Gravity balancing: a non-survey approach to subnational MRIOs

Mattia Cai

mattia.cai@oecd.org

First draft - May 2023

1. Introduction

Very few National Statistical Offices produce survey-based input-output (IO) data with subnational ('regional') resolution. Regional analysts have debated ways of getting around this problem for several decades now, giving rise to a large and expanding literature on IO regionalization.

The greatest challenge in this area is represented by an almost complete lack of official data on interregional trade. Accordingly, one line of research has focused on estimating bilateral trade between the regions of a country from existing data sources. These approaches are often cumbersome to implement, as they either rely on information that is noisy and incomplete (e.g. freight transport data) or require datasets that are only available under special circumstances (e.g. ad-hoc surveys, pre-existing regional IO tables) (Többen 2017, Zheng at al. 2022).

Conversely, a large family of ('non-survey') techniques attempt to construct single-region IO models without having to estimate bilateral trade flows. For example, location quotient (LQ) methods derive the IO coefficients of a regional model by tweaking those of the country as whole. Even though their theoretical foundations and empirical performance have sometimes been called into question (Hermannsson 2016, Lamonica and Chelli 2018), LQs are very popular in applications thanks to their computational simplicity and minimal data requirements (Buendía et al. 2022, Flegg et al. 2021, Kwon and Choi 2023).

Another popular non-survey approach builds on the fundamental fact that in an input-output table total supply must match total use. Hence, commodity balance (CB) methods seek to work out interregional trade as a balancing item after all other components of the regional IO table (e.g., output, intermediate and final use) have been estimated (e.g. combining national IO and regional accounting data). In terms of minimum data requirements CB techniques are comparable with LQs, but they provide a more intuitive way of incorporating any additional information that may be available. The main drawback is that they yield estimates of net trade but cannot distinguish between imports and exports, although some workarounds have been proposed (Kronenberg 2009).

As many modern applications of IO analysis require multiregional models (e.g. greenhouse gas emission accounting and trade in value added analysis), non-survey regionalization methods are increasingly being generalized to accommodate multiple regions. Still, this tends to require additional assumptions whose relationship with the original single-region setup is not always transparent. Also, ensuring accounting consistency across regions often involves additional balancing steps, as a result of which these techniques are no longer as convenient as their original single-region version (Többen and Kronenberg 2015, Jahn 2017).

In spite of its size and variety, the existing literature does not seem to offer the type of tool that practitioners typically wish for, that is, a framework for constructing multiregional IO databases at the subnational level that would be general, flexible, theoretically coherent and easy to use. This paper sets out to develop one such framework. The procedure proposed here only requires data that in most countries are routinely available from the national statistical office (i.e. national and regional accounts) but is flexible enough to accommodate any relevant additional data. In addition, it relies on standard economic assumptions and well-known econometric methods. Finally, to the extent possible it retains the ease of implementation of traditional non-survey approaches.

In practice, our approach can be described as combining the CB method with a doubly constrained gravity (DCG) model of interregional trade. On one hand, the DCG represents a natural way of blowing up the net trade estimates computed by CB regionalization into a full set of bilateral flows, thus separating imports from exports. On the other, the logic behind the CB method provides us with a simple way of computing the marginal totals that are needed to operationalize the DCG model. The data requirements of the DCG model are kept to a minimum using a calibration procedure discussed in Cai (2021, 2022).

For brevity, we refer to our approach as gravity balancing (GB) regionalization. The remainder of the paper describes the GB method in detail and provides a simple numerical example of its use with simulated data. Eventually, the paper will be expanded with a 21-region, 29-industry and 63-product application to Italy. At the time of writing, however, that work is still ongoing.

2 Accounting framework

Let the country of interest consist of $R > 1$ geographical subdivisions ('regions'). Its economy comprises I classes of goods and services ('products') produced by J branches of activity (‘industries’), with $I \geq J$. There are K primary factors (e.g. labor, capital, etc.) and H final use categories (households, government,...). The system of MRSUTs we aim to estimate is composed of three types of tables: supply tables, use tables and interregional trade matrices. Below, we outline the main features of and the relationship between the three types of tables.

Table 1 displays a stylized supply table for some region r. Its main component is the make matrix $\mathbf{Q}^r$, whose generic element q_{ij}^r measures the amount of product i produced by industry in j region r. The remaining two columns account for products that become available in region r through import from the rest of the country ($\mathbf{m}^r$) or the rest of the world ($\mathbf{g}^r$).

[TABLE 1 ABOUT HERE]

The use table (table 2) describes how products and primary factors are used by region r’s economy. The matrices $\mathbf{Z}^r$ and $\mathbf{Y}^r$ represent the flow of products to intermediate and final users, respectively, whereas $\mathbf{W}^r$ is the value added block. It should be noted that exports are not included in $\mathbf{Y}^r$ but are accounted for separately in columns $\mathbf{e}^r$ (exports to other regions) and $\mathbf{f}^r$ (exports to the rest of the world). No distinction is made between products that are produced within region r and imported products.

[TABLE 2 ABOUT HERE]

In what follows, we will often need to refer to the marginal totals of the matrices displayed in tables 1 and 2. We thus find it convenient to introduce the following shorthand notation: whenever there is no risk of confusion, summation over an index will be denoted by replacing that index with an asterisk. For example, consider matrix $\mathbf{Q}^r$. The sum of the elements in its i-th row, $\sum_j q_{ij}^r$, represents region r’s total output of product i and will be denoted q_{i*}^r. Conversely, q_{*j}^r represents the j-th column total, that is, the total output of industry j in region r. Analogous notation will be used for the margins of $\mathbf{Z}^r$, $\mathbf{Y}^r$, and $\mathbf{W}^r$.

Finally, table 3 displays the origin-destination matrix of interregional trade in generic product i, which we denote $\mathbf{X}_i$. Its (r, s)-th entry, x_i^{rs}, represents the amount of product i produced in region r and used in region s. Given the important role the marginal totals of the trade matrices will play in our estimation procedure, we introduce a dedicated notation for those totals: the $R \times 1$ vectors of $\mathbf{X}_i$’s row and column sums will be denoted $\mathbf{u}_i$ and $\mathbf{v}_i$, respectively. Formally, $\mathbf{u}_i = \mathbf{X}_i \boldsymbol{\iota}$ and $\mathbf{v}_i = \mathbf{X}_i' \boldsymbol{\iota}$ and, where $\boldsymbol{\iota}$ represents a suitably sized vector of ones.

[TABLE 3 ABOUT HERE]

It is assumed that transactions are measured using the same valuation concept (e.g. either at basic prices or at purchasers' prices) throughout tables 1 to 3.

To be internally coherent, our MRSUTs must satisfy two sets of accounting identities. Firstly, in each region r the supply table and the use table must be consistent with each other. From a product perspective, this means that total supply must match total use:

$$q_{i*}^r + m_i^r + g_i^r = z_{i*}^r + y_{i*}^r + e_i^r + f_i^r \tag{1}$$

for all i and r. At the industry level, on the other hand, it is necessary that the sum of intermediate consumption and value added equal total output:

$$q_{*j}^r = z_{*j}^r + w_{*j}^r \tag{2}$$

for all j and r. Secondly, the trade matrices must be compatible with the interregional imports in the supply table and the interregional exports in the use tables. Thus,

$$\sum_{t \neq r} x_i^{tr} = m_i^r \tag{3}$$

and

$$\sum_{t \neq r} x_i^{rt} = e_i^r \tag{4}$$

for all i and r. Taken together, these two equations also guarantee that $\sum_{r=1}^{R} e_i^r = \sum_{r=1}^{R} m_i^r$ for all i, so that interregional trade is balanced at the national level.

If a complete system of MRUSTs is available – that is, if we have supply and use tables for all the R regions and trade matrices are available for all the I products – constructing a MRIO model is a matter of few standard calculations (e.g. transforming the regional SUTs into symmetric IOTs and linking those using regional supply proportions computed from the trade matrices) (Miller and Blair 2022).

2. Data availability considerations

When it comes to estimating the MRSUTs, the greatest difficulties invariably come from its interregional components – the $\mathbf{m}$'s, the $\mathbf{e}$'s, and the underlying $\mathbf{X}$'s. In most applied settings there are indeed no directly observable data at all on interregional trade. The remaining blocks of the RSUTs can also pose estimation challenges, but one can generally find some statistical information to build on. For example, the value added matrix $\mathbf{W}^r$ is often available from regional statistics for all r, even though the detail in terms of industry and primary factor resolution may be much coarser than desired. Similarly, the final use block $\mathbf{Y}^r$ may also be partially known. The regional make and intermediate use matrices $\mathbf{Q}^r$ and $\mathbf{Z}^r$ are not observed but their national counterparts are. Finally, information on regional transaction with the rest of the world, $\mathbf{g}^r$ and $\mathbf{f}^r$, may also be available in some circumstances. Given that data availability conditions can vary significantly across studies, we will not attempt a general discussion of how to exploit that mass of incomplete and indirect information. The interested reader will find numerous examples in the existing literature. In our case study of Italy (which is still work in progress), we will describe what data are available and how they are processed in some detail. At this stage, however, it will be assumed that all MRSUT components have already been estimated except for the $\mathbf{m}$'s, the $\mathbf{e}$'s, and the $\mathbf{X}$'s. Throughout the paper, estimates will be distinguished from the corresponding true values using a superimposed hat. Thus, for the remainder of this section it is assumed that the analyst has knowledge of the estimates $\widehat{\mathbf{Q}}^r$, $\widehat{\mathbf{Z}}^r$, $\widehat{\mathbf{Y}}^r$, $\widehat{\mathbf{W}}^r$, $\hat{\mathbf{f}}^r$ and $\hat{\mathbf{g}}^r$ – and can therefore compute the associated row and column totals – for $r = 1, \dots, R$.

3. Commodity balance regionalization in a single region setting

We begin the discussion of our estimation strategy by considering what a standard single-region application of the CB method would look like in our setup. In a nutshell, the CB approach revolves around the idea that using equation (1) interregional trade can be worked out as a residual balancing item after all the other RSUT components have been estimated. Let $b_i^r = e_i^r - m_i^r$ denote net interregional export of product i by generic region r. Substituting this into (1), rearranging terms and replacing the unknown right-hand side quantities with the corresponding estimates gives us the following estimate of b_i^r:

$$\hat{b}_i^r = \hat{q}_{i*}^r + \hat{g}_i^r - \hat{z}_{i*}^r - \hat{y}_{i*}^r - \hat{f}_i^r \tag{5}$$

Effectively, $\hat{b}_i^r$ is set equal to the value that balances region r's SUTs taking all other estimates as given.

Equation (5) represents the core of the CB approach. Unfortunately, it says nothing on how to decompose $\hat{b}_i^r$ into separate estimates of import ($\hat{m}_i^r$) and export ($\hat{e}_i^r$). One classic solution has been to assume that, in any given product category, region r can be either an importer or an exporter, but not both. Thus, one would set: $\hat{e}_i^r = \hat{b}_i^r$ and $\hat{m}_i^r = 0$ if $\hat{b}_i^r$ is positive; $\hat{m}_i^r = -\hat{b}_i^r$ and $\hat{e}_i^r = 0$ if $\hat{b}_i^r$ is negative. Unfortunately, economic statistics provide ample evidence of pervasive intra-industry trade (or 'cross hauling'). A popular and more modern approach is represented by the CHARM method (Kronenberg 2009). The starting point is the following observation: the reason we observe cross hauling in economic data is that statistical classifications lump together entire classes of differentiated (or outright different) products. Hence, the central element of the CHARM method is a product heterogeneity coefficient that can be calculated from observable country-level data. Using this coefficient, the amount of cross hauling in product i is quantified, which in turn makes it possible to split $\hat{b}_i^r$ into separate $\hat{e}_i^r$ and $\hat{m}_i^r$ estimates.

While subsequent improvements of the CHARM method have overcome most limitations of the original formulation, it is still true that in a multiregional context this technique does not by itself yield a complete MRIO system, but rather a set of single-region tables (Többen and Kronenberg 2015). In addition, some scholars have objected to the underlying assumption of a spatially uniform heterogeneity coefficient (Jackson 2014). For these reasons, we attempt to develop the CB method in a different direction.

4. Commodity balance with bilateral trade flows

Rather than attempting to estimate $\mathbf{m}^r$ and $\mathbf{e}^r$ directly, our approach is to start by estimating the underlying bilateral trade flows: first, we construct a trade matrix estimate $\hat{\mathbf{X}}_i$ for each product i; then, the elements of $\hat{\mathbf{m}}^r$ and $\hat{\mathbf{e}}^r$ can be obtained by straightforward adding up in analogy with equations (3) and (4), respectively.

Bilateral trade flows will be estimated one product at time. Let us consider generic product i and the corresponding interregional trade matrix $\mathbf{X}_i$. Under our data availability conditions the row and column totals of $\mathbf{X}_i$ are easy to estimate. To see this, use equations (3) and (4) to write those rows and column totals as

$$u_i^r = x_i^{rr} + e_i^r \tag{5}$$

and

$$v_i^r = x_i^{rr} + m_i^r \tag{6}$$

for all r.

Then, let us break down region r's total output of product i into three components, depending on whether the product is used within region r itself, in the rest of the country or abroad. Formally, $q_{i*}^{r} = x_i^{rr} + e_i^r + f_i^r$, or equivalently

$$x_i^{rr} = q_{i*}^{r} - e_i^r - f_i^r \tag{7}$$

Substituting (7) into (5) and plugging in the relevant estimates on the right-hand side gives the following estimates of $\mathbf{X}_i$'s r-th row total:

$$\hat{u}_i^r = \hat{q}_{i*}^{r} - \hat{f}_i^r \tag{8}$$

The column totals are just as easy to estimate. Substituting (7) into (6) gives

$$v_i^r = q_{i*}^{r} - e_i^r - f_i^r + m_i = z_{i*}^{r} + y_{i*}^{r} - g_i^r$$

where the second equality follows directly from equation (1). Thus, the r-th column total of $\mathbf{X}_i$ can be estimated as

$$\hat{v}_i^r = \hat{z}_{i*}^{r} + \hat{y}_{i*}^{r} - \hat{g}_i^r \tag{9}$$

Comparing the system of equations (8) and (9) with equation (5) highlights two things. The first one is that, from the practical point of view, computing $\hat{u}_i^r$ and $\hat{v}_i^r$ requires the same data as computing $\hat{b}_i^r$. In other words, the data requirements are no more demanding in our approach than with traditional CB. Secondly, given that $\hat{u}_i^r - \hat{v}_i^r = \hat{b}_i^r$, our row and column total estimates imply the same level of net interregional export as the CB method.

5. A doubly constrained gravity model of interregional trade

Once the marginal total estimates $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ have been obtained from equations (8) and (9), we can estimate $\mathbf{X}_i$ using a simple doubly constrained gravity model approach. As a detailed methodological discussion of the method is already be found in Cai (2021, 2022), the argument will only be sketched here.

The central premise of the method is that the entries of $\mathbf{X}_i$ follow a generalized gravity model of the form

$$x_i^{rs} = \alpha_i^r \beta_i^s (d^{rs})^{-\theta_i} \tag{10}$$

for $r, s = 1, \dots, R$. The symbol d^{rs} denotes the distance from region r to region s and is assumed to be known and strictly positive (including when $r = s$). The Greek letters represent various unobserved product-specific parameters that we also assume to be strictly positive: θ_i is the distance elasticity of trade, whereas α_i^r and β_i^s are origin and destination effects, respectively. This formulation is quite general, as many variables typically included in empirical gravity models vary either by origin only (in which case their effect can be absorbed in α_i^r) or by destination only (in which case their effect can be absorbed in β_i^s)[1].

Our estimation strategy is based on the following observation: if the distance elasticity parameter θ_i and the marginal totals $\mathbf{u}_i$ and $\mathbf{v}_i$ were known, then $\mathbf{X}_i$ could actually be solved for. To see this, rewrite the system of equations (10) as

$$\mathbf{X}_i = \langle \boldsymbol{\alpha}_i \rangle \mathbf{T}_i \langle \boldsymbol{\beta}_i \rangle \tag{11}$$

where the notation is as follows: $\mathbf{T}_i$ is an $R \times R$ matrix with generic element $(d^{rs})^{-\theta_i}$; $\boldsymbol{\alpha}_i = (\alpha_i^1, \alpha_i^2, \dots, \alpha_i^R)$ and $\boldsymbol{\beta}_i = (\beta_i^1, \beta_i^2, \dots, \beta_i^R)$ are R-vectors collecting the origin and destination effects, respectively; angled brackets denote diagonalization of the relevant vector into a matrix. With knowledge of θ_i available, $\mathbf{T}_i$ can be computed and our problem reduces to one of recovering the vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$. But if the marginal totals $\mathbf{u}_i$ and $\mathbf{v}_i$ are also known, then $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ can be calculated using the RAS algorithm (up to an innocuous multiplicative constant). Effectively, $\mathbf{X}_i$ could be found as the balanced matrix obtained by RAS specifying $\mathbf{T}_i$ as the matrix of initial values and the pair $(\mathbf{u}_i, \mathbf{v}_i)$ as the target marginal totals.

In real-world applications, θ_i, $\mathbf{u}_i$ and $\mathbf{v}_i$ are not observed, but they can be replaced with estimates. We have already discussed how to construct the marginal total estimates $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ in the previous section. In most applications, obtaining an estimate of the distance elasticity θ_i should also be easy. Several solutions are documented in the literature. They range from imputing a value that the analyst deems realistic (most often 1) (Sargento et al. 2012, Johansen et al. 2018, Distefano et al. 2019, Fournier Gabela 2020) to conducting econometric work on freight transport (Lindall et al. 2006, Boero et al. 2018) or international trade data (Jahn et al. 2017). In our case study of Italy the θ's are estimated from international trade data as in Cai (2022).

[1] Equation (10) incorporates the assumption of a constant elasticity distance decay process. Adopting a more flexible formulation would allow for richer patterns of spatial dependence but would also increase data requirements and ultimately complicate our exposition unnecessarily.

To sum it up, the GB procedure by which we envisage estimating the MRSUTs consists of the following five steps:

1. Drawing on preexisting data, construct estimates of the matrices $\widehat{\mathbf{Q}}^r$, $\widehat{\mathbf{Z}}^r$, $\widehat{\mathbf{Y}}^r$, $\widehat{\mathbf{W}}^r$, $\hat{\mathbf{f}}^r$ and $\hat{\mathbf{g}}^r$ for all the R regions of the country.
2. Use those estimates along with equations (8) and (9) to compute $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ for all I products in the accounting system.
3. For each product i, obtain a distance elasticity estimate $\hat{\theta}_i$ and use the data on interregional distances to compute the matrix $\widehat{\mathbf{T}}_i$ with generic element $(d^{rs})^{-\hat{\theta}_i}$.
4. Proceeding one product at a time, find $\widehat{\mathbf{X}}_i$ by scaling the matrix of initial values $\widehat{\mathbf{T}}_i$ to the estimated marginal totals $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ by means of the RAS algorithm.
5. Once $\widehat{\mathbf{X}}_i$ is available for all i, compute interregional import and export vectors $\widehat{\mathbf{m}}^r$ and $\hat{\mathbf{e}}^r$ for all r by analogy with equations (3) and (4).

At this point it is natural to inquire into the feasibility and balancing properties of our regionalization procedure.

6. Feasibility issues

The issue of feasibility revolves around the following question: how confident can we be that the RAS problem in step 4 will have a solution for all i? Ultimately, the answer depends on how step 1 of the procedure is carried out. As noted above, the non-trade components of the MRSUTs can be estimated in various ways depending on the data availability conditions in which one operates. Even so, the country-level SUTs are invariably a key source of information and play a central role in the process.

In general terms, the fundamental structure of the national SUTs mirrors that of their regional counterparts in tables 1 and 2. The main difference concerns the interregional import and interregional export columns, which are absent from the national tables. Let us then introduce the following notation for the constituents of the national SUTs: each country-level matrix will be denoted by the same symbol as was used for its regional equivalent but with the r superscript dropped and an overbar added. Thus, for example, $\overline{\mathbf{Q}}$ represents the national make matrix and $\bar{q}_{i*}$ the total of its i-th row.

In step 1 of GB regionalization, it is easy to ensure that the following conditions hold for all i:

$$\textstyle\sum_r \hat{q}_{i*}^r = \bar{q}_{i*}; \quad \sum_r \hat{z}_{i*}^r = \bar{z}_{i*}; \quad \sum_r \hat{y}_{i*}^r = \bar{y}_{i*}; \quad \sum_r \hat{g}_{i*}^r = \bar{g}_{i*}; \quad \sum_r \hat{f}_{i*}^r = \bar{f}_{i*} \qquad (12)$$

The conditions spelled out in (12) are merely requirements that, product by product, our regional estimates of output, intermediate use, final use, foreign imports and foreign exports add up to the appropriate national total from the country-level SUTs. In applied work regional estimates are generally obtained by splitting a national aggregate extracted from the SUTs in proportion to some economic variable that can be observed at the regional level (e.g. employment), so that the step 1 estimates satisfy (12) automatically.

Let us now come back to the question of whether RAS will be feasible in step 4 for generic product i. Because the matrix of starting values $\widehat{\mathbf{T}}_i$ is strictly positive, all it takes for RAS to converge is that $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ be compatible with each other (Bacharach 1970), i.e., that

$$\sum_r \hat{u}_i^r = \sum_r \hat{v}_i^r \tag{13}$$

Substituting equations (8) and (9) into this expression, rearranging terms and assuming that (12) holds for product i gives:

$$\bar{q}_{i*} + \bar{g}_i = \bar{z}_{i*} + \bar{y}_{i*} + \bar{f}_i \tag{14}$$

This condition is the country-level counterpart of the commodity balance requirement introduced in equation (1) and will be met by any legitimate national SUTs. Thus, to ensure that our regionalization procedure does not face any convergence problems in step 4, it is sufficient that the non-trade estimates computed in step 1 satisfy equation (12) for all products.

7. Balancing properties

We then turn to the question of balancing: to what extent can we expect the MRSUTs estimated by GB to constitute an internally consistent accounting system, i.e., one in which the empirical counterparts of equations (1-4) are satisfied for all r? The first thing to notice is that, because the trade components of the system have been calculated using equations (1) and (3-4), the estimated MRSUTs will satisfy those equations by construction. Conceivably, the system may still fail to balance if the value of industry output does not match between the supply and use table as required by equation (2). In practice, however, this is little cause for concern. In a typical application, the paucity of regional level data means that the estimation of $\widehat{\mathbf{Q}}^r$ and $\widehat{\mathbf{Z}}^r$ in step 1 can only be approached in one way: first estimate the industry outputs (the $\hat{q}_{*j}^r$'s) and then construct both $\widehat{\mathbf{Q}}^r$ and $\widehat{\mathbf{Z}}^r$ on the basis of those estimates (typically in combination with country-level data). Naturally, this way of proceeding produces MRSUTs in which equation (2) is also satisfied.

One further aspect of the balancing issue concerns whether the building blocks of the regional supply and use tables add up to their national equivalents. In analytical terms, this type of national balancing imposes the following set of conditions on the MRSUTs:

$$\sum_r \widehat{\mathbf{Q}}_r = \overline{\mathbf{Q}}; \quad \sum_r \widehat{\mathbf{Z}}_r = \overline{\mathbf{Z}}; \quad \sum_r \widehat{\mathbf{Y}}_r = \overline{\mathbf{Y}}; \quad \sum_r \widehat{\mathbf{W}}_r = \overline{\mathbf{W}}; \quad \sum_r \hat{\mathbf{f}}_r = \bar{\mathbf{f}}; \quad \sum_r \hat{\mathbf{g}}_r = \bar{\mathbf{g}} \tag{15}$$

Whether or not (15) holds in a system of MRSUTs generated by GB depends once again on how step 1 is carried out.

8. A simple numerical example

This section demonstrates the mechanics of GB regionalization using a highly simplified numerical example. At the center of the exercise is a fictional country consisting of four regions (North, East, South and West). The economy is comprised of three branches of economic activity – namely, Agriculture (AGR), Manufacturing (MNF) and Services (SRV) – each with its distinctive primary product.

Our goal is to construct for this country a system of MRSUTs akin to that of section 2. To keep the size of the example manageable, we introduce the following three assumptions. Firstly, the country is closed to international trade. In other words, the regions trade with each other but not with the rest of the world, so that $\mathbf{g}^r = \mathbf{f}^r = \mathbf{0}$ for all r and $\bar{\mathbf{g}} = \bar{\mathbf{f}} = \mathbf{0}$ in the national SUTs. Secondly, there are no secondary productions. Accordingly, all regional make matrices $\mathbf{Q}^r$ are diagonal, as well as their country-level analogue $\overline{\mathbf{Q}}$. Thirdly, we posit that all regional economies are proportionally scaled versions of the national economy. Thus, for all r

$$\mathbf{Q}^r = \rho^r \overline{\mathbf{Q}}; \qquad \mathbf{Z}^r = \rho^r \overline{\mathbf{Z}}; \qquad \mathbf{Y}^r = \rho^r \overline{\mathbf{Y}}; \qquad \mathbf{W}^r = \rho^r \overline{\mathbf{W}}; \tag{16}$$

with $\rho^r \in (0,1)$ and $\sum_r \rho^r = 1$.

The first source of data we assume available to the analyst are the country-level SUTs. The use table is displayed in table 4. For compactness, no breakdown is provided for either the value added row or the final use column. In addition, given that $\bar{\mathbf{f}} = \mathbf{0}$ the export column has been omitted. In our simplified setting, the national supply table can also be dispensed with, because – with $\bar{\mathbf{g}} = \mathbf{0}$ and $\overline{\mathbf{Q}}$ diagonal – it does not contain any information that is not already captured in the use table. The second piece of information we presume available is estimates of each region's share in the national economy. Let us thus suppose that the $\hat{\rho}$'s are given by 0.1, 0.35, 0.4 and 0.15 for North, South, East and West, respectively. Using these data, $\widehat{\mathbf{Q}}^r$, $\widehat{\mathbf{Z}}^r$, $\widehat{\mathbf{Y}}^r$ and $\widehat{\mathbf{W}}^r$ are computed for all r by analogy with (16). Hence, the only MRSUT elements we have left to estimate are those pertaining to interregional trade.

[TABLE 4 ABOUT HERE]

To estimate interregional trade, we rely on the DCG model framework of section 2.5. To begin with, we compute the target marginal totals for our RAS procedures by substituting the relevant estimates into equations (8) and (9). In the simplified setting of our example, those two equations respectively reduce to

$$\hat{u}_i^r = \hat{q}_{i*}^r \tag{8a}$$

and

$$\hat{v}_i^r = \hat{q}_{i*}^r \tag{9a}$$

for all i and r. The former follows trivially from setting $\hat{f}_i^r = 0$ in equation (8). To obtain equation (9a), on the other hand, note that – in the absence of imports from the rest of the world and with regional aggregates proportional to their national counterparts – equation (9) becomes $\hat{v}_i^r = \hat{z}_{i*}^r + \hat{y}_{i*}^r = \hat{\rho}^r(\bar{z}_{i*} + \bar{y}_{i*})$. Also, in a closed economy the country-level commodity balance condition of equation (14) simplifies to $\bar{z}_{i*} + \bar{y}_{i*} = \bar{q}_{i*}$. Hence, $\hat{v}_i^r = \hat{\rho}^r \bar{q}_{i*} = \hat{q}_{i*}^r$. In other words, in the simplified setting of our example, the row and column total of any given trade matrix are both given by the vector of regional outputs of the relevant product. For the example data, the value of $\hat{q}_{i*}^r$ is tabulated for all i and r in table 5.

[TABLE 5 ABOUT HERE]

To compute our bilateral trade estimates, we will need two additional pieces of information. The first one is a matrix of pairwise interregional distances. For our example it is given in table 6. The final requirement is a set of distance elasticity estimates for the various products of the economy. Let us assume that preliminary econometric work has given the following estimates: $\hat{\theta}_{\mathrm{AGR}} = 0.5$, $\hat{\theta}_{\mathrm{MNF}} = 1$ and $\hat{\theta}_{\mathrm{SRV}} = 1.5$.

[TABLE 6 ABOUT HERE]

Consider, for instance, the case of services. For i = SRV, our matrix of starting values is:

$$\widehat{\mathbf{T}}_{\mathrm{SRV}} = [d_{rs}^{-1.5}] = \begin{bmatrix} 12.1 & 3.2 & 1.0 & 1.0 \\ 3.2 & 4.4 & 1.5 & 1.7 \\ 1.0 & 1.5 & 37.0 & 1.6 \\ 1.0 & 1.7 & 1.6 & 27.4 \end{bmatrix} \times 10^{-3}$$

Applying the RAS algorithm to $\widehat{\mathbf{T}}_{\mathrm{SRV}}$ with the row and column totals appropriate for Services from table 5 yields the interregional trade matrix estimate $\widehat{\mathbf{X}}_{\mathrm{SRV}}$ displayed in table 7.

[TABLE 7 ABOUT HERE]

Using the same procedure for Agriculture and Manufacturing, we obtain

$$\widehat{\mathbf{X}}_{\mathrm{AGR}} = \begin{bmatrix} 0.5 & 1.1 & 0.6 & 0.3 \\ 1.1 & 4.1 & 2.4 & 1.2 \\ 0.6 & 2.4 & 6 & 1 \\ 0.3 & 1.2 & 1 & 1.2 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{X}}_{\mathrm{MNF}} = \begin{bmatrix} 3.9 & 5.6 & 1.6 & 0.9 \\ 5.6 & 24.6 & 7.3 & 4.5 \\ 1.6 & 7.3 & 36.6 & 2.6 \\ 0.9 & 4.5 & 2.6 & 10 \end{bmatrix}$$

It is worth noting that our trade matrices are symmetric. This is a consequence of the simplifying assumptions we have introduced in this example, which results in net exports being zero for all regions and products. Clearly, this will not hold in the general case.

Finally, we can compute interregional import and export vectors for all regions to complete our system of MRSUTs. For example, the supply and use tables for region South are given in tables 8 and 9, respectively.

[TABLE 8 ABOUT HERE]

[TABLE 9 ABOUT HERE]

References

Buendia Azorin, J. D., Martinez Alpanez, R., & Sanchez de la Vega, M. D. M. (2022). A new proposal to model regional input–output structures using location quotients. An application to Korean and Spanish regions. *Papers in Regional Science*, *101*(5), 1219-1237.

Cai, M. (2021). Doubly constrained gravity models for interregional trade estimation. *Papers in Regional Science*, *100*(2), 455-474.

Cai, M. (2023). A calibrated gravity model of interregional trade. *Spatial Economic Analysis*, *18*(1), 89-107.

Flegg, A. T., Lamonica, G. R., Chelli, F. M., Recchioni, M. C., & Tohmo, T. (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.

Hermannsson, K. (2016). Beyond intermediates: The role of consumption and commuting in the construction of local input–output tables. *Spatial Economic Analysis*, *11*(3), 315-339.

Jahn, M. (2017). Extending the FLQ formula: A location quotient-based interregional input–output framework. *Regional Studies*, *51*(10), 1518-1529.

Kwon, H., & Choi, S. G. (2023). An alternative approach to estimating regional input–output tables: the KFLQ method. *The Annals of Regional Science*, 1-18.

Kronenberg, T. (2009). Construction of regional input-output tables using nonsurvey methods: the role of cross-hauling. *International regional science review*, *32*(1), 40-64.

Lamonica, G. R., & Chelli, F. M. (2018). The performance of non-survey techniques for constructing sub-territorial input-output tables. *Papers in Regional Science*, *97*(4), 1169-1202.

Többen, J. (2017). On the simultaneous estimation of physical and monetary commodity flows. *Economic Systems Research*, *29*(1), 1-24.

Többen, J., & Kronenberg, T. H. (2015). Construction of multi-regional input–output tables using the CHARM method. *Economic Systems Research*, *27*(4), 487-507.

Zheng, H., Többen, J., Dietzenbacher, E., Moran, D., Meng, J., Wang, D., & Guan, D. (2022). Entropy-based Chinese city-level MRIO table framework. *Economic Systems Research*, *34*(4), 519-544.

Tables and figures

Table 1 – Supply table for region *r*

		Industries	Interregional imports	Foreign imports
	Dimensions	J	1	1
Products	I	$\mathbf{Q}^r = [q_{ij}^r]$	$\mathbf{m}^r = [m_i^r]$	$\mathbf{g}^r = [g_i^r]$

Table 2 – Use table for region *r*

		Industries	Domestic final demand	Interregional exports	Foreign exports
	Dimensions	J	H	1	1
Products	I	$\mathbf{Z}^r = [z_{ij}^r]$	$\mathbf{Y}^r = [y_{ih}^r]$	$\mathbf{e}^r = [e_i^r]$	$\mathbf{f}^r = [f_i^r]$
Primary factors	K	$\mathbf{W}^r = [w_{kj}^r]$			

Table 3 – Trade matrix for product *i*

Trade matrix ($\mathbf{X}_i$)		Destination: Region 1	Region 2	…	Region R	Row sums ($\mathbf{u}_i$)
Origin	Region 1	x_i^{11}	x_i^{12}	…	x_i^{1R}	u_i^1
	Region 2	x_i^{21}	x_i^{22}	…	x_i^{2R}	u_i^2
	⋮	⋮	⋮		⋮	⋮
	Region R	x_i^{R1}	x_i^{R2}	…	x_i^{RR}	u_i^R
	Column sums ($\mathbf{v}_i^{\top}$)	v_i^1	v_i^2	…	v_i^R	

Table 4 – Simulated use table of the national economy

	AGR	MNF	SRV	Final use	**Total output**
Agriculture (AGR)	4	6	0	15	**25**
Manufacturing (MNF)	5	50	20	45	**120**
Services (SRV)	6	24	30	40	**100**
Value added	10	40	50		**100**
Total output	**25**	**120**	**100**	**100**	

Table 5 – Estimated regional output of products

	North	East	South	West	Sums
AGR	2.5	8.75	10	3.75	25
MNF	12	42	48	18	120
SRV	10	35	40	15	100
Sums	24.5	85.75	98	36.75	

Table 6 – Simulated bilateral distances

	N	E	S	W
North (N)	19	46	100	99
East (E)	46	37	75	71
South (S)	100	75	9	74
West (W)	99	71	74	11

Table 7 – Estimated bilateral trade in Services

	North	East	South	West	Sums
North	4.5	4.6	0.6	0.4	10
East	4.6	24.4	3.5	2.5	35
South	0.6	3.5	34.9	1	40
West	0.4	2.5	1	11.1	15
Sums	10	35	40	15	

Table 8 – Estimated supply table for region South

	AGR	MNF	SRV	Interreg. import	Sum
Agriculture (AGR)	10	0	0	4.0	**14.0**
Manufacturing (MNF)	0	48	0	11.4	**59.4**
Services (SRV)	0	0	40	5.1	**45.1**
Total output	**10**	**48**	**40**	**20.5**	

Table 9 – Estimated use table for region South

	AGR	MNF	SRV	Final use	Interreg. export	Sum
Agriculture (AGR)	1.6	2.4	0	6	4.0	**14.0**
Manufacturing (MNF)	2	20	8	18	11.4	**59.4**
Services (SRV)	2.4	9.6	12	16	5.1	**45.1**
Value added	4	16	20			**40.0**
Total output	**10**	**48**	**40**	**40**	**20.5**	