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

ABSTRACT. A basic underlying assumption in most of the research to date is that intermediate industry accounts of the economies in multiregional input-output (MRIO) 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 less empirical information than is available for economies of nations. Moreover, intra-economy intermediate-industry flows are typically larger than the set of a region's commodity in- and out-flows. So, if intermediate industry flows in a set of MRIO accounts are noticeably mis-estimated, it follows that interregional trade coincidentally derived using them must be even more conspicuously in error.

We hypothesize as more information is used to estimate MRIO accounts, the better the estimates should be. We start our experiment by consolidating 2019 FIGARO accounts of the 27 member states of the European Union, while maintaining sectoral detail, to produce a "national account". We then test several approaches to constructing MRIO tables. The approaches distribute interregional trade fully by receiving industry, as in FIGARO, as well as strictly in the form of a diagonalized matrix as if the commodity inflows are competitive imports. To do this, both a gravity model and RAS are applied to each approach. We then test to see how well the approaches estimate main features of FIGARO's MRIO accounts and detail a rather consistent ranking of the relative accuracy of them. We also find that the level of error inherent to the estimated MRIOs is markedly similar across approaches, particularly for multipliers. Further, relaxing interregional trade to a diagonalized matrix tends to add very little error. The approach that uses the least data is, however, markedly worse in replicating countries' direct requirements matrices and Leontief inverses, which suggests its use in a more-limited set of applications.

Keywords: Input-Output Analysis, Multiregional Input-Output, RAS, Gravity Models, Regional Purchase Coefficients, Economic Modelling

1. Introduction

Not too long after Wassily 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 administrative 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. In fact, Isard's model was the first multiregional IO (MRIO) table conceived, at least in literature. It predates by decades any proposal in the literature for a global MRIO (GMRIO).[1]

The story is told that Leontief was at least mildly appalled by Isard's (1951) lack of empirics.[2] That is, Isard (1951) does not develop a set of actual interregional IO accounts 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 linkages but

[1] Meadows and Meadows (1972) appear to be the first to suggest such a framework, while Leontief (1974) formulates one. Although Leontief (1953a) certainly must have yearned for one when working on his trade paradox.

[2] Hulu and Hewings (1993) provide additional historical insight into subnational interregional IO modelling.

also for all types of final demands. This demand for data has caused few attempts at survey-based subnational interregional IO (IRIO) table construction.[3] Japanese (1960-2005) and South Korean (2005) IRIO tables may be notable exceptions (Miller & Blair, 2022).[4] A prime difficulty is collecting data on interregional commodity trade flows, a relatively easy task at an international level.[5]

Despite data scarcity, there has been a sort of revival in interest in subnational MRIO models.[6] Institutions and scholars continue to be motivated by the need to address regional heterogeneity within countries and/or trading blocs (Monfort, 2020). Added interest derives from a rising interest in the role of regions with global value chains. Different from Isard´s (1951) original idea, these latter-day models tend to combine different geographical scales such as regions and countries (Bachmann, Roorda, & Kennedy, 2015). In addition to efforts in South Korea and Japan, recent officially sponsored models include those by IDE-JETRO (Inomata, Yamano, & Meng, 2013), EUREGIO (Thissen, Lankhuizen, van Oort, Los, & Diodato, 2018) and the still experimental FIGARO-REG (López-Álvarez, 2023). While compiled by or for national and international statistical institutions, this new batch of MRIO models does not tend to rely heavily on much survey-based data.

---

[3] We feel compelled to acknowledge the contributions of Professor Karen Polenske (1970, 1972, 1980). They result from research conducted by the Harvard Economic Research Project for the Economic Development Administration, U.S. Department. of Commerce. These and other publications by Polenske and her research team culminated Isard's and Leontief's visions.

[4] 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, Dietzenbacher, Los, Stehrer, & de Vries, 2015)—the so-called "mirror puzzle of trade"; while irritating to analysts the "inconsistency" is typically relatively small in broader perspective, often simply relating to the 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.

[5] 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).

[6] A renewed interest in interregional IO analysis started after computing capacity improved, particularly with the advent of personal computing. Moreover, it helped that Beyers (1980) showed the magnitude of the interregional feedback effects could outsize intraregional indirect effects of a region in practical way, countering results from Miller's (1966, 1969) tests for U.S. states, which suggested that feedback effects tend to be small.

In global MRIO models, national direct requirements tend to be known. This is far less often the case for subnational equivalents.[7] That is, survey-based regional IO tables have become increasingly rare. When survey work is used to construct regional tables, they typically rely much more heavily on nonsurvey techniques and scant survey data.[8] Batten and Martellato (1985, p. 4) review the nature of 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." Rather than update Batten & Martellato's excellent review of approaches, we use their framework to home in upon two of Sargento's (2009) three dimensions of subnational IO models. By eliminating the single-region case that she discusses, we exclude all models that cannot account for interregional spillover and feedback effects. Thus, we focus on i) the way in which intraregional shipments are estimated, and ii) how interregional trade flows are detailed. We do this to identify the approaches that are best for estimating subnational MRIOs, given the general paucity of data available for the task.

2. Research Approach: a simple comparison across alternative methods

The general conceptual approach to our research is to apply techniques akin to those used by statistical agencies (see, e.g., Dalgaard & Gysting, 2004; Planting & Guo, 2004; Valderas-Jaramillo et al., 2019, 2021). Thus, herein, we do not examine some of the most sophisticated approaches available in academic literature. For example, for now, we purposely opt not to employ mathematical programming (Cai, 2021; Flegg et al., 2021) or artificial intelligence techniques (Pakizeh & Kashani, 2022). We limit ourselves to apply gravity models and biproportional adjustment techniques.

2.1. Intraregional shipments

On the matter of producing subnational IO tables, Miller & Blair (2022, p. 65) note that as early as Isard & 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

[7] 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 and are therefore only "multi-regional" in a more global sense.

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

are fulfilled by local supplies—and, therefore, naturally were applied row wise to the direct requirements matrix **A**.[9]

Garhart & Giarratani (1987) were likely 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 an industry's RPC is the row's average regional supply share when applied to technology (**A**) matrix and that the regional supply share for each element in the row undoubtedly varies from that average. Thus, ideally a full matrix of RPCs should be applied to **A**, and thus a diagonal matrix of a vector of RPCs is 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 & Giarratani's (1987) assertions (see, e.g., Malizia & Bond, 1974; McMenamin & Haring, 1974). Unfortunately, survey-based data on intermediate input and intermediate output totals by industry at a regional level are scarce, published intermittently, often limited to a subset of regions within national space and, when available at all, for an aggregated set of industries. While data on value added by industry are frequently available for subnational areas (e.g., for U.S. states and the EU's NUTS2 regions), they are available only for sectors that are aggregates of national benchmark tables; moreover, data on output and final demand by either industry or commodity tend not to be available at all.

In this vein, we take two fundamental approaches to our research—one is bottom-up and the other top-down. In the bottom-up approach, shipments that are strictly intra-regional are estimated a priori. Estimates of intra-regional shipments instead emerge from the top-down approach.

2.1.1. The top-down (integrated) approach

The first approach is the simplest—the top-down approach, a.k.a. "the integrated approach" (Boero, Edwards, & Rivera, 2018). It follows Leontief's (1953b) except for the way industry outputs and demands are allocated across regions.[10] Data requirements are minimal since the

[9] 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.

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

approach assigns national output to regions via each region's shares of the nation on some economic measure by industry, e.g., jobs or value-added.[11] And each region's share of the nation's final uses by sector is identified via 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 appropriate length, in this case n). Thus, we obtain total output and total demand by industry for each region, which (more likely than not) is unbalanced. Moreover, we assure that the sum across all regions of both output and demand sum to the nation's totals. In this case, however, note that we do not *yet* identify the set of intra-regional trade flows; This is done later, simultaneously with estimating interregional trade flows.

2.1.2. Bottom-up regionalization approaches

Over time, many approaches have been applied to estimate a single region's intraregional trade flows. Recent publications (e.g., Lahr, Ferreira, & Többen, 2020; Flegg et al., 2021) winnowed out some of the chaff; that is, many approaches have clearly proven to be inferior. Thus, we ignored those known inferior approaches. We make an exception for the supply/demand ratio truncated at values below 1.0 à la Haddad (2014) since it has been widely and recently used to produce subnational MRIO models in data-scarce contexts (Haddad, 2014; Haddad et al., 2017). We identify two parametric approaches; and as such, at least for the time being, used information on intraregional trade by industry to estimate the required parameters. Because of this, we treated the remaining two alternatives equally and as if much more information was "known." Both also account for possible commodity cross-hauling, the notion that commodities are simultaneously shipped into and out of a region.[12] In all cases, we assume no information is available for value added or final demand by industry. We reserve an examination of the influence of these additional data for future research. Such an examination will be worthwhile since such data are available for states within the U.S. as well as for NUTS2 regions in the EU. For now, however, we find an

---

region's share of all transactions for a particular industry. Included in this category is the work of Round (1978, 1979, 1983), which has since been extended by Boomsma & Oosterhaven (1992) and Hulu & Hewings (1993) who apply RAS in a multistage process that assures a balance of national accounts while also breaking out trade between two regions. A referee reports that this process simplifies the consistency between import and export estimates.

[11] Note, rather regional shares of jobs by industry, that, if available, compensation or value-added shares by industry should yield better allocations of national output across regions since they should better account for productivity differences.

[12] This tends to be the rule than the exception, even for services (Treyz & Stevens, 1981). Note that this can happen when the commodity being shipped is the same, e.g., electricity, and hence due to the tyranny of space, or when there are even the slightest differences in product variety i.e., the case of boiling potatoes being shipped into Idaho while that same state ships out baking potatoes.

investigation into approaches using minimal data useful for developing nations (e.g., Egypt as in Elshahawany, Haddad, & Lahr, 2017) and nations in economic transition (e.g., Voicu & Lahr, 2022).

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 (SD) 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 SD 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).[13] Including the F(c), Haddad then applies the resulting vector row wise to national flows to obtain estimates of intraregional shipments by industry for each region.

Let $\mathbf{x}$ and $\mathbf{w}$ be the output and value-added arrays, respectively. Vectors $\mathbf{x}$ and $\mathbf{w}$ can be either known through surveys or shared out from national data using a proxy (e.g., we use employment). For region r and commodity i our version of Haddad´s (2014) equation is:

$$SHIN_i^r = \mathrm{Min}\left\{\frac{\mathbf{x}-\mathbf{w}}{\mathbf{A}\hat{\mathbf{x}}\mathbf{i}}, 1\right\}\mathrm{F(c)} \tag{1}$$

The only change we introduce consists of considering total intermediate supply and demand by industry, thus neglecting final demand.

2.1.2.2. FLQs

Flegg & Tohmo (2019) evaluate 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. Note that the use of CILQs enables a full matrix of RPCs. The CILQs also have the potential to heavily distort technology (recipes for making commodities). That is, CILQs can potentially both increase and

[13] 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/non-base production. Note, however, that both Stevens, Treyz, & Lahr (1989) and Jahn, Flegg & Tohmo (2020) suggest a value closer to 0.3 is probably more appropriate than 0.5 for modifying location quotients or SD.

In their FORTRAN program, Treyz & 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 SD 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.

decrease values of individual elements within a column of the **A** that could essentially force technology to be producing something other than the targeted commodity, i.e., they are quite distant from the original representations of national average input structures. The CILQs are further modified by a parameter δ that its authors suggest adjusts for the degree of inherent cross-hauling. Empirical work (e.g., a linear search) is required to identify the proper value of δ, so that one must know actual trade flows to calibrate it. Jahn, Flegg, & Tohmo (2020) note that a series of empirical tests conducted across different geographies suggests 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. In one of the most recent FLQ papers (Flegg et al., 2021), δ is estimated based on a first stage, entropy-generate, regional tables. We provide further detail in Appendix B. This extension of the FLQ approach has been labelled as "FLQ+."

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

$$a_{ij}^r = a^n \mu_{ij}^r \lambda_i \tag{2}$$

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):

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

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. (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, i.e., they commit to a curve-fitting exercise rather than seek economic causation. We further note that output data are typically not officially reported for subnational regions, which makes the formal data requirements of the FLQ excessively demanding.

We examine two different versions of the FLQ based on different values of δ. In the first, which we identify as "FLQ," we set δ = 0.3 for every region as suggested Jahn, Flegg, & Tohmo (2020). In this way, it draws a parallel to the SD approach since it requires no advance knowledge of each region's direct requirements. Alternatively, we set δ to its optimal value, which in our case is 0.17 (see Appendix A)—that we calibrated via a linear search technique that assumes analysts know each region's direct requirements. We identify this second version as a particular case of FLQ+ in which the first estimates for regional direct requirements are not derived from entropy-based methods but, rather, are "known." FLQ and FLQ+ methods have been tested by comparing the estimated output multipliers with survey-based regional tables for 16 Korean regions (Flegg & Tohmo, 2019) and Finland (Flegg & Tohmo, 2013). Recent work by Haddad, Araújo, & Perobelli (2023) and Fernandez-Vazquez & Hewings (2023) suggests that instead of relying on a single parameter established arbitrarily, it is better to rely on a full distribution of the parameters across space and, then, to use averages.

2.1.2.3. Econometric RPCs

The third bottom-up approach with respect to estimating intraregional trade flows that we examine is a location theory-based econometric approach for goods (only) that was originated by Treyz & Stevens (1985) for U.S. states, revisited by Stevens, Treyz, & Lahr (1989), and, more recently, formalized by Lahr, Ferreira, & Többen (LFT, 2020) for the European Union. LFT's quasi-binomial estimates of RPCs are rows-only adjustments to **A** that estimate intraregional trade by industry. That is, LFT estimate RPCs, the regions' propensities to use local production. Leaning on Treyz & 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 SD, SLQ, an older rendition of the FLQ (Flegg & Tohmo, 2013) that does not use CILQs, and Többen & Kronenberg's (2015) CHARM approach.

For the research reported here, we follow closely the approach implemented by LFT (2020) to estimate econometric RPCs for the goods-producing industries. Here we opt for beta binomial regression (Cribari-Neto & Zeileis, 2010) instead of the quasi-binomial regression applied by LFT. While quite similar to the quasi-binomial, the beta binomial estimates parameters requires a maximum likelihood approach as opposed to a generalized linear model

(McCullagh & Nelder, 1989).[14] To obtain RPCs for service industries, we applied the same approach as used in Haddad's SD pool approach (see footnote 9). Appendix B provides further detail on how our econometric RPCs are estimated for goods.

2.2. Interregional trade flows: gravity models and RAS

Lately, at the subnational level, the act of estimating interregional trade flows has been receiving less attention than has estimating 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 near the location where 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, Ramos, & Hewings, 2012, p. 174).

Over time, many approaches have accumulated to estimate interregional trade flows (Sargento, Ramos, & Hewings, 2012). Despite the diversity of their theoretical foundations, however, Batten & Boyce (1986, p. 357) conclude that alternatives "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 review resources and move forward with a focus on approaches that use gravity models, the full family of which has been reviewed by Sen & Smith (1995), and Isard (1998) provides a primer.

The choice of selecting a bottom-up rather than a top-down approach depends on whether regional direct requirements matrices exist for the regions or not. If they do and are deemed good and reasonable, then the gravity model operates on each region's excess supplies and excess demands[15] 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 discussed by Leontief (1953b); Sargento, Ramos, & Hewings (2012); Boero, Edwards, & 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

[14] By limiting estimates of the dependent variable to lie between 0 and 1, these two formulations follow the same restrictions as the functional form employed by Treyz & Stevens (1985) to estimate RPCs.

[15] By "excess" supplies and demands, we mean those local supplies not used locally and those local demands not fulfilled by local supplies.

the commodity.[16] While some survey information of this sort is available in some countries, it is typically tough to apply to regions due to the small, seasonal samples available from surveys.

Consider a pair of origin (o) and destination (d) regions included in a set of geographical units $r = 1, \cdots, g$. We apply a gravity model akin to Isard (1960), Leontief & Strout (1963), Isard (1998), and Yamada (2015). It is:

$$t_i^{od} = k_i \frac{(t_i^{o\bullet})^{\alpha_i} \left(t_i^{\bullet d}\right)^{\beta_i}}{\left(tt_i^{od}\right)^{\gamma_i}} \tag{4}$$

where $t_i^{o\bullet}$ is region o's excess supply, $t_i^{\bullet d}$ is region d's excess demand for i, tt_i^{od} is the shipping cost from o to d 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 parameters for commodity i that researchers typically borrow from prior studies.

Gravity model estimates of trade are often supply-demand balanced using RAS-type methods (Fournier Gabela, 2020). Note, RAS is only one out of many alternatives for matrix balancing (Jackson & Murray, 2004). RAS ensures the consistency of estimates within an MRIO model. But RAS can hardly compensate for bad initial MRIO estimates (Wiebe & Lenzen, 2016).

3. Data and methods

3.1. Two benchmark models based on FIGARO

To test the various approached we decided to use FIGARO, a GMRIO data set focused on Europe (Remond-Tiedrez & Rueda Cantuche, 2019). We contend that it has a good balance of country and industry detail for all European Union (EU-27) countries. Compared with other GMRIOs available, FIGARO also has more-recent data with fully comparable accounts from 2010 through 2020 (Piñero, Mandras, Rueda-Cantuche, & Kutlina-Dimitrova., 2022).[17] We use the 2019 industry-by-industry model data as a benchmark since the 2020 data are undoubtedly tainted by a pandemic slump.[18] Using FIGARO, we produce two models.

---

[16] 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.

[17] We focus on industry-by-industry input-output accounts since the regional data we apply (employment) are available strictly by industry. Indeed, what little data exist are rarely, if ever, available by commodity, which disables proper accounting via Supply and Use tables (SUTs). It is conceivable that some of the approaches that we use in this paper could also be applied to produce subnational SUT estimates, which derive from similar statistical sources.

[18] In Appendix C we summarize the values of intranational and international transactions by EU country in 2019 using data from FIGARO. This data is an aggregate of all national industries. but LFT highlight how manufactured goods are more tradable than services within the EU.

Table 1. Description of the benchmark model with fully specified trade.

	Intermediate transactions			Final demand				
	1	$\cdots$	g	1	$\cdots$	g	$\mathbf{e}$	Σ
1	$\mathbf{T}^{11}$	$\cdots$	$\mathbf{T}^{1g}$	$\mathbf{F}^{11}$	$\cdots$	$\mathbf{F}^{1g}$	$\mathbf{e}^{1}$	$\mathbf{x}^{1\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
g	$\mathbf{T}^{g1}$	$\cdots$	$\mathbf{T}^{gg}$	$\mathbf{F}^{g1}$	$\cdots$	$\mathbf{F}^{gg}$	$\mathbf{e}^{g}$	$\mathbf{x}^{g\bullet}$
$\boldsymbol{m}$	$\mathbf{m}^{\bullet 1}$	$\cdots$	$\mathbf{m}^{\bullet g}$	$\mathbf{f}^{o\not\subset r,1}$	$\cdots$	$\mathbf{f}^{o\not\subset r,g}$	–	$\mathbf{m}^{\bullet\bullet}$
GVA	$\mathbf{w}^{\bullet 1}$	$\cdots$	$\mathbf{w}^{\bullet g}$	–	$\cdots$	–	–	$\mathbf{w}^{\bullet\bullet}$
Σ	$\mathbf{x}^{\bullet 1}$	$\cdots$	$\mathbf{x}^{\bullet g}$	$\mathbf{f}^{\bullet 1}$	$\cdots$	$\mathbf{f}^{\bullet g}$	$\mathbf{e}^{\bullet\bullet}$	

In Table 1, we permit 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 the *commodity* that is shipped and occasionally by the origin or shipping *industry* of exporting establishments. Depictions in GMRIOs of off-diagonal international trade arise as research teams consolidate commodity trade accounts. Rarely are the receiving (or destination) industries recorded.

Table 2. Description of the benchmark model with restricted trade.

	Intermediate transactions			Final demand		
	1	$\cdots$	g	Regional	Foreign	Σ
1	$\mathbf{T}^{11}$	$\cdots$	$\mathbf{T}^{1g}$	$\mathbf{F}^{1}$	$\mathbf{e}^{1}$	$\mathbf{x}^{1\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
g	$\mathbf{T}^{g1}$	$\cdots$	$\mathbf{T}^{gg}$	$\mathbf{F}^{g}$	$\mathbf{e}^{g}$	$\mathbf{x}^{g\bullet}$
$\mathbf{m}$	$\mathbf{m}^{\bullet 1}$	$\cdots$	$\mathbf{m}^{\bullet g}$	$\mathbf{f}^{o\not\subset r}$	–	$\mathbf{m}^{\bullet\bullet}$
GVA	$\mathbf{w}^{\bullet 1}$	$\cdots$	$\mathbf{w}^{\bullet g}$	–	–	$\mathbf{w}^{\bullet\bullet}$
Σ	$\mathbf{x}^{\bullet 1}$	$\cdots$	$\mathbf{x}^{\bullet g}$	$\mathbf{f}^{\bullet\bullet}$	$\mathbf{e}^{\bullet\bullet}$	

Thus, in addition to full interregional trade matrices, we also generate trade estimates that strictly populate the main diagonal of each trade block. Hence, we reproduce the same information required for the Chenery (1953), Moses (1955), and Reifler & Tiebout (1970) models (see Table 2). This fundamentally treats regional inflows of commodities as if they are noncompetitive, i.e., like noncompetitive imports in national accounts. In addition, in this second set of models we consider interregional trade flows of final demand sectors to be consolidated into a single column or sector of interregional exports for each region.

In both models, FIGARO industry-by-industry tables have $n = 64$ industries. The database includes $g = 27$ EU countries. Final demand matrices $\mathbf{F}$ have $l = 5$ columns: (i) household consumption, (ii) collective consumption, (iii) government spending, (iv) gross fixed capital formation and (v) inventory variations. Vector **e** denotes international exports, which are treated as a separate final-demand component. International imports are recorded in vector **m,** including does directly feeding final demand. Value-added is consolidated into a single vector, **w**. Gross output by industry is represented by **x**.

3.2. The gravity model

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.[19] When road transportation is not possible, maritime transportation is the preferred option. In this case, we calculate distances based on each country´s freight port with the greatest throughput volume. 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 & 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. These parameter values are close to those reported 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, Bengoa Calvo, & Sánchez-Robles Rute, 2012). After some minor testing,[20] we decided to let $\alpha = \beta = \gamma = 1$ for all industries.

Following LFT (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 interindustry accounts that depicts interindustry shipments for the entire EU. Using these data, we then fabricate EU MRIO tables in six different ways: the five described in subsection 2.1.1 and an additional, more data intensive, setup. 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 of the five cases—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, Owens, & Thompson, (1982), among others, and assume an internal average travel speed of 90 km/h:

---

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

[20] We tested both 0.5 and 1.0, and the difference was quite small between the two sets of resulting estimates. Indeed, the difference between "actual" and "estimated" was only very slightly lower for all EU countries when $\gamma = 1$.

$$tt^{o=d} = d^{o=d} = 1/90\sqrt{area^{o=d}/\pi} \tag{5}$$

For the remaining alternatives, 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 $o = d$. The four alternative approaches that assume intraregional intermediate flows are known, as previously described, are Haddad's SD, the FLQ & FLQ+, and econometric RPCs (An additional alternative with more "known" information is also examined but described in Section 4, as it is not an equivalent test case). In *Step 2*, we use each region's intermediate industry excess supplies (local 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). After the GRAS adjustment, we enter the intraregional trade estimates into the main diagonal blocks for each region.

4. Hypothesis: More information is better

Ultimately, we examine twelve different sets of subnational MRIO accounts. Recall the set of regionalization methods considered in this research consists of the integrated approach (i.e., a simple gravity model, SGM) and four bottom-up approaches: Haddad's (2014) supply/demand ratio approach (SD), the FLQ and FLQ+, and an econometrically estimated RPC approach (RPC). As briefly alluded to in the last section, we also estimate an MRIO using Riefler and Tiebout's approach, which identify as "HYBRID." Specifically, this approach assumes complete knowledge of regional technology ($\mathbf{T}^{o=d}$), international imports/exports ($\mathbf{m}, \mathbf{e}$), value-added ($\mathbf{w}$), gross output ($\mathbf{x}$) and domestic final demand ($\mathbf{F}^{o=d}$). Note, this set of conditions is almost identical to that confronted by global MRIO modelers who tend to know national IO accounts. Further, each of the six alternatives is implemented with full and restricted (diagonal only) trade specifications (see Tables 1 and 2 above).

Table 3 summarizes the information requirements of each alternative method. We draw a frontier between nonsurvey and hybrid methods depending on whether the regional **A** matrices are required to be known or not. The two versions of the MRIOs for each of the SGM, SD, FLQ,

FLQ+, RPC, and HYBRID approaches differ only in how interregional trade is estimated as explained in subsection 3.1.1.

Table 3. Information requirements for each alternative method.

Method	Code	Information requirements
Non-survey models		
Simple gravity model	SGM	- Regional employment - Regional population
Supply/demand ratio	SD	- National **A** matrix - Regional employment - Regional population
Jahn et al. (2020)	FLQ	- National **A** matrix - Regional employment - Regional population
Econometric RPC	RPC	- Actual RPCs by goods-producing industries for each region - National **A** matrix - Regional employment - Regional population - Additional regional information (e.g., region area, value/weight...)
Flegg et al. (2021)	FLQ+	- Regional **A** matrices → To estimate optimum δ - National **A** matrix - Regional employment - Regional population
Hybrid model		
Riefler-Tiebout (1970)	HYBRID	- Regional **A** matrices → Directly embedded in the model - Regional arrays for $\mathbf{e}$, $\mathbf{m}$, $\mathbf{w}$ and $\mathbf{x}$ - Regional domestic final demand

Our hypotheses of the present research are grounded in the intensity of use of primary data. That is, we hypothesize that as more primary data are applied, the resulting estimated arrays should be better at replicating their actual equivalents. This follows from Lynch (1979) and Szyrmer (1989) who, when using RAS, find that error inherent in estimated IO tables declines as those tables include more "known" information.[21] This is because within RAS algorithms "unknown" values within the tables are increasingly constrained as other information becomes "known." In this vein, we expect the most naïve approaches—those estimating diagonal interregional trade flows, those not adjusted for regional industry value added, and, particularly, SGM—will be associated with more error. And, hence, that those regionalization approaches that estimate full arrays of interregional trade flows, that do adjust for regional industry value added and that do use more regional data to alter intraregional intermediate transactions—FLQ,

[21] Batten & Martellato (1985) observe similarly, but it is unclear that they apply RAS.

FLQ+, and econometric RPCs—should perform better. Finally, the Riefler-Tiebout (HYBRID) model, which assumes direct requirements are known, should test out as best.

5. Test measures

We test our hypotheses using these six approaches from four different perspectives: output multipliers, value-added multipliers, the direct requirements matrix (**A**) and the Leontief inverse (**L**). That is, we examine both the holistic and partitive accuracy of the tables as suggested by Jensen (1980).[22] The multipliers relate to holistic accuracy—the degree to which estimated tables reproduce what is generally deemed the most important practical features of their actual counterparts. The two latter address the partitive accuracy of the tables—accuracy regarding the way tables display detailed interindustry interactions.

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 mean absolute percentage error (WMAPE, also denoted as WAPE in the literature).[23] We only report two here—MAPE and WMAPE. See section 3 in Temursho, Webb & Yamano (2011) for a description. 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 mimics rankings across countries and across industries, as well as 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 output multipliers, value-added multipliers, **A,** and **L** are cell specific. That is, each cell is divided by the total sum of each array. In this regard, differences between MAPE and WMAPE for multipliers should be minimal. Note, however, **A** and **L** are column-specific, i.e., they are calculated only for the columns of the arrays pertaining to the country or industry identified for testing. In summary, WMAPE weights larger elements more heavily relative to smaller elements. Emphasizing accuracy of larger elements in estimates of both **A** and **L** arrays is

[22] Given the set of countries and industries included in the database a more comprehensive analysis of the analytically relevant multipliers for each table was impractical. In any case, the simple distinction by industry and by country facilitates a summary view of differences across the approaches.

[23] There are many measures that can be used to compare arrays--see the appendix of Lahr (2001b) for a good number of alternatives. As pointed out by a referee, one is the weighted relative percentage difference (Arto, Rueda-Cantuche, & Peters, 2014). To date, most of measures used to compare IO tables are suited to comparisons of vectors and not for the interplay between technology, productivity, and trade in an IO matrix. Hence, it may well be that Owen's (2017) use of structural decomposition approaches (as suggested by Lahr & Yang, 2011) is best for tests that compare IO tables.

critical, according to Jensen & West (1980). MAPE, on the other hand, gives equal importance to the elements it measures. So, for estimates from a given modelling approach, obtaining a MAPE that is higher in value than its WMAPE equivalent implies the approach tends to reproduce larger elements with greater accuracy than it does smaller elements. In such cases, i.e., when MAPE> WMAPE, the approach conforms with principles suggested by Jensen & West (1980).

6. Empirical Outcomes

Output multipliers provide a comprehensive assessment of each modelling alternative since they are based on both intra and interregional flows. Tables 4 and 5 summarize our results measured by MAPE and WMAPE for each alternative method by country and industry. As predicted by theory, the most data-intensive alternative (HYBRID) yields, on average as well as for nearly all EU member countries, substantially better results. In contrast, the least data-intensive method—the simple gravity model—tends to be least accurate. Generally speaking, however, differences in the average error, both overall and by country, are not large. The gap between the average WMAPE best and worst-performing alternatives, excluding HYBRID, is less than 2.5%.

Among the nonsurvey methods, a clear ranking emerges in their relative accuracy, despite small differences across them in average error levels as identified via WMAPEs. From best- to worst-performing, it runs: FLQ+, the econometric RPC, S/D, FLQ, to SGM. Still, the small difference across all of the approaches suggests, when constructing an MRIO, that it is sufficient to employ a minimum of data (employment by region-industry and regions' populations shares) if the application strictly demands holistic accuracy.

On other economic measures (value-added multipliers, **A**, and **L**) a similar general picture emerges. That is, simple gravity models (SGM) perform worst, and Riefler-Tiebout models (HYBRID) perform best (see Table 6). Moreover, value-added multipliers are more accurate than are output multipliers across all MRIO forms. Also, in the case of HYBRID, regional direct requirements are known, so error in the intraregional **A** matrices is minimized. This accuracy then spills over to characteristics of the Leontief inverses, the **L**s. Different from what we observe for output multipliers, however, results vary more across the nonsurvey methods although the ranking among them remains. That is, on average the FLQ+ outperforms other nonsurvey techniques for value-added multipliers, followed by the econometric RPC, SD, FLQ, to SGM.

Table 4. The relative accuracy of output multipliers (minus 1.0) by country

(bold-faced font denotes industry's lowest-valued MAPE or WMAPE)

	SGM		S/D		FLQ		RPC		FLQ+		HYBRID	
	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE
AT	29.39	24.97	28.60	25.68	28.67	24.71	28.11	25.32	25.89	22.92	**19.02**	**17.09**
BE	22.10	16.95	24.21	20.32	22.54	17.84	23.79	20.01	22.64	18.57	**17.24**	**15.13**
BG	19.75	17.61	23.60	20.41	20.42	18.00	22.86	19.80	20.63	18.08	**16.11**	**15.91**
CY	36.88	35.85	27.80	22.61	31.82	28.84	29.03	23.24	29.41	25.12	**17.34**	**18.82**
CZ	24.91	21.39	24.36	21.93	21.58	19.05	23.56	21.28	20.26	18.49	**15.11**	**12.98**
DE	19.69	17.56	18.73	17.54	20.42	19.27	19.04	17.94	20.78	19.83	**13.13**	**11.92**
DK	24.76	21.36	26.90	23.29	22.61	19.18	26.02	22.43	21.56	18.32	**17.63**	**17.12**
EE	22.42	22.33	19.42	**18.33**	21.50	21.25	19.58	18.67	21.25	20.96	22.65	25.41
ES	22.55	17.08	22.34	16.24	21.72	16.27	22.33	15.78	21.70	15.33	**13.77**	**12.23**
FI	54.91	56.49	32.92	33.76	45.62	46.93	33.54	34.37	39.66	40.34	**18.65**	**17.90**
FR	17.68	14.18	16.19	13.50	15.75	13.29	15.75	13.16	15.52	13.21	**13.61**	**10.69**
GR	39.06	40.31	29.87	23.11	29.17	25.32	29.84	23.03	27.14	20.79	**15.16**	**13.09**
HR	25.78	20.40	28.57	22.95	26.54	20.95	27.50	21.96	**25.46**	20.13	161.65	**12.93**
HU	31.66	24.68	31.81	25.65	31.17	24.50	30.94	24.88	28.76	**22.84**	**21.48**	23.58
IE	64.50	35.46	85.03	43.93	77.33	38.27	83.60	43.80	83.40	40.98	**17.33**	**12.39**
IT	21.65	19.60	19.68	16.77	19.97	17.37	19.67	16.54	18.99	15.72	**12.53**	**11.27**
LT	30.52	27.01	37.50	33.32	30.89	27.42	36.40	32.26	30.49	27.20	**15.07**	**16.56**
LU	27.45	22.63	27.48	22.75	27.88	22.72	26.98	22.43	26.96	22.21	**16.08**	**15.99**
LV	21.17	19.33	23.22	19.72	21.22	19.16	22.34	19.19	21.03	19.10	**17.00**	**18.37**
MT	**23.09**	**26.04**	25.90	29.21	32.20	34.73	26.26	29.57	30.82	33.38	25.86	28.22
NL	43.36	36.82	45.69	40.33	42.99	37.29	45.28	40.06	42.46	37.27	**18.01**	**15.50**
PL	18.96	15.86	18.70	15.17	17.79	14.59	18.34	14.93	17.48	14.31	**12.99**	**12.41**
PT	20.70	19.57	20.29	16.57	19.51	17.15	20.21	16.65	19.36	**16.21**	**18.34**	19.58
RO	19.13	18.51	17.29	15.89	17.85	16.69	17.54	16.27	17.31	15.90	**12.42**	**13.61**
SE	25.75	25.79	19.92	18.79	22.81	22.48	19.73	18.52	21.26	20.56	**13.88**	**15.27**
SI	32.55	26.60	33.89	28.49	31.91	25.52	32.67	27.44	29.88	24.45	**19.17**	**18.32**
SK	27.67	23.89	26.06	23.13	26.46	22.84	25.11	22.30	23.83	20.97	**17.81**	**16.64**
mean	28.45	24.75	28.00	23.31	27.72	23.39	27.63	23.03	26.81	22.34	**22.19**	**16.26**

Table 5. The relative accuracy of output multipliers (minus 1.0) by industry

(bold-faced font denotes industry's lowest-valued MAPE or WMAPE)

	SGM		SD		FLQ		RPC		FLQ+		HYBRID	
Industry	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE	MAPE	WMAPE
A01	18.46	0.74	15.95	0.63	15.90	0.62	15.30	0.60	17.97	0.71	**12.15**	**0.47**
A02	45.42	1.20	49.26	1.20	52.80	1.22	51.25	1.23	48.14	1.24	**12.79**	**0.37**
A03	21.36	0.74	20.89	0.68	21.10	0.69	22.25	0.73	20.76	0.69	**8.56**	**0.32**
B	28.91	1.02	27.96	0.95	27.52	0.94	25.98	0.89	26.32	0.93	**9.87**	**0.35**
C10T12	15.15	0.57	13.79	0.51	14.00	0.51	**12.66**	**0.46**	14.93	0.55	17.42	0.65
C13T15	28.64	1.01	26.96	0.96	26.18	0.94	23.67	0.83	25.32	0.90	**21.67**	**0.80**
C16	16.52	0.62	16.25	0.60	15.74	0.58	**14.28**	**0.53**	15.38	0.58	15.75	0.57
C17	**16.85**	0.58	19.46	0.65	19.02	0.63	17.81	**0.57**	18.23	0.62	26.67	0.99
C18	15.19	0.56	13.59	0.50	13.35	0.49	**12.99**	**0.47**	15.13	0.56	21.75	0.80
C19	64.88	2.03	73.18	2.19	72.73	2.18	69.43	2.07	64.96	1.97	**18.42**	**0.73**
C20	21.14	0.72	19.78	0.64	20.03	0.65	18.35	0.56	20.49	0.66	**29.88**	**1.12**
C21	31.54	1.06	30.57	1.02	29.71	1.01	27.86	0.92	29.37	1.00	**19.17**	**0.71**
C22	16.77	0.60	17.06	0.62	16.52	0.60	**12.98**	**0.47**	16.23	0.58	30.96	1.15
C23	19.32	0.70	19.07	0.69	18.64	0.68	16.10	**0.58**	19.12	0.71	**15.94**	0.60

C24	**17.58**	**0.59**	19.55	0.65	18.79	0.61	21.51	0.62	20.64	0.61	29.56	1.11
C25	16.67	0.61	16.50	0.59	16.29	0.59	**13.63**	**0.49**	16.40	0.60	24.40	0.90
C26	34.40	1.11	31.47	1.00	31.29	1.00	29.14	**0.89**	31.35	1.01	**26.99**	1.07
C27	21.66	0.77	19.53	0.69	19.40	0.69	**16.44**	**0.57**	19.55	0.70	29.06	1.08
C28	18.19	0.63	17.19	0.60	16.87	0.59	**13.80**	**0.48**	15.32	0.55	25.85	0.95
C29	21.61	0.70	17.44	0.56	16.98	0.55	**13.80**	**0.46**	14.40	0.49	35.36	1.35
C30	57.67	0.79	74.53	0.78	71.06	0.78	89.03	0.81	76.11	**0.77**	**29.73**	0.86
C31_32	22.12	0.80	19.71	0.72	19.58	0.71	**15.19**	**0.55**	19.08	0.69	**19.86**	**0.73**
C33	17.15	0.60	19.84	0.69	18.90	0.66	16.93	0.60	17.47	0.62	**15.26**	**0.55**
D35	30.45	1.09	37.37	1.22	36.92	1.21	32.58	1.12	30.83	1.09	**9.46**	**0.34**
E36	33.33	1.07	36.63	1.10	35.92	1.08	33.45	0.99	34.41	1.09	**8.83**	**0.30**
E37T39	19.89	0.73	19.49	0.70	18.94	0.69	16.80	0.62	19.69	0.73	**7.94**	**0.30**
F	17.87	0.66	16.04	0.58	15.93	0.57	14.69	0.54	17.87	0.66	**10.65**	**0.39**
G45	19.95	0.70	17.75	0.61	17.23	0.60	16.26	0.57	19.40	0.68	**11.12**	**0.41**
G46	19.85	0.73	20.57	0.73	20.10	0.71	16.49	0.58	19.84	0.73	**7.61**	**0.28**
G47	21.62	0.77	19.40	0.68	18.98	0.66	15.34	0.54	19.97	0.71	**11.35**	**0.41**
H49	18.41	0.68	17.24	0.63	16.82	0.62	14.60	0.53	17.62	0.66	**8.67**	**0.32**
H50	50.11	1.61	44.53	1.28	44.26	1.28	38.42	1.09	39.58	1.16	**15.16**	**0.56**
H51	20.51	0.71	18.60	0.63	18.50	0.63	**17.13**	**0.58**	18.40	0.62	18.42	0.66
H52	23.95	0.89	25.26	0.90	24.79	0.88	22.07	0.79	24.47	0.90	**8.12**	**0.30**
H53	33.79	1.02	37.18	1.10	36.33	1.08	31.47	0.91	33.76	1.02	**10.99**	**0.38**
I	21.65	0.79	20.08	0.70	19.66	0.68	18.71	0.64	21.88	0.78	**8.07**	**0.31**
J58	23.50	0.81	20.24	0.68	19.95	0.67	18.23	0.62	21.06	0.73	**10.82**	**0.39**
J59_60	26.77	0.90	29.47	0.96	28.82	0.94	24.53	0.81	26.83	0.91	**8.09**	**0.30**
J61	28.94	0.96	29.33	0.95	28.63	0.93	22.89	0.74	26.80	0.89	**8.65**	**0.31**
J62_63	37.33	1.26	35.81	1.20	34.93	1.17	29.68	1.02	35.17	1.21	**13.13**	**0.51**
K64	48.69	1.67	49.10	1.65	48.06	1.62	43.88	1.50	47.72	1.65	**19.95**	**0.78**
K65	32.47	1.08	31.84	1.05	31.08	1.03	27.74	0.93	29.74	0.99	**10.18**	**0.37**
K66	37.63	1.09	39.53	1.09	38.78	1.07	37.55	1.05	38.21	1.06	**15.11**	**0.49**
L68	43.21	1.33	45.87	1.26	45.78	1.27	48.32	1.33	45.85	1.35	**34.35**	**0.99**
M69_70	34.10	1.14	35.62	1.14	34.88	1.12	30.60	0.98	32.81	1.09	**12.32**	**0.45**
M71	21.75	0.75	21.06	0.69	20.78	0.69	18.78	0.62	20.98	0.72	**7.67**	**0.29**
M72	37.03	1.22	40.30	1.20	39.28	1.17	41.54	1.18	41.96	1.27	**12.71**	**0.43**
M73	23.61	0.77	20.87	0.67	20.72	0.66	19.54	0.62	22.70	0.75	**8.32**	**0.30**
M74_75	31.88	0.98	30.17	0.91	29.91	0.91	27.65	0.85	30.67	0.95	**8.90**	**0.31**
N77	30.76	0.99	28.01	0.87	27.84	0.87	25.47	0.79	29.52	0.94	**11.44**	**0.39**
N78	56.38	1.94	53.16	1.81	52.83	1.81	51.92	1.79	55.82	1.90	**46.22**	**1.15**
N79	42.06	1.26	46.83	1.27	46.55	1.27	43.23	1.21	42.04	1.24	**11.26**	**0.38**
N80T82	23.05	0.78	22.44	0.74	22.26	0.73	21.65	0.71	21.68	0.74	**8.32**	**0.30**
O84	26.26	1.01	23.32	0.86	22.78	0.85	**18.86**	**0.71**	24.23	0.93	20.99	0.73
P85	35.86	1.26	28.17	0.94	28.02	0.93	**27.74**	**0.93**	33.88	1.17	42.98	1.45
Q86	29.70	1.05	24.29	0.86	24.27	0.86	21.15	0.75	27.70	0.98	**13.57**	**0.47**
Q87_88	31.40	1.14	23.57	0.86	23.44	0.86	19.88	0.74	25.36	0.93	**18.06**	**0.59**
R90T92	50.58	1.19	50.59	1.06	50.24	1.05	46.29	0.95	49.54	1.14	**17.37**	**0.48**
R93	23.61	0.81	19.55	0.66	19.70	0.66	19.56	0.67	22.58	0.78	**8.44**	**0.30**
S94	35.17	1.23	31.19	1.09	31.27	1.10	31.84	1.14	34.37	1.21	**11.16**	**0.34**
S95	27.07	0.94	23.27	0.81	22.95	0.80	20.80	0.73	25.83	0.90	**11.42**	**0.43**
S96	26.32	1.01	22.29	0.85	21.74	0.83	18.63	0.73	24.16	0.93	**16.57**	**0.62**
T	**6.36**	3.64	6.37	3.64	6.38	3.64	6.44	**3.64**	6.41	3.64	353.09	19.72
U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mean	27.97	0.97	27.53	0.92	27.17	0.91	25.36	0.84	27.25	0.93	**16.61***	**0.59**

Notes: (1) Industry **T** is excluded from MAPE and WMAPE means calculations because small cells cause perversely high error levels.
(2) The set of dashes across the table delineates goods from services.

Table 6. Mean error by subnational MRIO table production approach

(bold-faced font denotes lowest-valued MAPE or WMAPE in the column)

Approach | Measure		Output multipliers		Value-added multipliers*		Direct requirements		Leontief inverse	
		Full	Diagonal	Full	Diagonal	Full	Diagonal	Full	Diagonal
SGM	MAPE	28.45	65.53	30.70	40.01	411.40	660.12	155.98	>9,999
	WMAPE	24.75	55.08	21.69	32.65	94.73	97.07	33.04	33.35
SD	MAPE	28.00	54.74	57.09	56.12	1,504.00	1,503.99	1,303.71	>9,999
	WMAPE	23.31	42.26	19.92	23.67	69.47	69.47	22.27	22.26
FLQ	MAPE	27.72	45.25	76.63	71.91	907.13	907.13	438.15	>9,999
	WMAPE	23.39	37.02	19.82	25.30	71.66	71.66	25.19	25.17
RPC	MAPE	27.63	50.33	55.95	55.50	1,462.27	1,462.27	1,007.61	>9,999
	WMAPE	23.03	37.73	19.63	23.24	67.60	67.60	21.54	21.57
FLQ+	MAPE	26.81	36.03	75.12	66.73	1,324.87	1,324.87	719.80	>9,999
	WMAPE	22.34	29.30	18.61	22.64	68.45	68.45	22.24	22.24
HYBRID	MAPE	**22.19**	**9.34**	**13.16**	**4.47**	0.00	0.02	**2.71**	**1.74**
	WMAPE	**16.26**	**9.24**	**6.89**	**4.11**	0.00	0.02	**0.13**	**0.06**

Note: Bulgaria and Cyprus are excluded from MAPE means' calculations because their large number of very small elements yield disturbingly high MAPE values.

For intraregional flows (direct requirements, **A**), the ranking changes, and differences become less negligible. Gauged by WMAPEs, the econometric RPC marginally, but consistently, outperforms the FLQ+ on average; and the simple gravity model is outperformed by all other approaches by nearly 25 percentage points. This gap suggests that MRIOs produced via the SGM approach that we apply here are liable to be less desirable (than MRIOs produced via other approaches) for research in which technology or input mix is a focus—supply-chain and structural decomposition analyses, for example.

Econometrically estimated RPC models present the lowest WMAPEs for the Leontief inverse **L** matrices as well, again marginally but consistently outperforming equivalents created via an FLQ+, SD, or FLQ approach. Note, however that in moving from an examination of direct requirements to one of the Leontief inverse (total requirements) that the error gap inherent to the SGM when compared to the other approaches reduces to 10 percentage points; While the gap is somewhat narrower, it remains substantial, albeit one that is less alarming.

When interregional trade is expressed strictly via diagonal elements—a reality rather than a mere practicality—the ability of estimates to replicate the actuals worsens (except for those emanating from HYBRID), as might be expected. The HYBRID exception comes with no surprise. When we restrict trade in final demand, we consolidate interregional exports into an additional column in **F**. This column would be known alongside international exports, provided a regional IO model exists.[24] Imports to final demand ($\mathbf{f}^{o\not\subset r}$) are normally included in regional IO

[24] For instance, all EU national IO tables break imports and exports into EU and non-EU trade; Spanish regional IO tables include imports/exports to the rest of Spain, to rest of the EU, and to rest of the world.

models as well. The additional constraint, thus, explains the better performance of HYBRID models. Of course, "knowing" the regional **A** matrices is the overriding factor in the superior performance of HYBRID compared to the other approaches.

7. Conclusions

Multiregional input-output tables are valuable for understanding how impacts of economic shocks transmit across regions. In the absence of official data on intra- and inter-regional transshipments, it is critical to know how choices among different methods options affect the accuracy of estimating such impacts. Our results for the HYBRID approach suggest that estimates of multiregional accounts greatly benefit when generated via surveys that are statistically representative.

More generally, we analyze the relative accuracy of techniques within the literature that estimates subnational MRIO models. We do so by observing how well different fundamental approaches can replicate intra- and inter-national interindustry trade flows across the EU-27. The simplest of them (the simple gravity model, SGM) uses only employment information by industry in each nation's estimates of shipping distances within a region and between region pairs, and EU population shares that are used to estimate final demands. 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 (SD) pool approach, two versions of the FLQ approach (Flegg et al., 2021), and an econometrically estimated RPC approach that employs a binomial beta regression. 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-industry by 64-industry meta-account. This account effectively forms the "national" transactions matrix, a fundamental requirement for producing subnational IO tables: one that assumes knowledge transfers freely across geographic space.

We compare all ten alternatives to the FIGARO GMRIO accounts for 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 in estimating multipliers, despite its very limited data requirements. That said, it also performs rather poorly in replicating direct requirements matrices and, hence, Leontief inverses.

The level of error suggests limited use of such a model, i.e., it is not preferred for applications that examine technology or industry mix.

From the same two perspectives—direct requirements matrices and Leontief inverses—the econometric RPC approach tends to perform best. But this approach also owns the steepest set of data requirements. In general, there are only small differences among the other approaches examined herein—SD, FLQ, FLQ+, and the econometric RPC— in their ability to replicate "known" MRIO accounts and multipliers. Clearly, as judged by the excellent performance of the HYBRID option, it is key to know regions' intra-regional shipments by or at least for as many industries and regions as possible, which comports with findings of Lahr (2001b) and Temurshoev (2010). Thus, we confirm our hypothesis that the application of more "knowable" data improves the accuracy of subnational MRIO tables, although the added improvement appears to be almost negligible, particularly in the case of nonsurvey methods. This suggests that constructors of MRIO accounts can afford to be honest when articulating interregional trade across sectors, which is generally very data-limited. Additionally, the performance of the HYBRID method shows that the accuracy of regional input-output tables and the estimations produced by those who use them would greatly benefit if regional statistical offices had the capacity to produce their own regional tables.

Limiting interregional trade to the diagonal of off-diagonal partitions adds error. What is a surprise, however, is how little error it induces into the models. 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 populated interregional trade matrices result from industry aggregation of import accounts for which only the commodity code of the imported shipment is known.

This current paper is a first stab at analyzing ways of producing subnational MRIO tables. More needs to be done. We assume that, for industries, MRIO tables constructors only "know" job counts for each subnational region. We also assume they have reasonable estimates of population shares across the relevant subnational regions. In three cases, we also further assume that table constructors can produce or have already produced regionalized intraregional IO tables. But we did not go so far as to assume that any components of regional value added or regional final demand by industry are knowable *a priori*. We also do not assume table constructors can "know" the size or destinations of regional industry outflows of commodities. Such data are, in fact, published by the EU for its NUTS2 regions and by the U.S. Bureau of Economic Analysis for U.S. states. Our future near-term plan is to perform experiments on subnational MRIOs that also incorporate such data, believing that what we have demonstrated herein is a strong start.

We note that, using the same data that we have applied here, researchers could apply more-sophisticated estimating techniques—e.g., doubly constrained gravity models and mathematical programming approaches, which could yield MRIO estimates that are even more accurate. As we noted at the start, we purposely limited our set of tests to approaches that use simple algorithms. Again, as with added data, so many alternatives exist that just one or two papers focused on such tests would be inadequate to cover the full spectrum of possibilities. It would be good to test approaches that do not use RAS at all. Also, while it undoubtedly can heal all wounds, RAS also can enable error to exaggerate and fester.

REFERENCES

Arto, Iñaki, José M. Rueda-Cantuche & Glen P. Peters (2014). Comparing the GTAP-MRIO and WIOD databases for carbon footprint analysis, *Economic Systems Research*, 26, 3, 327–353. https://doi.org/10.1080/09535314.2014.939949

Bachmann, Chris, Matthew J. Roorda, & Chris Kennedy. (2015). Developing a multi-scale multi-region input–output model. *Economic Systems Research* 27, 172–193. https://doi.org/10.1080/09535314.2014.987730

Batten, David F., & David E. Boyce. (1986). Spatial interaction, transportation, and interregional commodity flow models. In P. Nijkamp (Ed.), *Handbook of Regional and Urban Economics: Vol. I* (pp. 357–406). Elsevier Science Publisher. https://doi.org/10.1016/S1574-0080(00)80012-7

Batten, David F., & Dino Martellato. (1985). Classical versus modern approaches to interregional input-output analysis. *Annals of Regional Science*, 19(3), 1–15. https://doi.org/10.1007/BF01294827

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. https://doi.org/10.1016/j.jinteco.2020.103383

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. https://doi.org/10.1007/s00181-019-01727-z

Beyers, William B. (1980). Migration and the development of multiregional economic systems. *Economic Geography*, 56, 320-334. https://doi.org/10.2307/143329

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

Boomsma, Piet & Jan Oosterhaven. (1992). A double-entry method for the construction of bi-regional input-output tables, *Journal of Regional Science*, 32, 269–284. https://doi.org/10.1111/j.1467-9787.1992.tb00186.x

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. https://doi.org/10.1086/694292

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. https://doi.org/10.18637/jss.v034.i02

Dalgaard, Ebsen & Christian Gysting. (2004). An algorithm for balancing commodity-flow systems. *Economic Systems Research*, 16, 169–190. https://doi.org/10.1080/0953531042000219295

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. Available online in March 2024 at https://iioa.org/conferences/29th/papers/files/4669.pdf

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/

Fernandez-Vazquez, Esteban & Geoffrey J.D. Hewings (2023) No more gambling: Estimating non-survey regional and interregional input-output tables by averaging location quotients. An unpublished paper submitted for presentation at the 70th Annual North American Meetings of the Regional Science Association International, San Diego, USA, November 16.

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. https://doi.org/10.1186/s40008-021-00242-8

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. https://doi.org/10.1007/978-981-13-1507-7_13

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. https://doi.org/10.1111/j.1467-9787.1987.tb01157.x

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

Haddad, Eduardo, El-Hattab, Fatna., & Ail, Abdelaaziz A. (2017). *A Practitioner's Guide for Building the Interregional Input-Output System for Morocco, 2013* (17/02; OCP Policy Center Research Papers). Available oonlijne in March 2024 at https://www.policycenter.ma/sites/default/files/2021-01/RP%20-%2017-02%20%28Abdelaziz%29%20v2.pdf

Haddad, Eduardo A., Inácio F. Araújo, & Fernando S. Perobelli. (2024) Geographical proximity and technological similarity. *Structural Change and Economic Dynamics*, 70, 302-318. https://doi.org/10.1016/j.strueco.2024.02.007.

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

Hulu, Edison, & Geoffrey Hewings (1993). The development and use of interregional input-output models for Indonesia under conditions of limited information. *Review of Urban & Regional Development Studies*, 5(2), 135–153. https://doi.org/10.1111/j.1467-940X.1993.tb00127.x

Inomata, Satoshi, Norihiko Yamano, & Bo Meng. (2013) Developing international input–output databases: IDE-JETRO and OECD Experiences. Chapter 9 in Aaditya Mattoo, Zhi Wang, & Shang-Jin Wei (eds.), *Trade in value added: Developing new measures of cross-border trade.* Washington, D.C.: The International Bank for Reconstruction and Development, The World Bank, pp. 221–251.

Isard, Walter. (1951). Interregional and regional input-output analysis: a model of a space economy. *Review of Economic Statistics*, 33, 318–329. https://doi.org/10.2307/1926459

_____. (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.

_____. 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, Massachusetts: The M.I.T Press.

_____. (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, Vermont, Ashgate Publishing, pp. 243–280.

Isard, Walter, & Robert E. Kuenne. (1953). The Impact of Steel Upon the Greater New York-Philadelphia Industrial Region. *The Review of Economics and Statistics*, 35(4), 289–301. https://doi.org/https://doi.org/10.2307/1924389

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. https://doi.org/10.1111/j.1435-5597.1992.tb01750.x

Jackson, Randall W., & Alan T. Murray. (2004). Alternative input-output matrix updating formulations. *Economic Systems Research*, 16(2), 135–148. https://doi.org/10.1080/0953531042000219268

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

Jensen, Rodney C. (1980). The concept of accuracy in regional input-output models. *International Regional Science Review*, 5(2), 139–154. https://doi.org/10.1177/016001768000500203

Jensen, Rodney C. & Guy R. West. (1980). The effect of relative coefficient size on input-output multipliers. *Environment and Planning A,* 12, 659–670. https://doi.org/10.1068/a12065

Junius, Theo, & Jan Oosterhaven. (2003). The solution of updating or regionalizing a matrix with both positive and negative entries. *Economic Systems Research*, 15(1), 87–96. https://doi.org/10.1080/0953531032000056954

Keeble, David, Peter L. Owens, & Chris Thompson. (1982). Regional accessibility and economic potential in the European Community. *Regional Studies* 16, 419–432. https://doi.org/10.1080/09595238200185421

Lahr, Michael L. (2001a). Reconciling domestication techniques, the notion of re-exports, and some comments on regional accounting. *Economic Systems Research*, 13, 165–179. https://doi.org/10.1080/09537320120052443

_____. (2001b). A strategy for producing hybrid regional input-output tables. In Michael L. Lahr and Erik Dietzenbacher (eds.) *Input-output analysis: frontiers and extensions*. New York: Palgrave, pp. 211–244.

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

Lahr, Michael L. & Ling Yang. (2011). Comparing input-output tables of different economies: A new application for decomposition techniques? Presentation at the 19th International Input-Output Conference, Alexandria, Virginia, USA, June 14-17.

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/

Leontief, Wassily. (1936). Quantitative Input–Output Relations in the Economic System of the United States, *Review of Economics and Statistics*, 18, 105–125. https://doi.org/10.2307/1927837

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

_____. (1953a). Domestic production and foreign trade: the American capital position re-examined. *Proceedings of the American Philosophical Society*, 97, 332–349. https://www.jstor.org/stable/3149288

_____. (1953b). 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.

_____. (1974). Structure of the world economy: Outline of a simple input-output formulation (Nobel Memorial Lecture, December 11, 1973), *Swedish Journal of Economics,* 76, 387–401. https://www.jstor.org/stable/1815236

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

López-Álvarez, Jorge M. (2023). FIGARO-REG: A new dimension to measure bilateral regional exposure to value added. Presentation at the 14th Geoffrey J.D. Hewings Regional Economics Workshop, WIFO (Austrian Institute of Economic Research, Vienna, Austria, October 2-3, 2023.

Lynch, Robert G. (1986). An assessment of the RAS method for updating input-output tables. In Ira Sohn (ed), *Readings in Input-Output Analysis: Theory and Applications,* New York: Oxford University Press, pp. 271–284.

Malizia, Emil & Daniel L. Bond. (1974). Empirical tests of the RAS method of interindustry coefficient adjustment. *Journal of Regional Science,* 14, 355–365. https://doi.org/10.1111/j.1467-9787.1974.tb00458.x

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. https://www.redalyc.org/

McCullagh, Peter., & John A. Nelder, (1989). *Generalized Linear Models*. Routledge. https://doi.org/10.1201/9780203753736

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. https://doi.org/10.1111/j.1467-9787.1974.tb00442.x

Meadows, Donella and Dennis Meadows. (1972). *Limits to growth*. New York: Signet Press.

Miller, Ronald E. (1957). The Impact of the Aluminium Industry on the Pacific Northwest: A Regional Input-Output Analysis. *The Review of Economics and Statistics*, 39(2), 200–209. https://doi.org/10.2307/1928537

_____. (1966). Interregional Feedback Effects in Input-Output Models: Some Preliminary Results. *Papers of the Regional Science Association*, 17, 105–125. https://doi.org/10.1111/j.1435-5597.1966.tb01344.x

_____. (1969). Interregional Feedbacks in Input-Output Models: Some Experimental Results. *Western Economic Journal*, 7, 41–50.

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

Monfort, P. (2020). *Convergence of EU regions redux. Recent Trends in Regional Disparities* (02/2020; European Commission Working Papers).

Moses, Leon. (1955). The stability of interregional trading patterns and input-output analysis. *American Economic Review*, 45, 803–832. https://www.jstor.org/stable/1821380

Oosterhaven, Jan. (1981). *Interregional input-output analysis and Dutch regional policy problems.* Aldershot, U.K.: Gower Publishing Company, Ltd.

Owen, Anne. (2017). *Techniques for evaluating the differences in multiregional input-output databases: A comparative evaluation of CO_2 consumption-based accounts calculated using Eora, GTAP and WIOD*. Dordrecht, Germany: Springer International Publishing. https://doi.org/10.1007/978-3-319-51556-4

Pakizeh, Amir Hossein & Hamed Kashani. (2022). Application of machine-learning models to estimate regional input coefficients and multipliers. *Spatial Economic Analysis*, 17, 178–205. https://doi.org/10.1080/17421772.2021.1959046

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. Abstract available at: https://www.iioa.org/conferences/28th/

Planting, Mark & Jiemin Guo. (2004). Increasing the timeliness of US annual input-output accounts. *Economic Systems Research,* 16, 157–167. https://doi.org/10.1080/0953531042000219286

Polenske, Karen R. (1970). *A multiregional input-output model for the United States*. Report No 21 prepared for the Economic Development Administration, U.S. Department of Commerce under contract #7-35212. December. Available online in March 2024 at: https://researchrepository.wvu.edu/

_____. (1972). The implementation of a multiregional input-output model for the United States, in Andras Brody and Anne P. Carter (eds.) *Input-output techniques*, Amsterdam: North Holland Publishing Co., pp. 171–189.

_____. (1980). *U.S. multiregional input-output accounts and model*. Lexington, Massachusetts: D.C. Heath and Company.

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. https://doi.org/10.1111/j.1467-9787.1970.tb00042.x

Round, Jeffrey I. (1978). An Interregional Input-Output Approach to the Evaluation of Nonsurvey Methods. *Journal of Regional Science*, 18, 179–194. https://doi.org/10.1111/j.1467-9787.1978.tb00540.x

_____. (1979). Compensating Feedback Effects in Interregional Input-Output Models. *Journal of Regional Science*, 19, 145–155. https://doi.org/10.1111/j.1467-9787.1979.tb00581.x

_____. (1983). Nonsurvey Techniques: A Critical Review of the Theory and the Evidence. *International Regional Science Review*, 8, 189–212. https://doi.org/10.1177/016001768300800302

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/

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.

Stevens, B. H., George I. Treyz, & Michael L. Lahr. (1989). On the comparative accuracy of RPC estimating techniques. Chapter 18 in Ronald E. Miller, Karen R. Polenske, & Adam Z. Rose (Eds.), *Frontiers of input-output analysis.* New York: Oxford University Press, pp. 245–257.

Szyrmer, Janusz M. (1989). Trade-off between error and information in the RAS procedure. Chapter 19 in Ronald E. Miller, Karen R. Polenske, & Adam Z. Rose, *Frontiers of input-output analysis*, New York: Oxford University Press, pp. 255–278.

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. https://doi.org/10.1111/roie.12178

Temurshoev, Umed. (2010). Identifying optimal sector groupings with the hypothetical extraction method. *Journal of Regional Science*, 50, 872–890. https://doi.org/10.1111/j.1467-9787.2010.00678.x

Temurshoev, Umed, Colin Webb, & Norihiko Yamano. (2011). Projection of supply and use tables: Methods and their empirical assessment. *Economic Systems Research*, 23(1), 91–123. https://doi.org/10.1080/09535314.2010.534978

Thissen, Mark, Maureen Lankhuizen, Frank van Oort, Bart Los, & Dario Diodato. (2018). *EUREGIO: The construction of a global IO database with regional detail for Europe for 2000–2010* (2018-084/VI; Tinbergen Institute Discussion Paper). https://www.econstor.eu/

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

_____. (1985). The TFS regional modelling methodology. *Regional Studies*, 19, 547–562. https://doi.org/10.1080/09595238500185531

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. https://doi.org/10.1080/09535314.2021.1883553

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. https://doi.org/10.1080/09535314.2018.1545221

Voicu, Alexandru & Michael L. Lahr. (2022). NRRP-induced occupation/skill shifts for Romania's NUTS-2 regions. 9th Hispanic-American Conference on Input-Output Analysis, Aveiro, Portugal, September 22.

Wiebe, Kirsten. S., & Manfred Lenzen. (2016). To RAS or not to RAS? What is the difference in outcomes in multi-regional input–output models? *Economic Systems Research*, 28(3), 383–402. https://doi.org/10.1080/09535314.2016.1192528

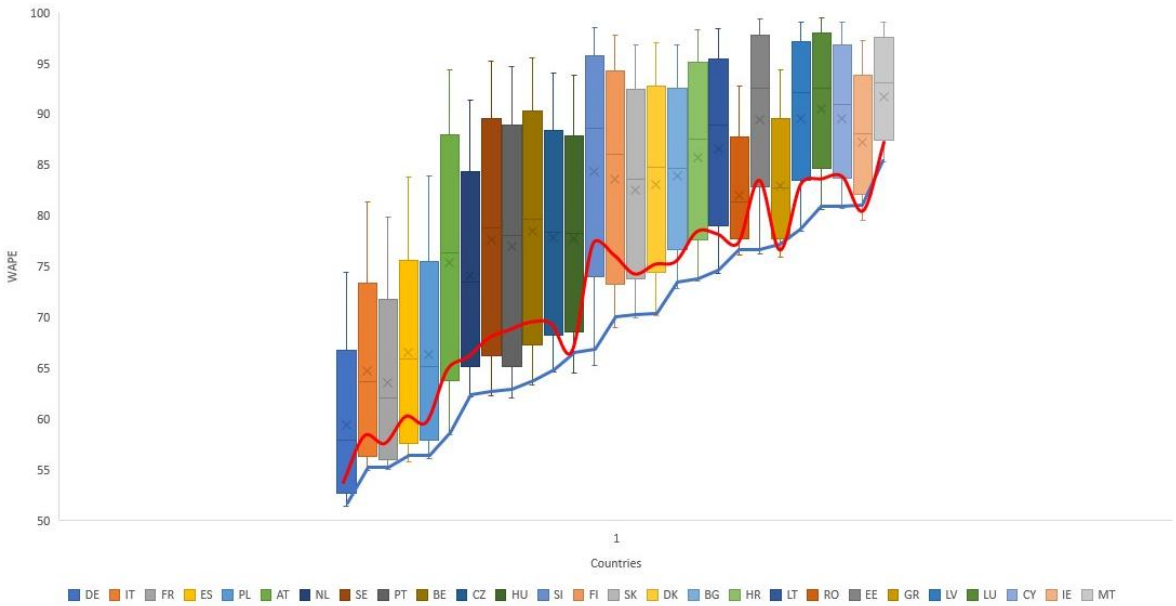
Yamada, Mitsuo. (2015). Construction of a multi-regional input-output table for Nagoya metropolitan area, Japan. *Journal of Economic Structures*, 4(11). https://doi.org/10.1186/s40008-015-0022-7

APPENDIX A

FLQ+ is a FLQ variant presented by Flegg et al. (2021). Its main novelty over the FLQ is that it seeks optimum δ values. The authors generate an initial set of regional **A** matrices using entropy-based methods. They then regress this set of regional **A** matrices to obtain a best-fitting δ value. The resulting optimal δ is then introduced into the standard FLQ equations to attain the goal of the exercise—estimated regional **A** matrices.

We similarly search for an optimum δ value. We calculated 99 FLQ-based sets of regional **A** matrices by ranging δ's value from 0.01 to 0.99 by increments of 0.01. We calculate location quotients and economy size using employment data.[25] We then contrast each of the 99 sets against the true **A** matrices reporting WMAPE as a difference measure. The δ value that minimizes average WMAPE across EU countries in FIGARO is 0.17. Note that for all countries but Greece (GR) and Ireland (IE) the optimal δ yields more accurate results than when δ is set to 0.3 (c.f. the database "global" optimal value is the blue line in Figure A.1, while the WMAPE when $\delta = 0.3$ is depicted by the red line). Note further in Figure A.1 that the optimum δ value minimizing the WMAPE for each country is typically close to that which the global optimal yields.

Figure A.1 – Box & whisker plots of WMAPE on coefficients across EU for all δ values
(Listed left to right by country WMAPE when $\delta = 0.17$, the blue line. The red line is WMAPE when $\delta = 0.3$)



[25] FLQ literature typically uses regional ***output***, data that are rarely published by statistical agencies.

APPENDIX B

Cribari-Neto & 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, Havranek, Irsova, & Schwarz, 2020; Benkovskis Bluhm, Bobeica, Osbat, & Zeugner, 2020) among other applications in economics.

Table B.1. Beta binomial regression estimates for RPCs of goods-producing industries in the EU 27 in 2019[26]

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 **
SD (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 B.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 (and the equation's R^2 is quite similar). This is despite FIGARO's more recent vintage (2019 rather than 2016) and larger number of goods-producing sectors. Note that 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 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 SD, then the more able a sector is to satisfy its county's local intermediate industry and final demands. Further, binary variables of some industries reveal regional preferences to consume locally produced products: primary industries, Food,

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

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 fulfil 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 B.1. Their distribution across countries is displayed in Figure B.2. Note with a few exceptions that means tend to be close to zero and the dispersion of errors is generally limited.

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

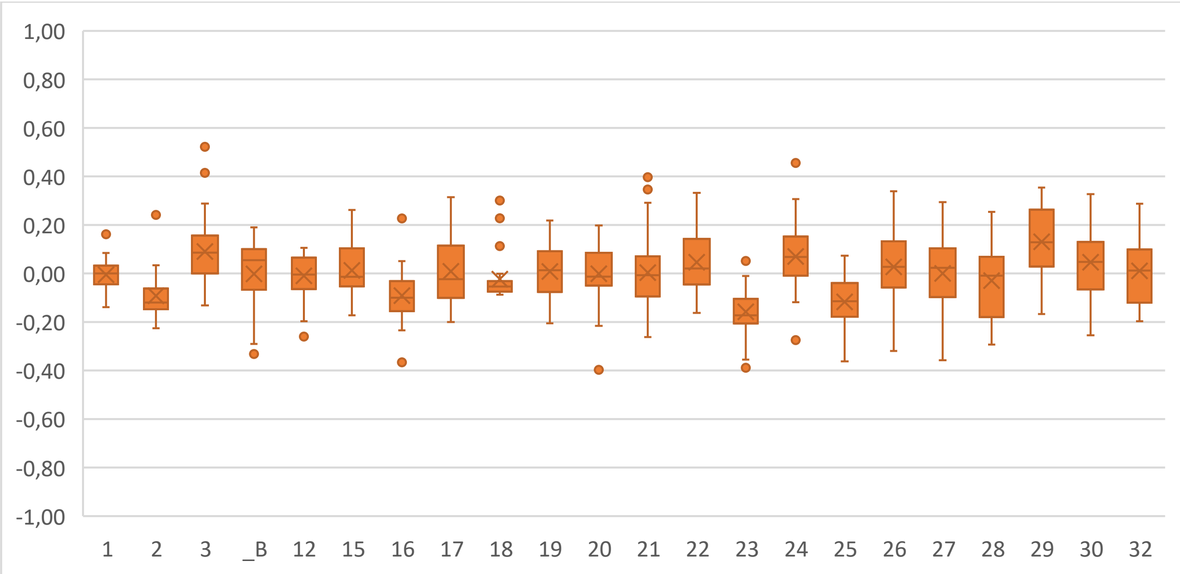
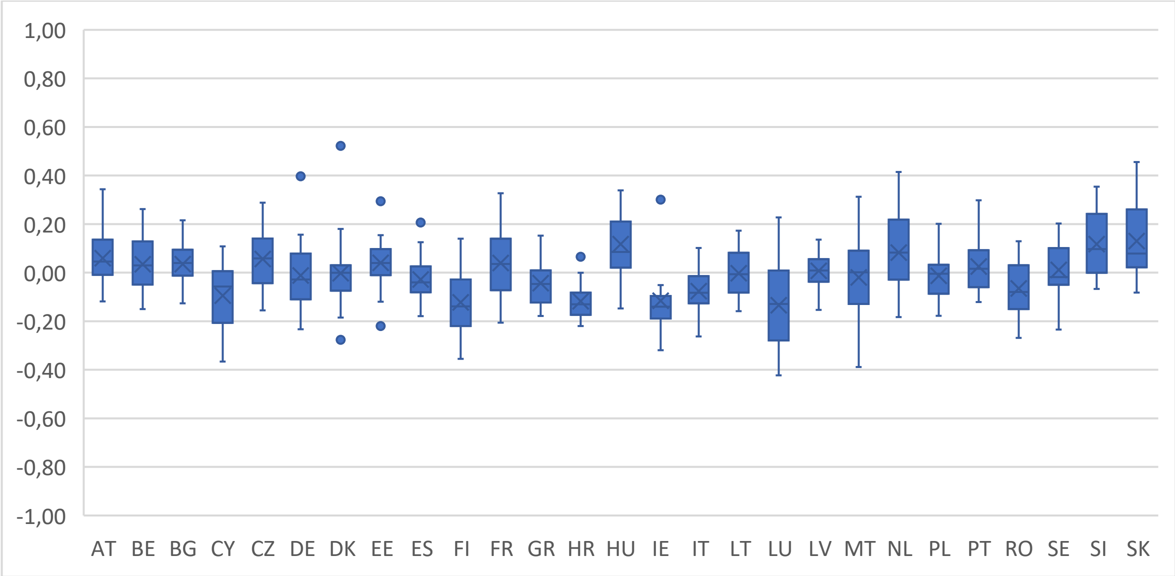


Figure B.2 - Error by country using binomial beta regression.



APPENDIX C

Table C.1. Share of intranational vs international flows by EU country in 2019 according to FIGARO

	Share of transshipments	
Country	**Intra-national**	**International**
AT	75.9%	24.1%
BE	73.1%	26.9%
BG	70.4%	29.6%
CY	73.5%	26.5%
CZ	71.3%	28.7%
DE	79.8%	20.2%
DK	73.2%	26.8%
EE	70.9%	29.1%
ES	83.7%	16.3%
FI	80.5%	19.5%
FR	84.6%	15.4%
GR	82.2%	17.8%
HR	87.2%	12.8%
HU	63.6%	36.4%
IE	58.2%	41.8%
IT	84.8%	15.2%
LT	68.3%	31.7%
LU	57.5%	42.5%
LV	76.8%	23.2%
MT	52.9%	47.1%
NL	72.1%	27.9%
PL	75.9%	24.1%
PT	80.9%	19.1%

RO	80.6%	19.4%
SE	78.5%	21.5%
SI	66.7%	33.3%
SK	66.1%	33.9%