Revisiting Methods for Estimating Interregional Input-Output Accounts: It's Not Just About Trade Flows

June 19, 2023

Ana Lúcia Marto Sargento, CARME, School of Technology and Management, Polytechnic Institute of Leiria, Leiria, PT

Michael L. Lahr, Bloustein School of Planning & Public Policy, Rutgers University, New Brunswick, NJ, USA

João Pedro Ferreira, Weldon Cooper Center for Public Service, University ofVirginia, Charlottesville, VA, USA

Fernando de la Torre Cuevas, Department of Applied Economics, University of Santiago de Compostela, Santiago de Compostela, ES

ABSTRACT. Interregional input-output (IO) tables largely differ in the quantity and nature of the set of available information pertaining to interregional trade. With respect to ability to replicate interregional trade accurately, research to date suggests decreasing returns to scale persist regarding both more theoretical expectations and added empirical data. A basic underlying assumption is that intermediate industry accounts of the economies in interregional IO tables exist and are accurate. In fact, if they exist at the subnational level, such accounts are, at best, roughly estimated and predicated on far lessempirical information than is available for economies of nations. Moreover, intra-economy intermediate-industry flows are typically markedly larger than the set of a region's commodity in- and out-flows. So, if intermediate industry flows in a set of interregional IO accounts are noticeably mis-estimated, it follows that interregional trade coincidentally derived using them must be even more conspicuously in error.

In this piece, we identify a few approaches typically used by researchers worldwide to develop subnational interregional IO models. We start by consolidating all accounts of the 27 member states for the year 2019, while maintaining sectoral detail, to produce a "national account". We then apply (1) a unified approach that uses only regional populations, national direct requirements coefficients, geographic distances, jobs by regional industry to distribute all transactions. In the others, we assume that value added shares of output for each regional industry are known and apply additional information for the year 2019 for each region's industries. In approach (2), we also "know" each regional industry's supply/demand ratio to produce intraregional transactions. In (3), we estimate Flegg and Tohmo's (2016) FLQ for each industry in each region to produce intraregional intermediate transactions. And in (4) we estimate econometric RPCs of the sort discussed in Lahr, Ferreira, and Többen (2020). Using this information, we construct interregional trade flows using a gravity model and RAS in all cases. We then test to see how well the approaches estimate interregional IO accounts of member states in the European Union (EU) as presented by the FIGARO database. We then apply the eight approaches—(1a), (1b), (2a), (2b), (3a), (3b),(4a) and (4b), where (a) denotes the full set of trade flows, (b) a diagonal set of trade flows only. We compare each to actual interregional accounts of EU member states published in FIGARO and, in turn, examine the benefits and trade-offs inherent to them.

1. Introduction

Not too long after Leontief (1936, 1941) published the first input-output (IO) table thoughts ran toward generating such a table for subnational geographic units. Walter Isard (1951) was the first to formally propose one. He recognized the less-than-desirable nature of political regions for economic analysis, the lack of trade data at the subnational level, and the huge hurdle to applied interregional analyses that would arise from (then) computational limits. The story is told that Wassily Leontief was at least mildly appalled by Isard's (1951) lack of empirics.

Isard (1951) does not develop an interregional IO table since his proposal is hopelessly data intensive. It not only requires information on the origins and destinations of trade flows by both region and industry for all intermediate industry production but also for all types of final demands. It is no surprise, then, that subnational interregional IO (IRIO) tables of this genre have seldom been attempted. Japanese (1960-2005) and South Korean (2005) survey-based IRIO tables are notable exceptions (Miller & Blair, 2022).[1] A prime difficulty is collecting interregional commodity trade flows, a relatively easy task at an international level.[2]

The trade data are critical at the subnational level since subnational trade across regions can be quite volatile, as producers and consumers alike seek better and more diverse supplies. Moreover, use of Leontief models assumes their coefficients are relatively stable. In Isard's (1951) conceptual treatment, the **A** matrices (the direct requirements matrices) for each region are likely nearly as stable as those for the nation (D. Batten & Martellato, 1985). But each is composed of intra-regional direct-

---

[1] Oddly enough, above the national level, the set of international trade flows are likely of better quality than intranational interindustry shipments, which are roughly estimated (see, e.g., Planting & Guo, 2004; Dalgaard & Gysting, 2004). This is despite the well-known inconsistency of import and export flows between country trade pairs (Timmer et al., 2015)—the so-called "mirror puzzle of trade"; while irritatingto analysts the "inconsistency" is typically relatively small in broader perspective, often simply relating tothe accounting (or not) of logistical carrying costs. Moreover, Jackson, Israilevich, and Comer (1992) note that even when access to regional technology data is granted, estimating intraregional IO tables for all regions of a nation in a manner consistent with a published national table would undoubtedly create so many complex problems to render the problem intractable. At the national level, these authors reveal that issues arise simply matching changes national IO technology as revealed through census data to changes in IO technology available in balanced national accounts.

[2] At best nations undertake seasonal surveys at selected times of day, which yields biased results on freight flows (Lau, 1995). The survey action itself also can bias the responses of freight handlers. For example, it is well known in the U.S. that truckers communicate with each other when and where weigh stations are stopping vehicles to undertake such surveys, so they can avoid delays and fines (for carrying too much weight in their vehicle or for carrying goods they ought not be carrying).

requirementscoefficients a_{ij}^{s}, final-demand deliveries f_{i}^{rs}, and trade coefficients c_{ij}^{rs}, where r and s,

respectively, denote the origin and destination regions and i and j, respectively, indicate the origin and destination industries and sectoral output is denoted.

$$x_{i}^{r} = \sum\nolimits_{s=1}^{m} \left\{ \left[\sum_{j=1}^{n} (a_{ij}^{s} c_{ij}^{rs} x_{j}^{s}) \right] + \sum_{k=1}^{l} f_{ik}^{rs} \right\} \quad (1)$$

for all r and i. The temporal stability of trade coefficients for a given *r-s* pair for any given *i-j* flow is less tenable.

Batten and Martellato (1985, p. 4) classify interregional IO models "[i]n order of increasing generality and difficulty to implement… (1) Leontief, (2) Leontief and Strout, (3) Chenery-Moses, (4) Riefler-Tiebout, and (5) Isard." Table 1 briefly reviews these classical approaches. Subsequent paragraphs discuss data requirements and the set of hypotheses in relation to trade coefficients with respect to recent best practice in subnational multiregional IO (MRIO) table estimation techniques.[3]

The implementation of global MRIO models assumes the availability and verity of underlying national IO tables. Official subnational equivalents tend not to be readily available.[4] In fact, regional IO tables are seldom published. When they are, they typically rely more heavily on non-survey data than on survey data[5] or contain only estimated blocks predicated upon primary data. There has been literature on hybrid methods for producing regional IO tables, and much of centers on finding the right balance between insertion of direct information and nonsurvey estimates (e.g., Lahr, 2001). We do not intend to replicate here an exhaustive literature review of these methods, which can be found, for instance, in Miller and Blair (2022). Rather we focus on more recent contributions that have been tested for their accuracy, particularly relating to the two main issues involved in MRIO table construction: 1) estimation of intraregional flows and 2) estimation of interregional trade flows.

[3] For a complete description of these models, please refer to Miller and Blair (2022).

[4] For a matter of clarification, we herein refer to MRIO systems that deal with subnational units. In the recent years, some integrated systems of national IO tables – such as WIOD or EORA - have been built using the same multiregional IO modelling principles. These are sometimes called MRIO models, although they incorporate and link national IO tables.

[5] e.g., the series of tables produced for the State of Washington.

Table 1. Classical approaches in interregional IO models – summary table.

Interregional IO modelling approach	Data requirements	Hypotheses regarding trade coefficients (t_{ij}^{rs})	Main limitation
(1) Leontief (1953)	• Distinction between “regional” and “national” commodities” • Regional outputs and regional final demands for both types of commodities • Proxies (only) of market share of “national” production yield estimates of regional production	Market share of each region in providing each “national” product is a constant proportion of the national output of the same product	It relies on the use of ***net*** trade flows across regions, which underestimates interregional feedback effects
(2) Leontief and Strout (1963)	• Regional supply (supply pool) for each product • Regional demand (demand pool) for each product • Some measure of impedance (for instance, related to distance) between region origin-destination pairs	Import proportionality assumption: the same trade coefficient applies to all uses (both intermediate and final) in the destination region.	Prior information on interregional trade flows is required to estimate the gravity model parameters. Import proportionality assumption: in fact, different industries and final users have different import propensities
(3) Chenery (1953)**-Moses** (1955)	• Origin-destination matrix depicts intra- and interregional shipments of each commodity’s output • Interindustry flow matrix for each region, in which the generic element considers all geographic origins of input *i*, except for international imports • Each commodity in final demand is specified by each source region • Regional output for each commodity	Import proportionality assumption: the same trade coefficient applies to all uses (both intermediate and final) in the destination region.	Import proportionality assumption: in fact, different industries and final users have different import propensities
(4) Riefler and Tiebout (1970)	• Import matrix and export matrix needed for each region • Interregional trade statistics by commodity • Each commodity in final demand is specified by each source region • Regional output for each commodity	Trade coefficients’ stability, i.e., when a shift in final demand occurs, trade patterns are unaltered	Demands data that are rarely available, e.g., import and export matrices for each region.
(5) Isard (1951)	• Regional production of each product • Final demand of each product satisfied by regional production • Intermediate consumption flows, distinguishing by industry of origin and region of origin (the same or each of the others) • International imports used as intermediate consumption	Trade coefficients’ stability, i.e., when a shift in final demand occurs, trade patterns are unaltered	Very high demand of seldom available data

Source: Authors’ elaboration.

2. Research Approach

We follow Sargento's (2009) three dimensions of subnational IO models which classifiesthem as i) the number of regions considered: single- or many-region models, i.e., the recognition (or not) of interregional linkages; ii) the way in which regional directrequirements matrices are estimated, and iii) the extent to which interregional trade flowsare detailed. A good deal of recent work has focused on (ii) above in the single region case—the estimation of single regional direct requirement matrices. So, we limit the focusof the present paper to the coalescence of all three dimensions in the many-region case only; we, thus, exclude all models that cannot account for interregional spillover and feedback effects.

2.1. Intraregional shipments

On the matter of producing subnational IO tables, Miller and Blair (2022, p. 65) note that as early as Isard and Kuenne (1953) and Miller (1957) used "national technical coefficients… in conjunction with an adjustment procedure… to capture some of the characteristics of the regional economies". The adjustments have been called both "regional supply percentages" and "regional purchase coefficients" (RPCs). They were defined as the shares of local demands that are fulfilled by local supplies—and naturally were applied row wise to the direct requirements matrix **A**.[6]

Garhart and Giarratani (1987) were probably the first to formally declare that regionalists could do better, at least from a conceptual perspective if not a pragmatic one.That is, they note that the RPCs were each row's average regional supply share; the regional supply share for each element in the row undoubtedly varies from that average. Thus, in reality a full matrix of RPCs is the ideal, and the diagonalized matrix of the vector of RPCs was imprecise, albeit approximately correct. Strictly RAS-based studies of regionalization in which a region's industry totals for intermediate inputs and intermediate outputs are perfectly known confirm Garhart and Giarratani's (1987) assertions (see, e.g., Malizia & Bond, 1974; McMenamin & Haring, 1974). Unfortunately, we are aware of no real-world instances in which such totals are available.

[6] Note, the difference between a technical or technology matrix and a direct requirements matrix is the treatment of imports, particularly noncompetitive international imports, which are not separated out distinctly in a "true" technology matrix. In a regional setting the implication is that such imports are usedby all regions in the same proportion and allocated with output. This is, of course, a rather strong assumption, particularly for large countries with rather large economies, like Russia, China, Canada, United States, Brazil, India, Mexico, Indonesia, Australia, and Japan. In such cases, more remote and interior regions are less apt to import goods from abroad than are core and coastal regions.

As a result, we test three basic ways to estimate the intra-regional direct requirement matrices. We bypass some of the most sophisticated approaches in this piece. We purposely opt not to employ mathematical programming (Flegg et al., 2021) or artificial intelligence techniques (Pakizeh & Kashani, 2022), at least for now. The idea is not to use approaches methodologically more advanced than those typically used by statistical agencies (see, e.g., Dalgaard & Gysting, 2004; Planting & Guo, 2004; Valderas-Jaramillo et al., 2019, 2021). Still, all three basic approaches assume that information exchange is free and perfect, i.e., that national industry technology is the same everywhere in a nation.

2.1.1. The integrated approach

The first of the three (1) is the simplest. It is what Boero, Edwards, and Rivera (2018) call "an integrated approach" and derives directly from Leontief (1953) except for the way industry outputs and demands are allocated across regions.[7] In this instance, we start by assuming that minimal data are available for regions. So, the approach assigns national output to regions via each region's shares of national labor use by industry,[8] assumes total productivity is the same nationwide (value added's share of output is spatially constant) and each region's share of final uses by sector is estimated well by the region's share of the nation's population (c.f., Treyz & Stevens, 1985). Knowing output for each sector in each region, we then estimate intermediate output by each sector in each region by post-multiplying $\mathbf{A}$ by the diagonalized vector of regional output $\hat{\mathbf{x}}^s$ and summing across of the resulting matrix, i.e., $\mathbf{A}\hat{\mathbf{x}}\mathbf{i}$ where $\mathbf{i}$ is a summation vector (a vector of 1s of appropriatelength, in this case n). Thus, we obtain total output and total demand by industry for eachregion, which (more likely than not) is unbalanced. Moreover, we are assured that the sum across all regions of both output and demand sum to the nation's totals. In this case,note, however, that we have not *yet* identified the set of intra-regional trade flows; This is done later, simultaneously with estimating interregional trade flows.

2.1.2. Other regionalization approaches

Again, many approaches have been applied. One thing that recent publications have doneis winnow out the chaff; that is, recent tests have tended to show some of the many approaches

[7] Over time, a long list of MRIO table builders have used this general approach, including Sargento, Ramos, and Hewings (2012), Haddad (2014) and Elshahawany, Haddad, and Lahr (2017). In many cases, however, authors make rather strong assumptions about intra-regional intermediate transactions as a region's share of all transactions for a particular industry.

[8] Note, rather shares of jobs, that compensation or value-added shares ought to produce better allocations of national output across regions since they should better account for productivity differences.

are clearly inferior. Thus, we ignore those known inferior approaches except for a supply/demand ratio truncated at values below 1.0 à la Haddad (2014). The two approaches we selected are parametric; and as such, at least at present, use information on intraregional trade by industry to set the parameters. Because of this, we treat the remaining two alternatives equally and as if much more information is "known". Both also account for possible commodity cross-hauling. In all cases, we assume information is available for at least three components of value added (labor compensation, indirect business taxes, and other value added) by industry as well as for all final demand by industry. The reason we do so is because most of such information is available for stateswithin the US as well as for NUTS2 regions in the EU.[9]

2.1.2.1. The supply/demand ratio

The first and simplest of the three alternatives is one used by Haddad (2014). He lays out the groundwork for use of a supply/demand ratio (S/D) that is censored so that its maximum value is 1.0 when producing all regional direct requirements matrices for an MRIO model of Lebanon (Haddad, 2014, Equation 2). He modifies the S/D further by using what we can only assume is an expert-based "fudge factor," a parameter F(c) that articulates "the extent of tradability of a given commodity" (p. 17).[10] Including the F(c) Haddad then applies the resulting vector rowwise to national flows to obtain estimates of intraregional shipments by industry for each region.

2.1.2.2. FLQs

Flegg and Tohomo (2019) test a family of methods called FLQs. The most recent and accurate of them, according to the tests, consists of parametric transformations of cross-industry location quotients (CILQs) and accounts for the relative economic size of a region. They are further modified by a parameter δ that adjusts for the degree of inherent cross-hauling. Empirical work

---

[9] We understand that not all final demand data are, in fact available, by NUTS2 region or U.S. state.

[10] For basic goods, Haddad (2014) set F(c) to 0.9, while for export-base goods he set it to 0.5. He assumed no interregional trade at all for public administration. This basically reflects Leontief's (1953) suggestion to split on base/nonbase production. Note, however, that both Stevens, Treyz, and Lahr (1989) and Jahn, Flegg and Tohmo (2020) suggest a value closer to 0.3 is probably more appropriate than 0.5 for modifying location quotients or S/D.

In their FORTRAN program, Treyz and Stevens (1985) used 0.95 rather than 0.9 as a top value for basic goods, although they also applied an algorithm using the untruncated LQ or S/D value to allow the top value to asymptotically approach 1.0. They also set the ratio for the lodging industry to 0.5, figuring it tended to be an export-based industry because, except for overly interesting cases, locals tend not to use local-area hotels.

We further note that Fournier Gabela (2020, fn 15) applies a strictly algorithmic approach to estimate trade for "not easily tradeable" commodities, although he provides no economic intuition to support it. Because of this missing element, we used Haddad's brute-force approach instead. That is, it could be that Fournier Gabela's algorithm yields "better" results.

is required to identify the proper value of δ. Jahn, Flegg, and Tohmo (2020) note that a series of empirical tests carried out in different geographies using FLQs suggest that $\delta = 0.3 \pm 0.1$. Thus, Flegg et al. (2021, p. 2) note that "the choice of a suitable value of δ … has limited the practical use of the FLQ." Its lack of economic content is another.

Following Flegg et al. (2016), Flegg et al. (2021, eq. 48) estimate any given elementof the regional direct requirement matrix (their latest FLQ):

$$a_{ij}^r = a^n \mu_{ij}^r \lambda_i \qquad (2)$$

where $\mu_{ij}^r = \begin{cases} CILQ_{ij}^r & \text{if} \quad i \neq j \\ SLQ_i^r & \text{if} \quad i = j \end{cases}$ and $\lambda_i = \log_2 \left(1 + \frac{x_i^r}{x_i^n}\right)^{\delta_i}$. The parameter δ_i is estimated via a transformed version of (2) (Flegg et al., 2021, eq. 49):

$$\ln\left(\frac{a_{ij}^r}{a_{ij}^n \mu_{ij}^r}\right) = \delta_i \lambda_i + \varepsilon_{ij} \qquad (3)$$

This approach transforms national technology coefficients to regional direct coefficients by applying the region's cross-industry location quotient (CILQ) to off-diagonal elements of the national direct requirements matrix or by applying the industry supply location quotient (SLQ) to its diagonal elements and, then, multiplying the results by the log base 2 of the sum of unity added and the industry's national output share contained within the region. Flegg et al. (2016, 2021) report that λ adjusts for region size in terms of output, but do not reveal rationales for taking the log base 2 of the industry output share, why unity is added to it, or why that sum must be modified by δ. Presumably, these various assertions simply assure the CILQs and SLQs fit better. Note that the use of CILQs enables a full matrix of RPCs as extolled by Garhart and Giarratani (1987).

2.1.2.3. Econometric RPCs

The second is a location theory-based econometric approach originated by Treyz and Stevens (1985) for U.S. states, revisited by Stevens, Treyz, and Lahr (1989), and formalized by Lahr, Ferreira, and Többen (LFT, 2020), who formulated it for the European Union. Their quasi-binomial estimates of RPCs are rows-only adjustments to **A** that estimate intraregional trade by industry. That is, LFT estimate RPCs (regional purchase coefficients or regional supply percentages), the regions' propensities to use local production. Leaning on Treyz and Stevens (1985), LFT estimate RPCs as a function of regional geographic size, total demand per sector, supply-demand ratio, hotel room- nights per capita, and other sectoral and regional variables. LFT treat EU member states as "regions" of an amalgamated EU "nation" and find their approach is relatively more accurate (contrasted to the nations' true RPCs) vis-à-vis conventional rows-only trade-adjusting approaches, such as the S/D, SLQ, an older rendition of the FLQ (Flegg & Tohmo, 2013) that does not use CILQs, and Többen and Kronenberg's (2015) CHARM approach.

2.2. Interregional trade flows

Estimating interregional trade flows has received less attention of late at the subnational level than have intraregional direct requirements matrices. This is because much of the focus of the latter has been on single region IO models. Data on interregional trade between firms at the subnational level are rarely available (Hewings & Jensen, 1986). When they are available, they are complicated by logistical transfer points, i.e., the tendency of final users to buy from wholesalers and warehouses, which need not be nearthe location at which the commodity is ultimately used. Recall, "the aim is to estimate a set of flows among several origins and several destinations, separated in space" (Sargento et al., 2012, p. 174).

Over time, an accumulation of approaches has amassed to estimate interregional trade flows (Sargento et al., 2012). Despite the diversity of their theoretical foundations, however, Batten and Boyce (1986, p. 357) conclude that "are more notable for their similarities than for their differences." Ultimately, all apply a gravity model of some sort. Considering this, we point the reader to the two above reviewresources and move forward with a focus on approaches that use gravity models, the fullfamily of which has been reviewed by Sen and Smith (1995), and Isard (1998) provides a primer.

Again, we chose to ignore mathematical programming approaches, at least now in this exercise. We made this choice knowing that models like Cai's (2021) doubly constrained gravity model should yield greater precision. Instead, we chose to work witha gravity model + RAS approach.

How the approach is effected depends on whether regional direct requirements matrices exist for the regions or not. If they do, then the gravity model operates on each region's excess supplies and excess demands by industry as in Yamada (2015), Fournier Gabela (2020), and Fei (2020). If not, it operates on each region's shares of national supplies and demands by industry as in Leontief (1953), Sargento, Ramos, and Hewings (SRH, 2012), Boero, Edwards, and Rivera (2018), and others. In addition to information on supplies and demands by commodity, the gravity model uses shipping costs between region pairs, or their proxies—travel times, or distances. Some analysts (e.g., Fei, 2020) have also broken out shipping costs that depend upon the transport mode used for the commodity.[11] While some survey information of this sort is available in some countries, it is typically tough to apply to regional work due to small, seasonal samples applied in the survey work.

Isard (1960), Leontief and Strout (1963), Isard (1998), and Yamada (2015) apply the same basic gravity model. It is:

$$t_i^{rs} = k_i \frac{(t_i^{r\bullet})^{\alpha_i}(t_i^{\bullet s})^{\beta_i}}{(tt_i^{rs})^{\gamma_i}} \tag{4}$$

where $t_i^{r\bullet}$ is region r's excess supply, $t_i^{\bullet s}$ is region s's excess demand for i, tt_i^{rs} is the shipping cost from r to s of commodity i, k_i is a gravity coefficient that appropriately scales the gravity relationship for commodity i, and α_i, β_i, and γ_i are estimated parameters for commodity i that are sometimes borrowed from prior studies.

3. Data: FIGARO

To test the various approached we decided to use FIGARO, a GMRIO data set focused on Europe (Remond-Tiedrez & Rueda Cantuche, 2019). We believe it has a good balance of country and industry detail for all European Union (EU-27) countries. FIGARO also has recent data compared with other GMRIOs available with fully comparable accounts from 2010 through 2020 (Piñero et al., 2022). We use the 2019 product-by-product model data as a benchmark.

All national FIGARO commodity-by-industry tables have $n = 64$ commodities. The database includes $g = 45$ countries plus a "rest of the World" region (FIG, according to FIGARO codes). Final demand matrices $\mathbf{F}^{rs}$ have $l = 5$ columns: (i) household consumption, (ii) collective consumption, (iii) government spending, (iv) gross fixed capital formation and (v) inventory variations. National value-added matrices $\mathbf{W}^s$ have $p = 3$ different rows: (a) compensation of employees, (b) gross operating surplus and (c) other net taxes on production.

For the sake of the gravity model, we assume road transportation as the default option. We set the locations of every country in their capital cities and use Google Maps travel times.[12] When road transportation is not possible, maritime transportation is our second option. In this case, we

[11] Clearly, bulk products like grains and minerals cost less to ship as they tend to be transported by slower, less expensive modes like water transport and rail, rather than truck. On the opposite end of the spectrum, when freight carrying costs—breakage, insurance, and storage (including time in transit)— make up relatively high shares of transportation costs, faster, more expensive modes of transport (e.g.,air freight) are likely engaged.

[12] We assume truckers expense highway tolls and use the fastest routes.

calculate distances based on each country´s biggest freight port. We calculate the shortest path between ports and assume an average speed of 40 km/h (21.60 nautical knots per hour).

Recently, de la Torre Cuevas and Lahr (2023) estimated the gravity equation (4) for some small and large EU countries and found that $\alpha \cong 1$, $\beta \cong 1$, and $\gamma \cong 0.5$ on average across shipped commodities, which are close to those found by Yamada (2015). Using a broader set of countries or regions others have found all parameters are close to 1 (Chaney, 2018; Hillberry & Hummels, 2003; Martínez San Román et al., 2012). After some minor testing, we decided to let $\alpha = \beta = \gamma = 1$.[13]

Following the logic of Lahr, Ferreira and Többen (2020), we aggregate all EU-27 countries to form a "national" model. That is, we sum up the 27 tables cell by cell to form a single 64×64 set of commodity accounts that depicts interindustry shipments for the entire EU. Using these data, we then fabricate EU MRIO tables in the four different ways described in Section 2.1. In all cases, we approximate regional industry output using regions' shares of national industry employment. Industry demand is estimated via the product of the national direct requirements coefficients matrix, for both intermediate and final uses, and outputs, where outputs of final demand are estimated using regions' shares of the nation's population.

In one case of the four—the integrated approach, we apply a gravity model and estimate all trade flows simultaneously using a supply-demand pool approach. We use an internal distance measure applied by Keeble et al. (1982), among others, and assume an internal average travel speed of 90 km/h:

$$d^{rr} = 1/90\sqrt{area^r/\pi}\,. \qquad (5)$$

In the others, we estimate intraregional shipments and use a four-step process. The three other estimating techniques differ only in the way intraregional shipments are estimated in *Step 1* and as described in Subsection 2.1.1. That is, we estimate the intraregional blocks in which $o = d$. Recall, the three alternative approaches are Haddad's S/D, CILQ-based FLQs,[14] and econometric RPCs. In *Step 2*, we use each region's intermediate industry excess supplies (local

[13] We tested both 0.5 and 1.0, and the difference was quite small between the two sets of resulting estimates. In fact, the difference between "actual" and "estimate" was very slightly lower the set of all EU countries with γ =1.

[14] We have not yet "optimized" the FLQ approach on δ. For the time being, we simply applied the rough value of 0.3, the midpoint value suggested by Jahn, Flegg, and Tohmo (2016). We understand from Flegg and Tohmo (2016, 2019) and Flegg et al. (2023) that this lack of present diligence could result findings biased against the FLQ+. Of course, the point of any FLQ is to keep things simple, otherwise a full econometric approach to estimating any sort of trade intra- or inter-regional should yield superior results.

industry-wise outputs not used locally) and excess demands (the region's industry-wise demands not fulfilled by local producers). In essence, from total industry supplies and demands by region, net out the row/column sums of our intraregional estimates. We then apply a gravity model to allocate the excess supplies and demands across the other regions. In *Step 3*, we perform a similar supply-demand pool allocation for final demands. In *Step 4*, we balance all flows generated in Step 2 and Step 3 using GRAS (Junius & Oosterhaven, 2003). Then after the GRAS adjustment, we enter the intraregional trade estimates into the main diagonal blocks for each region.

The above is performed in a way that permits the full population of cells in the interregional trade blocks ($\mathbf{T}^{o\neq d}$). Of course, even nations do not typically track which industries demand imports; that is, trade data are typically only available by commodity and only sometimes by the origin or shipping industry of exporting establishments. Rarely are the receiving (or destination) industries known. Depictions in GMRIOs like FIGARO of off-diagonal international trade arise during consolations of commodity trade accounts. Thus, in addition to full interregional trade matrices, we also opted to generate trade estimates that strictly populate the main diagonal of each trade block. In the case of final demand, we consolidate trade flows into a single column of interregional exports for each region. To conclude, international exports are shared out according to employment data in all models.

Figure 1. Description of a symmetric MRIO model

	Intermediate transactions			Final demand		
	1	⋯	m	Regional	Foreign	Σ
1	$\mathbf{T}^{11}$	⋯	$\mathbf{T}^{1m}$	$\mathbf{F}^{1}$	$\mathbf{e}^{1}$	$\mathbf{x}^{1}$
⋮	⋮	⋱	⋮	⋮	⋮	⋮
m	$\mathbf{T}^{m1}$	⋯	$\mathbf{T}^{mm}$	$\mathbf{F}^{m}$	$\mathbf{e}^{m}$	$\mathbf{x}^{m}$
m	$\mathbf{m}^{1}$	⋯	$\mathbf{m}^{m}$	$\mathbf{m}^{f}$	–	$\mathbf{m}^{\bullet}$
GVA	$\mathbf{w}^{1}$	⋯	$\mathbf{w}^{m}$	–	–	$\mathbf{w}^{\bullet}$
Σ	$\mathbf{x}^{1}$	⋯	$\mathbf{x}^{m}$	$\mathbf{f}^{\bullet\bullet}$		

Source: author´s elaboration.

The structure of our model is depicted in Figure 1. For the sake of simplicity, we restrict trade in final demand. Matrix **F** only contains two columns of interregional and international exports instead of a full trade pattern as in **T**. We also sum value-added components into value-added vectors **w**. Finally, we account for international imports and exports in a specific row (**m**) and column (**e**). Imports include those shipped to final demand.

4. Provisional Outcomes

We examine seven different sets of subnational MRIO accounts. Recall the seven sets are the simple gravity model (SGM) approach plus two each for Haddad's (2014) S/D pool approach, the CILQ-based approach by Flegg and others (FLQ+), and an econometrically estimated RPC approach (see Appendix A for a description). Estimated RPCs for service industries in the RPC approach followed those applied in Haddad's S/D pool approach (see footnote 10). As mentioned earlier, we set the FLQ+ δ parameter to .3. The two versions of the MRIOs for each of the S/D, FLQ+, and econometric RPC approaches differ only in how interregional trade is estimated. That is, we retain the intraregional trade (on-diagonal) matrix partitions. But rather than assume one can estimate the full-set of interregional trade in the off-diagonal matrix partitions, we instead assume that we can at best only know only the industry disposition of imports at the destination region by the industry that exports them at the region of origin, and so force interregional trade to be on the diagonal only of the off-diagonal partitions.

We probe the veracity of these seven approaches using four different perspectives: output multipliers, value-added multipliers, the Leontief inverse (L), and the direct requirements matrix A. Recall that SGM is at a disadvantage in that its regional A matrices are not productivity-adjusted to reflect local-area value-added shares; the A matrices for the other three approaches have been so adjusted.

To explore these four perspectives, we calculated four different distance measures: mean absolute deviation (MAD), weighted absolute deviation (WAD), mean absolute percentage error (MAPE), and weighted absolute percentage area (WMAPE). But we only report two here—MAPE and WMAPE. We do so because they are easier to interpret since they report percentages; also, it turns out, at least in our reams of results, that the WMAPE tends to mimic rankings across counties and even magnitudes of distance as measured by MAD and WAD. Multiplier WMAPEs are weighted using regions' industry output values and as suggested by Oosterhaven (1981), we subtract 1.0 (the direct effects) from the output multiplier so the distance measures focus on the multiplier effects rather than the relative size of the direct effects. Weights for A and L are cell-specific; that is, each cell of the interregional **T** matrix is divided by its column total. In summary, big industries in big countries get bigger weights.

Table 2 shows the performance of all four approaches with a full **T** on output multipliers (minus 1.0—the direct effects) are very close but tend all to be off target on the order of 23% to 30%. The econometric RPC approach is marginally best and, for different reasons, SGM (it wins more best-in-country awards) and S/D are close seconds. The FLQ+ yields results that are apparently least accurate, on average, but likely difficult to identify as much different from the other three alternatives. Note that all approaches fail to do well for Finland, Ireland, and The Netherlands, and perform much better for Germany, France, Poland and Romania. In any case, given its minimal data requirements (no value-added information, just jobs), the SGM performs extraordinarily well.

Table 2. The relative accuracy of output multipliers (minus 1.0) by country
(bold-faced font denotes country's lowest-valued MAPE or WMAPE)

	SGM		**S/D**		**FLQ+**		**Econometric**	
	MAPE	**WMAPE**	**MAPE**	**WMAPE**	**MAPE**	**WMAPE**	**MAPE**	**WMAPE**
AT	29.6	**25.1**	28.8	25.9	33.6	28.3	**28.3**	25.5
BE	**22.1**	**16.9**	24.2	20.3	24.3	18.4	23.7	20.0
BG	**19.5**	**17.5**	23.2	20.1	21.0	18.3	22.6	19.6
CY	37.5	37.2	**28.3**	**23.8**	36.4	35.5	29.5	24.4
CZ	25.3	21.7	24.7	22.2	28.1	23.8	**23.8**	**21.5**
DE	20.0	17.8	18.9	17.7	**17.9**	**16.0**	19.2	18.1
DK	**25.0**	**21.6**	27.2	23.6	28.7	24.6	26.3	22.7
EE	22.6	22.6	**19.4**	**18.4**	20.6	20.1	19.6	18.8
ES	22.5	17.1	22.3	16.3	22.3	18.6	**22.3**	**15.8**
FI	55.6	57.2	**33.2**	**34.1**	51.9	53.5	33.9	34.7
FR	17.6	14.2	16.1	13.4	16.2	13.2	**15.7**	**13.1**
GR	38.9	40.0	30.0	23.1	37.4	37.1	**29.9**	**23.0**
HR	**25.8**	**20.4**	28.7	23.0	29.2	23.1	27.6	22.0
HU	31.9	24.8	32.0	25.8	35.9	28.0	**31.1**	**25.1**
IE	**64.6**	**35.5**	85.2	44.0	68.9	37.6	83.8	43.9
IT	21.8	19.9	19.8	17.0	23.1	21.9	**19.8**	**16.7**
LT	**29.6**	**26.2**	36.6	32.5	33.6	29.6	35.5	31.4
LU	27.6	22.7	27.6	22.8	29.9	23.7	**27.1**	**22.5**
LV	**21.0**	19.4	23.0	19.6	22.1	19.3	22.1	**19.1**
MT	**23.0**	**25.9**	25.8	29.1	30.0	33.2	26.2	29.5
NL	**43.9**	**37.3**	46.2	40.8	46.3	39.3	45.8	40.5
PL	19.0	15.9	18.7	15.2	19.0	15.7	**18.3**	**14.9**
PT	20.7	19.6	20.3	16.6	20.7	18.9	**20.2**	**16.6**
RO	19.1	18.5	**17.3**	**15.9**	18.9	18.1	17.5	16.3
SE	25.9	26.0	19.9	18.8	23.4	23.1	**19.8**	**18.6**
SI	**32.7**	**28.3**	34.2	30.2	37.1	31.4	33.0	29.2
SK	27.9	24.1	26.3	23.3	31.2	26.6	**25.3**	**22.5**
mean	28.5	24.9	28.1	23.5	29.9	25.8	**27.7**	**23.2**

On other economic measures (value-added multipliers, **L**, and **A**), however, econometric RPCs and S/D yield half the error on average of SGM and FLQ+ (see Table 3). This speaks well for all but FLQ+, which is the worst and yet has data requirements verging on those of the econometric approach. RPC approach is best for most counties and at worst ranked second for those for which it is not.

Table 3. Mean error by subnational MRIO table production approach
(bold-faced font denotes lowest-valued MAPE or WMAPE in the row)

	SGM		S/D		FLQ+		Econometric	
	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE
Full IRIO								
Output multipliers	28.5	24.9	28.1	23.5	29.9	25.8	**27.7**	**23.2**
Value added multipliers	26.3	21.1	25.0	19.3	26.9	21.8	**24.7**	**19.0**
Leontief inverse	156.7	1.0	1,303.3	0.7	**99.1**	1.1	1,007.2	**0.7**
Direct requirements	415.1	2.7	1,503.8	2.0	**104.2**	2.8	1,462.0	**1.9**
Diagonal IRIO								
Output multipliers	67.4	57.4	**54.5**	**42.7**	94.9	77.5	64.4	50.0
Value added multipliers	67.4	60.8	**30.1**	**23.1**	94.3	84.8	64.4	55.3
Leontief inverse	109.3	1.0	1,303.8	0.7	**97.0**	1.1	1,460.8	**0.7**
Direct requirements	204.4	2.7	1,503.8	2.0	**104.2**	2.8	1,537.4	**2.0**

When we permit trade only on the main diagonal—a reality rather than a practicality—accuracy gets worse as might be expected (see lower half of Table 3). For output multipliers, they worsen least for RPCs and most for FLQ+, for which error nearly triples in the average WMAPE (~27%=> 78%). The others less than double in the average WMAPE (~25%=>~50%). Again, SGM appears to be the big winner because of its minimal data requirements.

In value-added multipliers, S/D appears to be the best by far. Its errors are almost the same as for its full-fledged interregional equivalent. Accuracy findings for the other three appear to echo equivalents for output multipliers. That is, those for FLQ+ remain by far worst.

For **L** and **A** things do not change much with a move to just diagonal flows, which is surprising and yet good news. It is evident from differences between MAPE and W**M**APE that S/D and RPC approaches tend to match up with larger-valued cells better than do the other two. But they also **<u>really</u>** fail to get smaller values right. In this regard, FLQ+ is downright non-discriminatory on cell values, where average MAPE errors across countries range from just 94.7% to 126.0%. in S/D and RPC the range is more like 56% to 18,000% and SGM is 75.6% to 331.7%.

5. Conclusions

Herein, we analyze the relative accuracy of basic techniques typically used nowadays to estimate subnational MRIO models. For now, we purposely bypass techniques that employ mathematical programming (optimization) techniques. We do so by seeing how well different fundamental approaches can replicate intra- and inter-regional interindustry trade flows in the EU-27. The simplest of them (the simple gravity model, SGM) uses only employment information for each industry in each region and estimates of shipping distances within a region and between region pairs. In others, we assume that a superior estimate of intraregional interindustry trade flows can be constructed; to do so, we use Haddad's (2014) supply/demand (S/D) pool approach, the FLQ+ approach (Flegg & Tohomo, 2016), and an econometrically estimated RPC approach that employs a binomial beta regression. In effecting these three approaches, we assure the regional direct requirements are also productivity-adjusted by column-wise re-estimating national direct requirements, so they reflect each region's industry value-added shares. For each, we first assume that officials know the sectors that import interregional shipments and use a gravity model and RAS to populate a full interregional trade matrix. We create a second set of MRIOs to reflect the actual case—regional officials at best know the extent to which industries in their region export. In this case, we limit interregional shipments so that receiving industries in destination regions are the same as the origin industries of the regions that shipped them. We estimate these MRIOs for the EU-27 in 2019 by aggregating the 27 national accounts into a single 64-commodity by 64-commodity meta-account. This account effectively forms the "national" transactions matrix, a fundamental requirement for producing subnational input-output tables: one that assumes knowledge transfers freely across geographic space.

We compare all eight alternatives to FIGARO GMRIO accounts of the EU-27. To compare, we rely on the weighted mean absolute percentage error (WMAPE) measure informed by the mean absolute percentage error (MAPE) measure. We find that SGM performs relatively well despite its very limited data requirements. The FLQ+ performs worst (often 50% worse) via several perspectives; the RPC econometric approach tends to perform best, as expected due to its steeper set of data requirements.

Limiting interregional trade to the diagonal of off-diagonal partitions induces substantial error. This is no surprise since one must compare zero-valued cells in the estimated table to cells with nonzero values in the actual (2019 FIGARO) table. This point deserves more exploration since imports by using industry are not typically known, even in global and international accounts. Note WIOD and FIGARO report them as if they are known, but at best their fully population interregional trade matrices result from industry aggregation of import accounts for which only the commodity code of the shipment is known.

We must note that better approaches than those that we test here likely exist—e.g., doubly constrained gravity models—since we limited our set of tests to approaches that use simple algorithms. Also, while RAS likely heals all sore spots, it also can possibly enable error to exaggerate and fester.

REFERENCES

Batten, David, & Dino Martellato. (1985). Classical versus modern approaches to interregional input-output analysis. *Annals of Regional Science*, 19(3), 1–15.

Bajzik, Josef, Tomas Havranek, Zuzana Irsova, and Jiri Schwarz. (2020). Estimating the Armington elasticity: The importance of study design and publication bias. *Journal of International Economics* 127, 103383.

Benkovskis, Konstantins, Benjamin Bluhm, Elena Bobeica, Chiara Osbat, and Stefan Zeugner. (2020). What drives export market shares? It depends! An empirical analysis using Bayesian model averaging. *Empirical Economics*, 59, 817-869.

Boero, Riccardo, Brian K. Edwards, & Michael K. Rivera. (2018). Regional input–output tables and trade flows: an integrated and interregional non-survey approach. *Regional Studies*, 52, 225–238. https://doi.org/10.1080/00343404.2017.1286009

Cai, Mattia. (2021). Doubly constrained gravity models for interregional trade estimation. *Papers in Regional Science*, 100, 455–474. https://doi.org/10.1111/pirs.12581

Chaney, Thomas. (2018). The gravity equation in international trade: An explanation. *Journal of Political Economy,* 126, 150–177.

Chenery, Hollis. (1953). Regional analysis. In Hollis Chenery, Paul G. Clark, & Vera Cao-Pinna, eds., *The structure and growth of the Italian economy*. U.S. Mutual Security Agency, pp. 98–139.

Cribari-Neto, Francosco & Achim Zeileis. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2), 1–24. doi: 10.18637/jss.v034.i02

Dalgaard, Ebsen & Christian Gysting. (2004). An algorithm for balancing commodity-flow systems. *Economic Systems Research*, 16, 169–190.

de la Torre Cuevas, Fernando & Michael L. Lahr. (2023). Simplifying gravity equations to embed a region within world input-output models. An unpublished paper submitted for presentation in Parallel Session 10 at the 29th International Input-Output Conference, Alghero, Sardinia Italy, June 30.

Elshahawany, Dina N., Eduardo A. Haddad, & Michael L. Lahr (2017). Accessibility, transportation cost, and regional growth: A case study for Egypt, *Middle East Development Journal*, 9, 256–277. http://dx.doi.org/10.1080/17938120.2017.1366773

Fei, Da. (2020). *U.S. State-level Consumption-based Accounting of Greenhouse Gas Emissions: A Scenario Analysis*. Unpublished Ph.D. Dissertation. Edward J. Bloustein School of Planning and Public Policy, School of Graduate Studies Rutgers, The State University of New Jersey. https://rucore.libraries.rutgers.edu/rutgers-lib/65106/PDF/1/play/

Flegg, Anthony T., Guiseppe R. Lamonica, Francesco M. Chelli, Maria C. Recchioni, & Timo Tohmo. (2021). A new approach to modelling the input-output structure of regional economies using non-survey methods. *Journal of Economic Structures* 10(1), 1–31.

Flegg, Anthony T., Xesús Pereira, Napoleón Guillermo Sánchez, Chóez, Fernando de la Torre Cuevas, & Timo Tohmo. (2023). Curve shapes and parameters in FLQ regional modelling: some alternative approaches. An unpublished paper submitted for presentation at the 29th International Input-Output Conference, Parallel Session 10, Alghero, Sardinia, Italy, June 30.

Flegg, Anthony T. & Timo Tohmo. (2013). Regional input-output tables and the FLQ formula: A case study of Finland. *Regional Studies*, 47, 703–721. https://doi.org/10.1080/00343404.2011.592138

Flegg, Anthony T. & Timo Tohmo. (2018). The regionalization of national input–output tables: A review of the performance of two key non-survey methods. In Kakali Mukopadhyay, ed., *Applications of the input-output framework.* Singapore: Springer, pp. 347–386.

Flegg, Anthony T., & Timo Tohmo. (2019). The regionalization of national input - output tables: A study of South Korean regions. *Papers in Regional Science*, 98, 601–620. https://doi.org/10.1111/pirs.12364

Fournier Gabela, Julio Gustavo. (2020). On the accuracy of gravity-RAS approaches used for inter-regional trade estimation: evidence using the 2005 inter-regional input–output table of Japan. *Economic Systems Research*, 32, 521–539. https://doi.org/10.1080/09535314.2020.1753662

Garhart, Robert E. & Frank Giarratani. (1987). Nonsurvey input-output estimation techniques: Evidence on the structure of errors. *Journal of Regional Science*, 27, 245–253.

Haddad, Eduardo. (2014). Trade and interdependence in Lebanon: An interregional input-output perspective. *Journal of Development and Economic Policies*, 1(16), 5–45.

Hewings, Geoffrey J.D. & Rodney C. Jensen. (1986). Regional, inter-regional and multiregional analysis. In Peter Nijkamp, ed., *Handbook of regional and urban economics. Vol. 1.* Amsterdam, Free University, pp. 295–355.

Hillberry, Russell & David Hummels. (2003). Intranational home bias: Some explanations. *Review of Economics and Statistics*, 85, 1089–1092. https://doi.org/10.1162/003465303772815970

Isard, Walter. (1951). Interregional and regional input-output analysis: a model of a space economy. *Review of Economic Statistics*, 33, 318–329.

Isard, Walter. (1953). Some empirical results and problems of regional input-output analysis. Chapter 5 in Wassily Leontief & Hollis B. Chenery, Paul G. Clark, James S. Duesenberry, Allen M. Ferguson, Anne P. Grosse, Robert N. Grosse, Mathilda Holzman, Walter Isard, & Helen Kistin, eds., *Studies in the structure of the American economy: Theoretical and empirical explorations in input-output analysis*. New York: Oxford University Press, pp. 116–181.

Isard, Walter in association with David F. Bramhall, Gerald A.P. Carruthers, John H. Cumberland, Leon H. Moses, Daniel O. Price, & Eugene W. Schooler. (1960). *Methods in regional analysis: An introduction to regional science*. Cambridge, MA: The M.I.T Press.

Isard, Walter. (1998). Gravity and spatial interaction models. In Walter Isard, Iwan J. Azis, Matthew P. Drennan. Ronald E. Miller, Sidney Salzman, & Erik Thorbecke, eds., *Methods of inter-regional and regional analysis*. Brookfield, VT, Ashgate Publishing, pp. 243–280.

Jackson, Randall W., Philip R. Israilevich, & Jonathan C. Comer. (1992). A note on the role of survey data and expert opinion in constructing input-output tables. *Papers in Regional Science,* 71, 87–93.

Jahn, Malte. (2017). Extending the FLQ formula: a location quotient-based interregional input-output framework. *Regional Studies*, 51, 1518–1529.

Jahn, Malte, Anthony T. Flegg, & Timo Tohmo. (2020). Testing and implementing a new approach to estimating interregional output multipliers using input–output data for South Korean regions. *Spatial Economic Analysis*, 15, (2), 165–185. https://doi.org/10.1080/17421772.2020.1720918

Kowalewksi, J. (2015). Regionalization of National Input-Output Tables: Empirical Evidence on the Use of the FLQ Formula. *Regional Studies*, 49(2), 240–250.

Lahr, Michael L. (2001). Reconciling domestication techniques, the notion of re-exports, and some comments on regional accounting. *Economic Systems Research*, 13, 165–179.

Lahr, Michael L., Joao Pedro Ferreira, & Johannes R Többen. (2020). Intraregional trade shares for goods-producing industries: RPC esimates using EU data. *Papers in Regional Science*, 99, 1583–1605. https://doi.org/10.1111/pirs.12541

Lau, Samuel W. (1995). *Truck travel surveys: A review of the literature and state of the art*. Metropolitan Transportation Commission. Oakland, CA. Available online in May 2023 at https://rosap.ntl.bts.gov/view/dot/13308/dot_13308_DS1.pdf.

Leontief, Wassily. (1936). "Quantitative Input–Output Relations in the Economic System of the United States," *Review of Economics and Statistics*, 18, 105–125.

_____. (1941). *The structure of American economy 1919–1939*. New York: Oxford University Press.

Leontief, Wassily. (1953). "Interregional Theory". Chapter 4 in Wassily Leontief & Hollis B. Chenery, Paul G. Clark, James S. Duesenberry, Allen M. Ferguson, Anne P. Grosse, Robert N. Grosse, Mathilda Holzman, Walter Isard, & Helen Kistin, eds., *Studies in the structure of the American economy*. New York: Oxford University Press, pp. 93–115.

Leontief, Wassily & Alan Strout. (1963). Multiregional input-output analysis. In Tibor Barna, ed., *Structural interdependence and economic development*. Palgrave Macmillan, London, pp. 119–150. https://doi.org/10.1007/978-1-349-81634-7_8

Malizia, Emil & Daniel L. Bond. (1974). Empirical tests of the RAS method of interindustry coefficient adjustment. *Journal of Regional Science,* 14, 355–365.

Martínez San Román, Valeriano, Marta Bengoa Calvo, & Blanca Sánchez-Robles Rute. (2012). European Union and trade integration: Does the home bias puzzle matter? *Revista de Economía Mundial,* 32, 173–188. Retrieved from https://www.redalyc.org/articulo.oa?id=86625395007

McCullagh, Peter & John A. Nelder, FRS. (1989). *Generalized linear models,* 2nd ed. Monographs on Statistics and Applied Probability 37. Chapman and Hall, Boca Raton, FL. https://www.utstat.toronto.edu/~brunner/oldclass/2201s11/readings/glmbook.pdf

McMenamin, David G. & Joseph E. Haring. (1974). An appraisal of nonsurvey techniques for estimating regional input-output models. *Journal of Regional Science*, 14, 191–205.

Miller, Ronald E. & Peter D. Blair. (2022). *Input-output analysis: Foundations and extensions*. 3rd ed. Cambridge, UK: Cambridge University Press.

Moses, Leon. (1955). The stability of interregional trading patterns and input-output analysis. *American Economic Review*, 45, 803–832.

Oosterhaven, Jan. (1981). Interregional input-output analysis and Dutch regional policy problems (University of Groningen). University of Groningen. https://research.rug.nl/en/publications/interregional-input-output-analysis-and-dutch-regional-policy-pro

Pakizeh, Amir Hossein & Hamed Kashani. (2022). Application of machine-learning models to estimate regional input coefficients and multipliers. *Spatial Economic Analysis*, 17, 178–205.

Piñero, Pablo, Giovanni Mandras, José Manuel Rueda-Cantuche, & Zornitsa Kutlina-Dimitrova. (2022). Trade between EU countries and the world: Deviations in official inter-country input-output tables and their implications for policy. 28th International Input-Output Conference. Langkawi Island, Malaysia, August 28- September 2. https://www.iioa.org/conferences/28th/papers/files/4342.pdf

Planting, Mark & Jiemin Guo. (2004). Increasing the timeliness of US annual input-output accounts. *Economic Systems Research,* 16, 157–167.

Remond-Tiedrez, Isabelle & José Manuel Rueda-Cantuche, eds. (2019). *EU Inter-country Supply, Use and Input-output Tables: Full International and Global Accounts for Research in Input-output Analysis (FIGARO).* Luxembourg: Publications Office of the European Union. https://doi.org/10.2785/008780

Riefler, Robert & Charles Tiebout. (1970). Interregional input-output: an empirical California-Washington model. *Journal of Regional Science*, 10, 135–152.

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

https://www.academia.edu/71493693/Introducing_Input_output_Analysis_at_the_Regional_Level_Basic_Notions_and_Specific_Issues

Sargento, Ana L.M., Pedro N. Ramos, & Geoffrey J.D. Hewings. (2012). Inter-regional trade flow estimation through non-survey models: an empirical assessment. *Economic Systems Research*, 24, 173–193. https://doi.org/10.1080/09535314.2011.574609

Sen, Ashish & Tony E. Smith (1995). *Gravity models of spatial interaction behavior*. Berlin, Springer-Verlag

Többen, Johannes R. & Tobias H. Kronenberg. (2015). Construction of multi-regional input-output tables using the CHARM method. *Economic Systems Research*, 27, 487–507. https://doi.org/10.1080/09535314.2015.1091765

Treyz, George I. and Benjamin H. Stevens. (1981). Location analysis for multiregional modeling. In F. Gerard Adams & Norman Glickman, eds., *Modeling the multiregional economic system*. Lexington, MA: Lexington Books, pp. 75–87.

Treyz, George I. & Benjamin H. Stevens. (1985). The TFS regional modelling methodology. *Regional Studies*, 19, 547–562. https://doi.org/10.1080/09595238500185531

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, & Gaaitzen J. de Vries. (2015). An illustrated user guide to the World Input–Output Database: the case of global automotive production, *Review of International Economics*, 23: 575–605.

Valderas-Jaramillo, Juan Manuel, José Manuel Rueda-Cantuche, & Joerg Beutel. (2021). The Euro and SUT-RAS methods: some further considerations. *Economic Systems Research*, 33, 276–286.

Valderas-Jaramillo, Juan Manuel, José M. Rueda-Cantuche, Elena Olmedo, & Joerg Beutel. (2019). Projecting supply and use tables: new variants and fair comparisons. *Economic Systems Research,* 31, 423–444.

Yamada, Mitsuo. (2015). Construction of a multi-regional input-output table for Nagoya metropolitan area, Japan. *Journal of Economic Structures*, 4(11). https://doi.org/10.1186/s40008-015-0022-7

Appendix A

For the research reported here, we opted for beta binomial regression (Cribari-Neto & Zeileis, 2010) instead of the quasi-binomial regression approach employed by Lahr, Ferreira, and Tobben (LFT, 2020). The beta binomial is quite similar to the quasi-binomial but estimates parameters using maximum likelihood (ML) approach as opposed to a generalized linear model (GLM) (McCullagh & Nelder, 1989). Limiting values to lie between 0 and 1, these two formulations follow the same restrictions as the functional form employed by Treyz and Stevens (1985). Cribari-Neto and Zeileis (2010) explain that the beta binomial yield beta-distributed parameters depend on a mean and an estimated precision parameter. As in GLMs, a variable's mean is linked to the responses through a link function and a linear predictor. Additionally, the beta distribution's precision parameter is linked to another (potentially overlapping) set of regressors through a second link function, resulting in a model with variable dispersion. The approach has been used previously in analyses of trade (Bajzik et al., 2020; Benkovskis et al., 2020) among other applications in economics.

Table A.1. Beta Binomial Regression Estimates for RPCs of Goods-producing Industries in the EU 27 in 2019[15]

Variables	Coefficients	*p*-value
Constant	-3.866	< 2e-16 ***
ln(land area)	0.095	9.84e-06 ***
ln(hotel room nights)	0.099	0.000379 ***
Location quotient (max = 1)	0.072	0.002809 **
S/D (max =1)	2.234	< 2e-16 ***
weight/value of good	0.052	3.70e-05 ***
Agriculture and mining	1.037	< 2e-16 ***
Food, beverage, & tobacco	0.551	2.62e-05 ***
Textiles	-0.493	0.000621 ***
Printing	1.609	< 2e-16 ***
Chemicals	-0.459	7.05e-06 ***
Electrical components and machinery	-0.617	3.39e-12 ***
R^2	.646	

Table A.1 above and Table 3 in LFT use the same set of variables from the same and explain similar amounts of variance in the actual RPCs (R2 is quite similar). This is even though FIGARO's data are more recent (2019 rather than 2016) and have a few more goods-producing sectors. An employment LQ replaces the supply LQ in LFT. Its direction of influence (sign) is expected to be the same, however. The bigger the LQ is, bigger the RPC value should be. The direction of influence remains for all other nonbinary independent variables. The greater the

[15] The R algorithm that we used is available at https://cran.utstat.utoronto.ca/web/packages/betareg/

region's land area, the more overnights spent by tourists, the higher the weight/value ratio of the commodity (the more commoditized the good is), and the larger the S/D, then the more likely it is to satisfy a greater share of the region's local intermediate industry and final demands. Further, binary variables of some industries reveal regional preferences to consume locally produced products: primary industries, Food, beverage, and tobacco manufacturing, and the printing industry (note their positive coefficients point to higher-valued RPCs). Other sectors tend to be more involved in global value chains: textiles, chemicals, and electrical components and machinery industry, and, thus, are more apt to have lower-valued RPCs, i.e., local intermediate industry and final demands are more likely to be satisfied by suppliers from abroad and to fulfill demands of other regions. That is, they are industries in which cross-hauling is more likely to be the rule than the exception.

The sectoral distribution of errors in our binomial beta regression is displayed in Figure A.1. Their distribution across countries is displayed in Figure A.2. Note with a few exceptions that means tend to be close to zero and the dispersion of errors is generally limited.

Figure A.1. Error by sector using binomial beta regression

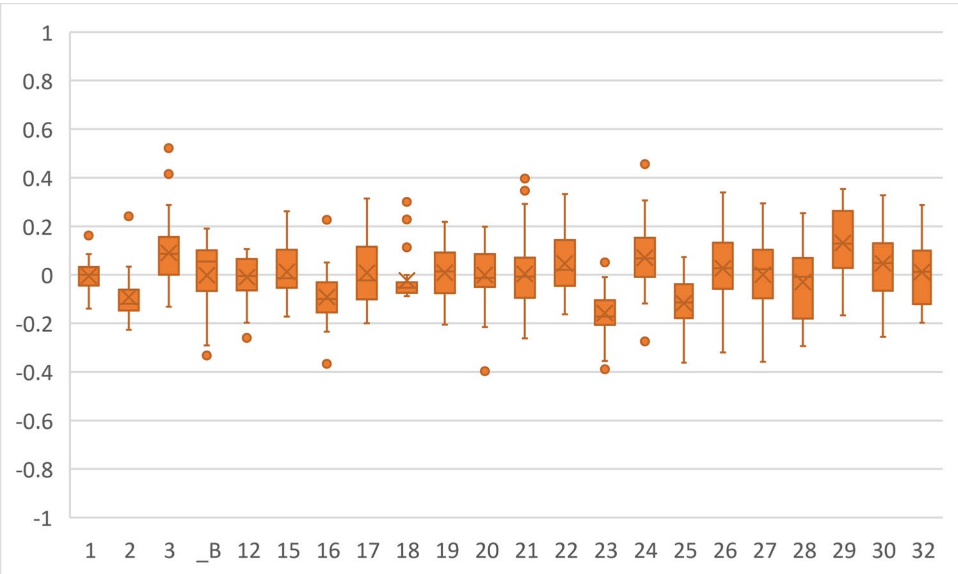

Figure A.2 - Error by country using binomial beta regression.