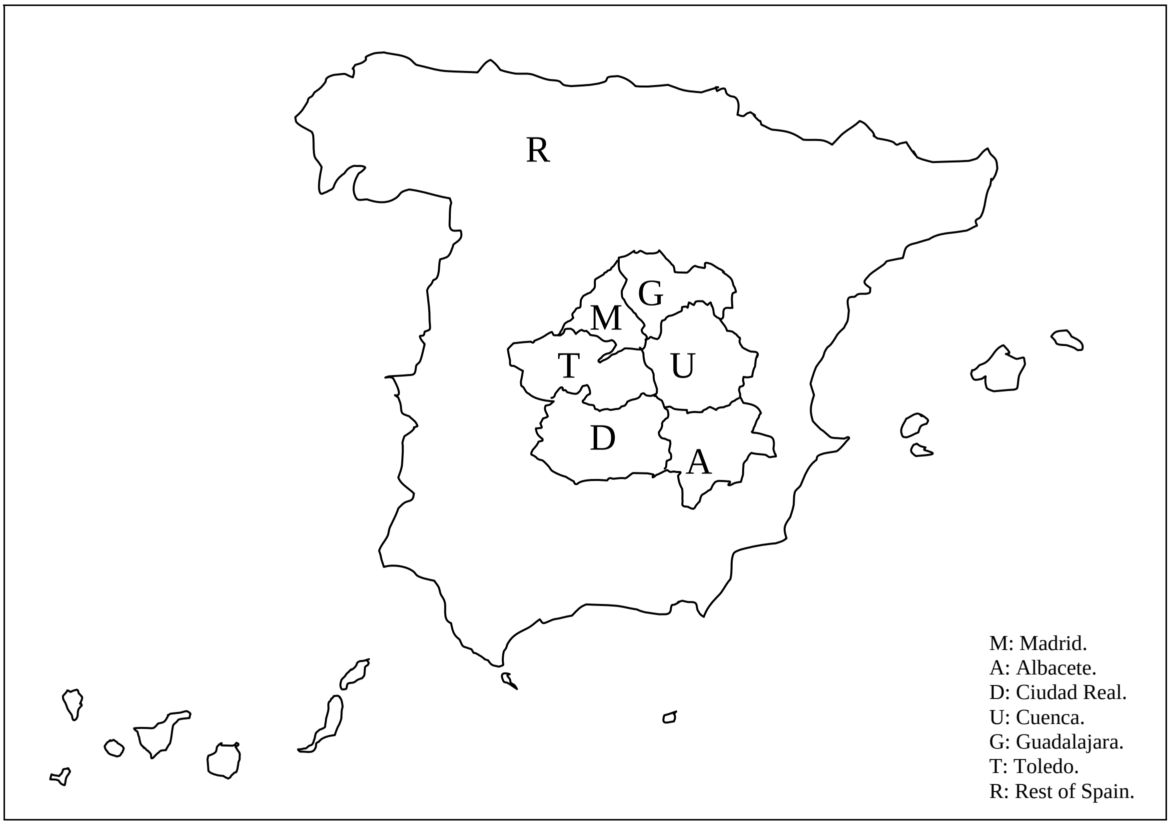

Figure 1. Seven regions of the Spanish interregional input-output table.

Given the geographical characteristics and the data availability, the core of the construction problem and the core of its solution are three-fold. First, provincial sectoral and final demand totals have to be estimated. Second, regional technical coefficients and regional final expenditure coefficients have to be estimated. Third, regional purchase coefficients have to be estimated. With the these three sets of estimates, the intra-regional intermediate and final purchases Z^{RR} and F^{RR} can be estimated.

2.1 Provincial employment, income and population shares

In this first step, total output, total use and the various components of value added for the 60 sectors with provincial employment data, and the totals of domestic provincial final demand by category, have to be estimated for the five CLM provinces.

Obviously, the point of departure are the 6-sector provincial value added totals and the 29-sector regional value added totals. The core of the solution of the further subdivision of these totals, to the 60-sector level of the provincial employment statistics, is to assume that the sectoral differences in labour productivity within CLM, are proportional to either those of Spain minus Madrid (E-M) or those of comparable Spanish regions. Total 60-sector output, use and value added for RoS can be calculated as a residual, and its implied labour productivity levels can be used to check plausibility. This gives $\mathbf{x}^{\mathbf{r}}$ and $\mathbf{V}^{\mathbf{r}}$ in Table 1. Previously we have to define the following parameters:

- n, number of sectors of the economic structure of Spain and its different regions.
- w_i^P, employment of sector i of region P.
- x_i^P, total output of sector i of region P.
- v_i^P, total value added of sector i of region P.
- $mclm$, number of provinces of the global region Madrid + Castilla La Mancha.

In this way, total outputs in the five provinces of Castilla-La Mancha will be approximated as a relation between the employment of the province and the employment of the global region Spain less Madrid:

$$\hat{x}_i^P = w_i^P \frac{x_i^{E-M}}{w_i^{E-M}}, \; P = A,C,U,G,T \; ; \; i = 1...n, \qquad \textbf{(1)}$$

And for the Rest of Spain:

$$x_i^R = x_i^E - \sum_{b=1}^{mclm} x_i^b \qquad \textbf{(2)}$$

we can use $x_i^R = w_i^R \dfrac{x_i^E}{w_i^E}$ to check plausibility.

On the other hand, total value added in the five provinces of Castilla-La Mancha will be:

$$\hat{v}_i^P = w_i^P \frac{v_i^{E-M}}{w_i^{E-M}}, \; P = A,C,U,G,T \; , \; i = 1...n \qquad \textbf{(3)}$$

Once we have those values, if we know for each province the value added for a number of sectors p and for each one of those sectors we know regional value added disaggregated into m sectors, we may apply RAS in the following way p times:

$$z^0 = \begin{pmatrix} \hat{v}^A_{p_1} & \hat{v}^A_{p_2} & \dots & \dots & \hat{v}^A_{p_m} \\ \hat{v}^D_{p_1} & \hat{v}^D_{p_2} & \dots & \dots & \hat{v}^D_{p_m} \\ \hat{v}^U_{p_1} & \hat{v}^U_{p_2} & \dots & \dots & \hat{v}^U_{p_m} \\ \hat{v}^G_{p_1} & \hat{v}^G_{p_2} & \dots & \dots & \hat{v}^G_{p_m} \\ \hat{v}^T_{p_2} & \hat{v}^T_{p_2} & \dots & \dots & \hat{v}^T_{p_m} \end{pmatrix};\ u = \begin{pmatrix} v^A_p \\ v^D_p \\ v^U_p \\ v^G_p \\ v^T_p \end{pmatrix};\ v = \begin{pmatrix} v^{CLM}_{p_1} & v^{CLM}_{p_2} & \dots & \dots & v^{CLM}_{p_m} \end{pmatrix} \quad \textbf{(4)}$$

For the Rest of Spain:

$$v_i^R = v_i^E - \sum_{b=1}^{mclm} v_i^b \quad \textbf{(5)}$$

we can use $v_i^R = w_i^R \dfrac{v_i^E}{w_i^E}$ to check plausibility.

Lacking total provincial final demand by category, obviously it has to be estimated by assuming it to be equal to the provincial population, income or industrial output shares in the corresponding national total. Again the RoS totals result as a residual and will be used to check plausibility. This gives $\mathbf{y^r}$ and $\mathbf{V_y}^{\mathbf{r}}$ in Table 1. Therefore, if we specify previously the final demand categories and parameters:

- $q = 1$, Final Households Consumption Expenditure.
- $q = 2$, Final Non-profit Institutions Serving Households Consumption Expenditure.
- $q = 3$, Final Government Consumption Expenditure.
- $q = 4$, Gross Fixed Capital Formation and Valuables.
- gw^P, government employment in region P.
- $mclmr$, number of regions in which Spain is divided for this study.
- r^P, household income of region P.

We can estimate the provincial totals of final demand in their different categories:

$$\sum_{a=1}^{mclmr} y_{iq}^{aP} =$$

$$y_{iq}^{E-M} \frac{r^P}{r^{E-M}},\ q = 1,2; \quad \textbf{(6)}$$

$$y_{iq}^{E-M} \frac{gw^P}{gw^{E-M}},\ q = 3; \quad \textbf{(7)}$$

$$y_{iq}^{E-M} \frac{x_i^P}{x_i^{E-M}},\ q = 4;\ P = A,C,U,G,T \quad \textbf{(8)}$$

With respect to the rest of Spain (R):

$$\sum_{a=1}^{mclmr} y_{iq}^{aR} = y_{iq}^{E} - \sum_{v=1}^{mclm} \sum_{a=1}^{mclmr} y_{iq}^{av} \text{ , } q = 1,2,3,4 \tag{9}$$

we can use $y_{iq}^{E} \frac{r^R}{r^E}$, $q = 1,2$; $y_{iq}^{E} \frac{gw_i^R}{gw_i^E}$, $q = 3$; and $y_{iq}^{E} \frac{x_i^R}{x_i^E}$, $q = 4$, to check plausibility.

With respect to the Changes in Inventories,

$$ys_i^P = ys_i^{E-M} \frac{x_i^P}{x_i^{E-M}} \text{ ,} \tag{10}$$

as we suppose that $ys_i^{PR} = 0$, $P \neq R$, that is, there is not interregional transactions in the changes in Inventories.

Subtracting the detailed foreign export data and the subdivided national changes in stocks from the total output by sector and region gives the total domestic sales per sector for all regions, i.e. $\mathbf{x}^{\mathbf{r}} - \mathbf{F}^{\mathbf{r}}\,\mathbf{i}$ in Table 1.

- z_{ij}^{RS}, economic transaction between the sector i of region R and sector j of region S.
- e_i^P, exports in sector i of region P.
- ys_i^P, changes in inventories in sector i of region P.

The exports $\hat{e}^R$ obtained from the Spanish Foreign Trade Data Base[8] for any region or province R may scale those from Spanish IO table:

$$e_i^P = \hat{e}_i^P \frac{e_i^E - e_i^M}{\hat{e}_i^E - \hat{e}_i^M} \tag{11}$$

And besides it has to be fulfilled that the sum of the exports in a specific economic sector of the seven Spanish regions has to be the same that the exports of the whole nation:

$$\sum_{a=1}^{mclmr} e_i^a = e_i^E \tag{12}$$

As it was said above, the total domestic sales per sector per region will be the total output less the imports and the changes in inventories:

$$\sum_{a=1}^{mclmr} \sum_{j=1}^{n} z_{ij}^{Pa} = x_i^P - e_i^P - ys_i^P \text{ ; } P = M, A, C, U, G, T, R \text{ ; } i = 1 \ldots n \tag{13}$$

[8] This Data Base from the Spanish Ministry of Commerce is very complete and exhaustive. It brings data about every region of Spain (17) or every province (50). The products are classified in the Standard International Trade Classification, Revision 4 and the data are in kilograms or euros. Besides the data are referred to monthly periods from 1995 and the destination may be any country in the world.

Thus, at the end of Step 1 all margin matrices and vectors of Figure 1 will be estimated, which leaves the estimation of the intra-regional, interregional and foreign imports sub-matrices with intermediate and final demand for the rest of the method.

2.2 Technical and expenditure coefficients per sector and category

In this step next to the given matrices with total intermediate requirements and total final requirements for Madrid (i.e. the sum of the domestic and foreign origin matrices), comparable technical requirement matrices have to be estimated for the five CLM provinces.

The solution here is straightforward, as it can only be based on the assumption that the CLM technical coefficients by sector and the CLM expenditure coefficients by final demand category are equal to the corresponding coefficients of comparable other Spanish regions or of E-M. After that, it has to be checked that the total requirements by purchasing sector and region plus value added 1st step, are equal for total output. The intermediate and final requirement matrices for the RoS will result as a residual and its implied technical and expenditure coefficients can be used to check plausibility.

We may define RTC_{ij}^{P}, the Regional Technical Coefficient of sector j of region P with respect to sector i, as the total purchases of sector j of region P to the sector i of Spain and the rest of the world. In the case of the region composed by Spain less Madrid and if m_{ij}^{P}, means the imports of sector j of region P from sector i of RoW, the Regional Technical Coefficient is:

$$RTC_{ij}^{E-M} = \frac{\sum_{a=1}^{mclmr} z_{ij}^{a,E-M} + m_{ij}^{E-M}}{x_j^{E-M} - v_j^{E-M}} \qquad \textbf{(14)}$$

If we define too r_{ij}^{P} as total requirements of sector j of region P from sector i, regardless its location, we will have that the total requirements in the five provinces of Castilla-La Mancha can be approximated as:

$$r_{ij}^{P} = \sum_{a=1}^{mclmr} z_{ij}^{aP} + m_{ij}^{P} = \left(x_j^{P} - v_j^{P}\right)\frac{\sum_{a=1}^{mclmr} z_{ij}^{a,E-M} + m_{ij}^{E-M}}{x_j^{E-M} - v_j^{E-M}}, \qquad \textbf{(15)}$$

$$P = A, C, U, G, T\ ;\ i, j = 1 \ldots n$$

Later we check that the total requirements are the total output less the value added:

$$\sum_{i=1}^{n} r_{ij}^{P} = x_j^{P} - v_j^{P} \qquad \textbf{(16)}$$

The total requirements in the Rest of Spain can be approximated as:

$$r_{ij}^{R} = \sum_{a=1}^{mclmr} z_{ij}^{aR} + m_{ij}^{R} = z_{ij}^{E} + m_{ij}^{E} - \sum_{b=1}^{mclm}\left(\sum_{a=1}^{mclmr} z_{ij}^{ab} + m_{ij}^{b}\right);\ i, j = 1...n, \quad \textbf{(17)}$$

and we can use $\sum_{a=1}^{mclmr} z_{ij}^{aR} + m_{ij}^{R} = \left(x_{j}^{R} - v_{j}^{R}\right)\frac{\sum_{a=1}^{mclmr} z_{ij}^{a,E-M} + m_{ij}^{E-M}}{x_{j}^{E-M} - v_{j}^{E-M}}$ to check plausibility and also it has to be fulfilled, as it was with a province P of CLM:

$$\sum_{i=1}^{n} r_{ij}^{R} = x_{j}^{R} - v_{j}^{R} \quad \textbf{(18)}$$

2.3 Estimation of regional purchase coefficients

In relation to the given intra-regional transactions matrix for Madrid, comparable matrices with intra-regional intermediate and final transactions have to be estimated for the five CLM provinces.

The most common way to estimate these matrices is by multiplying the total requirement matrices from step 2 with provincial self-sufficiency ratios, also labelled as regional purchase coefficients (RPCs, Stevens & Trainer, 1980). The Spanish IO data are unique in the sense that 60x64=3840 cell-specific RPCs can be calculated for at least nine Spanish regions. To keep the construction methodology tractable, we have chosen to explain the (row) aggregate RPCs for the 60 supplying sectors by means of regression analysis with sectoral fixed effects, and to use the average row pattern of the cell-specific RPCs to differentiate the aggregate RPCs by purchasing industry and purchasing category of final demand. To explain the aggregate sectoral RPCs from the nine regional IOTs, we will use the aggregate sectoral RPCs from the regional trucking survey, regional sectoral employment shares and geographical indicators, such as land surface share and the regional sectoral employment densities (cf. Oosterhaven, 2005). This will give a good estimate for the aggregate RPCs for the CLM region, and will give a maximum estimate for the aggregate RPCs for the five CLM provinces.

We may define $ARPC_{i}^{P}$, the Aggregate Regional Purchase Coefficient of sector i of region P, as the total purchases of region P to the sector i of region P with respect to the total purchases of region P to the sector i of the world. For example, for the region P we will have:

$$ARPC_{i}^{P} = \frac{\sum_{j=1}^{n} z_{ij}^{PP}}{\sum_{j=1}^{n} z_{ij}^{PP} + \sum_{\substack{a=1 \\ a\neq S}}^{mclmr} \sum_{j=1}^{n} z_{ij}^{aP} + \sum_{j=1}^{n} m_{ij}^{P}} \quad \textbf{(19)}$$

Therefore, as it has been said above, we undertake a regression analysis (RA) and use the trucking survey to check the past results (Wilson, 2000):

$$ARPC_i^{P-RA} = \alpha_i + \beta_i \frac{ts_i^{PP}}{ts_i^{PP} + \sum_{\substack{a=1 \\ a \neq P}}^{mclmr} ts_i^{aP} + ts_i^{FP}} + \gamma_1 \frac{w_i^P}{w^P} + \gamma_2 \frac{L^P}{L^E} + \gamma_3 \frac{w_i^P}{L^P}, \forall i \quad \textbf{(20)}$$

Being:

- α_i, fixed effects constants.
- β_i, fixed effects regression coefficients.
- $\gamma_1, \gamma_2, \gamma_3$, coefficients to be obtained in the regression analysis.
- ts_i^{RS}, trucking survey weight of good i transported from region R to region S.
- L^R, surface of the region R.
- i, type of commodity transported, $i = 1...s$

We are to use the IO data from the nine Spanish regions to calibrate the $2s+3$ unknowns $\alpha_1,...,\alpha_i,...\alpha_s, \beta_1,...,\beta_i,...,\beta_s, \gamma_1, \gamma_2, \gamma_3$ so our objective has to be that the difference between the aggregate regional purchase coefficient obtained from the IO tables, $ARPC_i^P$, and the ARPC obtained from the regression analysis, $ARPC_i^{P-RA}$, is as less as possible.

If we use the method of least squares we will have to minimize the sum of the square of that difference in all the Spanish regions and in all the types of commodities, that is,

$$\underset{\alpha_i, \beta_i, \gamma_1, \gamma_2, \gamma_3}{Minimize}$$

$$f = \sum_{\forall P} \sum_{\forall i} \left(\alpha_i + \beta_i \frac{ts_i^{PP}}{ts_i^{PP} + \sum_{\substack{a=1 \\ a \neq P}}^{mclmr} ts_i^{aP} + ts_i^{FP}} + \gamma_1 \frac{w_i^P}{w^P} + \gamma_2 \frac{L^P}{L^E} + \gamma_3 \frac{w_i^P}{L^P} - \frac{\sum_{j=1}^{n} z_{ij}^{SS}}{\sum_{j=1}^{n} z_{ij}^{SS} + \sum_{\substack{a=1 \\ a \neq S}}^{mclmr} \sum_{j=1}^{n} z_{ij}^{aS} + \sum_{j=1}^{n} m_{ij}^{S}} \right)^2$$

(21)

But we may write it in a more simplified way:

$$\underset{\alpha_j, \beta_j, \gamma_1, \gamma_2, \gamma_3}{Minimize} \quad f = \sum_i \sum_j \left(\alpha_j + \beta_j a_{ij} + \gamma_1 b_{ij} + \gamma_2 c_i + \gamma_3 d_{ij} - e_{ij} \right)^2 \quad \textbf{(22)}$$

If we want to obtain the minimum of a function we have to get the derivatives of the function with respect to the unknowns, that is,

$$\frac{\partial f}{\partial \alpha_j} = 2 \sum_i \left(\alpha_j + \beta_j a_{ij} + \gamma_1 b_{ij} + \gamma_2 c_i + \gamma_3 d_{ij} - e_{ij} \right) = 0, \qquad \forall j \quad \textbf{(23)}$$

$$\frac{\partial f}{\partial \beta_j} = 2 \sum_i a_{ij} \left(\alpha_j + \beta_j a_{ij} + \gamma_1 b_{ij} + \gamma_2 c_i + \gamma_3 d_{ij} - e_{ij} \right) = 0, \qquad \forall j \quad \textbf{(24)}$$

$$\frac{\partial f}{\partial \gamma_1} = 2\sum_i \sum_j b_{ij}\left(\alpha_j + \beta_j a_{ij} + \gamma_1 b_{ij} + \gamma_2 c_i + \gamma_3 d_{ij} - e_{ij}\right) = 0 \tag{25}$$

$$\frac{\partial f}{\partial \gamma_2} = 2\sum_i \sum_j c_i\left(\alpha_j + \beta_j a_{ij} + \gamma_1 b_{ij} + \gamma_2 c_i + \gamma_3 d_{ij} - e_{ij}\right) = 0 \tag{26}$$

$$\frac{\partial f}{\partial \gamma_3} = 2\sum_i \sum_j d_{ij}\left(\alpha_j + \beta_j a_{ij} + \gamma_1 b_{ij} + \gamma_2 c_i + \gamma_3 d_{ij} - e_{ij}\right) = 0 \tag{27}$$

If we adapt this system of equations to a matrix form we will have $Ax = b$, that can be easily solved.

$$\left(\begin{array}{ccccccccccccc}
1 & ... & 0 & ... & 0 & \sum_i a_{i1} & ... & 0 & ... & 0 & \sum_i b_{i1} & \sum_i c_i & \sum_i d_{i1} \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
0 & ... & 1 & ... & 0 & 0 & ... & \sum_i a_{ij} & ... & 0 & \sum_i b_{ij} & \sum_i c_i & \sum_i d_{ij} \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
0 & ... & 0 & ... & 1 & 0 & ... & 0 & ... & \sum_i a_{is} & \sum_i b_{is} & \sum_i c_i & \sum_i d_{is} \\
\sum_i a_{i1} & ... & 0 & ... & 0 & \sum_i a_{i1}^2 & ... & 0 & ... & 0 & \sum_i a_{i1}b_{i1} & \sum_i a_{i1}c_i & \sum_i a_{i1}d_{i1} \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
0 & ... & \sum_i a_{ij} & ... & 0 & 0 & ... & \sum_i a_{ij}^2 & ... & 0 & \sum_i a_{ij}b_{ij} & \sum_i a_{ij}c_i & \sum_i a_{ij}d_{ij} \\
... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\
0 & ... & 0 & ... & \sum_i a_{is} & 0 & ... & 0 & ... & \sum_i a_{is}^2 & \sum_i a_{is}b_{is} & \sum_i a_{is}c_i & \sum_i a_{is}d_{is} \\
\sum_i b_{i1} & ... & \sum_i b_{ij} & ... & \sum_i b_{is} & \sum_i a_{i1}b_{i1} & ... & \sum_i a_{ij}b_{ij} & ... & \sum_i a_{is}b_{is} & \sum_i\sum_j b_{ij}^2 & \sum_i\sum_j b_{ij}c_i & \sum_i\sum_j b_{ij}d_{ij} \\
\sum_i c_i & ... & \sum_i c_i & ... & \sum_i c_i & \sum_i a_{i1}c_i & ... & \sum_i a_{ij}c_i & ... & \sum_i a_{is}c_i & \sum_i\sum_j b_{ij}c_i & s\sum_i c_i^2 & \sum_i\sum_j c_i d_{ij} \\
\sum_i d_{i1} & ... & \sum_i d_{ij} & ... & \sum_i d_{is} & \sum_i a_{i1}d_{i1} & ... & \sum_i a_{ij}d_{ij} & ... & \sum_i a_{is}d_{is} & \sum_i\sum_j b_{ij}d_{ij} & \sum_i\sum_j c_i d_{ij} & \sum_i\sum_j d_{ij}^2
\end{array}\right)
\begin{pmatrix} \alpha_1 \\ ... \\ \alpha_j \\ ... \\ \alpha_s \\ \beta_1 \\ ... \\ \beta_j \\ ... \\ \beta_s \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}
=
\begin{pmatrix} \sum_i e_{i1} \\ ... \\ \sum_i e_{ij} \\ ... \\ \sum_i e_{is} \\ \sum_i a_{i1}e_{i1} \\ ... \\ \sum_i a_{ij}e_{ij} \\ ... \\ \sum_i a_{is}e_{is} \\ \sum_i\sum_j b_{ij}e_{ij} \\ \sum_i\sum_j c_i e_{ij} \\ \sum_i\sum_j d_{ij}e_{ij} \end{pmatrix}$$

(28)

Once we have developed the regression analysis, we define the RPC_{ij}^P, Regional Purchase Coefficient of sector j of region P with respect to sector i, as the total purchases of sector j of region P to the sector i of region P with respect to the total purchases of sector j of region P to the sector i of the world. For example, for the region P we will have:

$$RPC_{ij}^P = \frac{z_{ij}^{PP}}{r_{ij}^P} = \frac{z_{ij}^{PP}}{\sum_{a=1}^{mclmr} z_{ij}^{aP} + m_{ij}^P} \tag{29}$$

Then, we establish a relation between the regional purchase coefficient and the aggregate regional purchase coefficient, that is,

$$\frac{RPC_{ij}^{P}}{ARPC_{i}^{P}} = \frac{z_{ij}^{PP} \Big/ \left(\sum_{a=1}^{mclmr} z_{ij}^{aP} + m_{ij}^{P} \right)}{\sum_{j=1}^{n} z_{ij}^{PP} \Big/ \left(\sum_{j=1}^{n} z_{ij}^{PP} + \sum_{\substack{a=1 \\ a\neq S}}^{mclmr} \sum_{j=1}^{n} z_{ij}^{aP} + \sum_{j=1}^{n} m_{ij}^{P} \right)} \qquad \textbf{(30)}$$

and assume that the average of this coefficient in Spain is the same in the provinces we are studying, that is:

$$\frac{RPC_{ij}^{P}}{ARPC_{i}^{P-RA}} = \frac{\sum_{u=1}^{f} \frac{RPC_{ij}^{u}}{ARPC_{i}^{u}}}{f} \qquad \textbf{(31)}$$

being f the number of regions with input output data available that in our Spanish case $f = 9$. Therefore, the regional purchase coefficient in the province P will be:

$$RPC_{ij}^{P} = ARPC_{i}^{P-RA} \frac{\sum_{u=1}^{f} \frac{RPC_{ij}^{u}}{ARPC_{i}^{u}}}{f} \; ; \; P = M, A, C, U, G, T, R \; ; \; i = 1 \ldots n \qquad \textbf{(32)}$$

And finally we will be able to approximate the intra-regional IO table for the five provinces of CLM:

$$z_{ij}^{PP} = ARPC_{i}^{P-RA} \frac{\sum_{u=1}^{f} \frac{RPC_{ij}^{u}}{ARPC_{i}^{u}}}{f} \left(\sum_{a=1}^{mclmr} z_{ij}^{aP} + m_{ij}^{P} \right) \qquad \textbf{(33)}$$

3. Results and discussion

As this is a complex and very detailed model we are to concentrate in this paper in the key point of the first phase, the regression analysis of Equation 1. In this regression we need the following data: Intermediate demand of IO tables, statistical returns in respect of the carriage of goods by road, employment data and land areas data. The two last aspects are easily obtained from the Spanish Statistics National Institute.

As it has been said above, the intermediate demand of IO tables are taken from nine Spanish regional IO tables which have full intra-regional matrix and rest of Spain and out of Spain import matrices. The economic sectors in these tables are referred to the two digits NACE 93 Classification[9], the official classification of the national

[9] Nomenclature generale des Activites economiques dans les Communautes europeennes (NACE) refers to the industrial classification as defined in Revision 1 which is used by Eurostat. It has 17 Sections (letters A to Q), 31 Subsections (2-character alphabetical codes), 60 Divisions (2-digit codes), 222 Groups (3-digit codes) and 503 Classes (4-digit codes). It is regulated in the “Council Regulation (EEC) No

accounts in the countries of the European Union. With respect to the trucking survey it is undertaken every year in all the countries of the European Union[10]. In the case of Spain, the survey carried out in 2000 had 227,931 data where the most important variables are the type of goods, according to the groups referring to the NSTR/67 classification, the weight of the goods (gross weight in 100 kg), and the region or province of loading and unloading of the goods (country in the case of imports or exports).

Description	**Regression sectors**	**Two digit CPA**	**NSTR**	**Trucking survey codes**
Agriculture, hunting, forestry and fish	1	01,02,05	00,01,02,03,04,05,06,09	1,2,3,4,5
Food, beverages and tobacco	2	15,16	11,12,13,14,16,17,18	6,7
Coal, crude petroleum, natural gas, uranium and coke	3	10,11,12,23	21,22,23,31,32,33,34	8,9,10
Metal ores and basic metals	4	13,27	41,45,46,51,52,53,54,55,56	11,12,13
Other mining and non metallic mineral products	5	14,26	61,62,63,64,65,69,95	14,15,22
Paper and chemical products	6	21,24	71,72,81,82,83,84,89	16,17,18,19
Textiles, leather, wood, straw, rubber, precision and optical instruments and furniture	7	17,18,19,20,22,25,33,36	96,97	23
Fabricated metal products	8	28	94	21
Machinery, office machinery, computers, electrical machinery, communication equipment, motor vehicles, other transport equipment and secondary raw materials	9	29,30,31,32,34,35,37	91,92,93,99	20,24
Agriculture, fish, mining, and manufactured products	10	01-37	00-99	1-24

Table 2. Coordination of CPA-96 and NSTR/67 classifications.

The first and main difficulty we have to overcome is the different product classifications between the input-output tables and the trucking survey as the IO tables

3037/90 of 9 October 1990 on the statistical classification of economic activities in the European Community", whose purpose is to establish a common statistical classification of economic activities within the European Community in order to ensure comparability between national and Community classifications and hence national and Community statistics.

[10] This is the subject of the "Council Regulation (EC) No 1172/98 of 25 May 1998 on statistical returns in respect of the carriage of goods by road". This regulation states that Member States shall compile statistical data relating to the following areas: a) vehicle-related data; b) journey-related data and c) goods-related data.

are referred to the official EU classification, the CPA 96[11], and the trucking survey is referred to the NSRT/67. The two digits CPA 96 classification has 60 economic sectors and the NSRT/67 classification has 52 two digits sectors. Nevertheless these 52 sectors have been grouped into 24 sectors in the trucking survey. When we face both classifications, CPA and NSTR, we realize that all the NSRT sectors only correspond to the first half of the CPA sectors, as the second half of the CPA sectors are related to services, not to products. One second important aspect is that both classifications are rather intertwined, so finally the option chosen was to group CPA sectors which were more or less identified with NSTR sectors, which were also linked to trucking survey sectors. As it can be seen in Table 2, the regression analysis has finally 10 sectors, nine of these correspond to the links between groups of sectors of the different classifications and the tenth to the relation between the totals of the classifications.

After this step, which is the most important one, and with the help of Matlab, we undertake the regression analysis and we obtain first the parameters of Equation 22 $a_{ij}, b_{ij}, c_i, d_{ij}, e_{ij}$. Then we solve the system of equations that has its matrix form in Equation 28 and we realize that the main matrix is a symmetric positive definite one, so the Cholesky decomposition may be applied. As we can see in the results shown in Table 3, the fixed effects constants are less than the fixed effects regression coefficients and with respect to the three last coefficients the one referred to the relation between the region and country areas shows its importance with a value of 1,62.

Parámetro	**Valor**	**Parámetro**	**Valor**	**Parámetro**	**Valor**
α_1	-0,057817	α_9	-0,008833	β_7	0,419393
α_2	-0,041797	α_{10}	0,041992	β_8	0,621985
α_3	-0,002625	β_1	0,414467	β_9	0,149616
α_4	0,006420	β_2	0,536967	β_{10}	-0,573253
α_5	-0,001603	β_3	0,155968	γ_1	0,589507
α_6	-0,009888	β_4	0,199953	γ_2	1,620864
α_7	-0,079881	β_5	0,481812	γ_3	0,005004
α_8	-0,089372	β_6	0,128193		

Table 3. Results of the regression analysis.

[11] Established in the Council Regulation (EEC) No 3696/93 of 29 October 1993 on the statistical classification of products by activity (CPA) in the European Economic Community. The statistical Classification of Economic Activities (NACE) and the statistical Classification of Products by Activity (CPA) in the European Community are part of the integrated system of statistical classifications. They have the same structure, their difference is that NACE is for industries and CPA is for products.

4. Conclusions

In this paper an interregional input-output model for Spain is presented and it is mainly based on the statistical returns in respect of the carriage of goods by road that is undertaken every year in the different countries of the European Union. These statistical returns provide us with abundant data (more than 200,000 surveys in Spain in 2000) in the economic sectors that produce goods like agriculture, fish, mining, and manufacturing but do not contribute with data in the services sector, so we have to assume that the economic transactions in this sector are proportional to the totals in the sectors interchange goods.

The key point of the model with respect to obtaining the intra-regional input-output matrices is a regression analysis between input-output tables of Spanish regions which have detailed regional, domestic imports and foreign imports and the statistical returns in respect of the carriage of goods by road, besides other parameters like employment and land areas. This regression analysis can be solved in an analytical way as its solution is constituted by a system of equations whose matrix is a symmetric positive definite one, so the Cholesky decomposition may be applied.

The available data to do the interregional model allow us to reach the level of NUTS 3 statistical regions of Eurostat. This is important as the vast majority of the input output tables carried out in Europe corresponds to the national and regional level (NUTS 1 and NUTS 2) and therefore with this provincial level IO table we are able to analyze economic relations like the interregional employment spillovers and feedbacks between the core region of Madrid and the less developed five provinces of Castilla-La Mancha.

Acknowledgements

We would like to deeply thank Roberto Mínguez who provided us with helpful assistance in the mathematical aspects of the paper. On the other hand Yves Mahieu and other colleagues from Eurostat provide us with useful correspondence tables between different economic classifications used in this paper.

References

Beemiller, R. M. (1990) Improving accuracy by combining primary data with RIMS: Comment on Bourque, **International Regional Science Review**, 13, 1-2, pp. 99-101.

Boomsma, P. & Oosterhaven J. (1992) A double-entry method for the construction of bi-regional input-output tables, **Journal of Regional Science,** 32, 3, pp. 269-84.

Bourque, P.J. (1990) Regional multipliers: WAIO vs. RIMS, **International Regional Science Review,** 13, 1-2, pp. 87-98.

Chenery, H.B. (1953) Regional analysis, in: H.B. Chenery, P.G. Clark & V.C. Vera (eds.) **The structure and growth of the Italian economy** (Rome, U.S. Mutual Security Agency).

Hewings, G. J. D. (1969) Regional input-output models using national data: The structure of the West Midlands economy, **The Annals of Regional Science,** 3, 1, pp. 179–191.

Isard, W. (1951) Interregional and regional input-output analysis, a model of the space economy, **Review of Economics and Statistics,** 33, pp. 318-28.

Jensen, R.C. (1980) The concept of accuracy in regional input-output models, **International Regional Science Review,** 5, 2, pp. 139-54.

Lahr, M. (1993) A review of the literature supporting the hybrid approach to constructing regional input-output models, **Economic Systems Research,** 5, 3, pp. 277-93.

Leontief, W. W. & Strout A. (1963) Multiregional Input-Output Analysis, in T. Barna (ed.) **Structural Interdependence and Economic Development** (New York, St. Martin's Press).

Linden, J. A. van der & Oosterhaven J. (1995) European Community Intercountry Input-Output Relations: Construction Method and Main Results for 1965-1985, **Economic Systems Research,** 7, 3, pp. 249-69.

Moses, L. N. (1955) The stability of interregional trading pattern and input-output analysis, **American Economic Review,** 45, 5, pp. 803-32.

Oosterhaven, J. (2005) Spatial Interpolation and Disaggregation of Multipliers, **Geographical Analysis,** 37, 1, pp. 69-84.

Round, J. I. (1978) An interregional input-output approach to the evaluation of non-survey method, **Journal of Regional Science,** 18, 2, pp. 179-94.

Schaffer, W. A. & Chu K. (1969) Non-survey techniques for constructing regional interindustry models, **Papers of the Regional Science Association,** 23, pp. 83-101.

Spijker, J. (1985) Een combinatie van directe en indirecte methoden, in: J. Oosterhaven en B.B.A. Drewes (eds.), **Constructie en actualisering van regionale en interregionale input-output tabellen**, (Arnhem, Economisch-Technologische Instituten).

Stevens, B. H. & Trainer G. A. (1980) Error generation in regional input-output analysis and its implications for non-survey models, in S. Pleeter (ed.) **Economic impact analysis: methodology and applications** (Boston, Martinus Nijhoff).

West, G. R. (1990) Regional trade estimation: A hybrid approach, **International Regional Science Review,** 13, 1-2, pp. 103-18.

Wilson, A. G. (2000) **Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis** (Devon, England, Northcote House Pub Ltd).