Flegg *et al. Economic Structures* (2021) 10:12
https://doi.org/10.1186/s40008-021-00242-8

Journal of Economic Structures

RESEARCH Open Access

A new approach to modelling the input–output structure of regional economies using non-survey methods

Anthony T. Flegg[1*], Guiseppe R. Lamonica[2], Francesco M. Chelli[2], Maria C. Recchioni[2] and Timo Tohmo[3]

*Correspondence:
tony.flegg@uwe.ac.uk
[1] Department of Accounting, Economics and Finance, University of the West of England, Coldharbour Lane, Bristol BS16 1QY, UK
Full list of author information is available at the end of the article

Abstract

This paper proposes a new approach to the regionalization of national input–output tables where suitable regional data are scarce and analysts are considering using location quotients (LQs). We focus on the FLQ formula, which frequently yields the best results of the pure LQ-based methods, and develop an enhanced way of implementing this approach. We use a modified cross-entropy (MCE) method, along with a regression model, to estimate values of the unknown parameter δ in the FLQ formula, specific to both region and country. An analysis of survey-based data for 16 South Korean regions reveals that the proposed FLQ+ approach yields more accurate estimates of both input coefficients and sectoral output multipliers than those from simpler LQ-based methods or the MCE approach alone. Sectoral outputs (or employment) are the only regional data required. The MCE method also clearly outperforms GRAS.

Keywords: FLQ, FLQ+, Cross-entropy, GRAS, Regional tables

1 Introduction

Many non-survey methods of regionalizing a national input–output table (NIOT) have been developed, with the aim of avoiding the high costs and lengthy delays associated with constructing regional tables via survey-based methods. Non-survey methods include those based on location quotients (LQs) and constrained matrix-balancing methods.

The LQ-based methods include the classical *simple* and *cross-industry* LQs, along with several refinements examined in the next section. These methods hinge on the assumption that regions and nations employ the same technology of production, with the implication that regional input coefficients differ from their national counterparts only so far as each region imports goods and services from other regions. The performance of LQ-based methods deteriorates significantly when this assumption of a common technology is violated.

By contrast, constrained matrix-balancing methods are less sensitive to violations of the assumption of identical technology. These procedures estimate unknown data from limited initial information, subject to a set of linear constraints. The most

popular of such techniques are RAS (Bacharach 1970; Stone 1961) and the cross-entropy (CE) method (Golan et al. 1994), both of which are known to work well (Davis et al. 1977; Golan and Vogel 2000; Hosoe 2014; Kapur and Kesavan 1992; Lamonica et al. 2020; Léony et al. 1999; Robinson et al. 2001; Vazquez et al. 2015). The constrained matrix-balancing procedures include methods based on minimizing squared or absolute differences (Pavia et al. 2009). However, such techniques are more time-consuming to implement than the pure LQ-based methods and normally require the solution of a constrained nonlinear optimization problem, whereas the LQ-based methods are very quick and simple to apply.

The present study focuses on the FLQ (Flegg's LQ), which is often found to be one of the most accurate pure LQ-based methods (Bonfiglio and Chelli 2008; Flegg and Tohmo 2016). Applications of the FLQ include Dávila-Flores (2015), Hermannsson (2016), Jahn (2017), Kronenberg and Fuchs (2021), Morrissey (2016), and Singh and Singh (2011). To apply the FLQ, a value for a crucial unknown parameter, δ, must be chosen. Although several empirical studies have sought to find appropriate values of δ, these studies have not been conclusive. Examples include analyses of data for Scotland (Flegg and Webber 2000), Germany (Kowalewksi 2015), Argentina (Flegg et al. 2016) and South Korea (Flegg and Tohmo 2019). Given these diverse outcomes, Jahn et al. (2020) recommend that analysts should consider regional characteristics when selecting a value of δ in the interval 0.3 ± 0.1. This range, which is suggested by earlier empirical work, can then be refined in the light of the results. Jahn et al. also formulate an econometric model that should assist in this process.

Nevertheless, it is evident that the choice of a suitable value of δ is still an open question, which has limited the practical use of the FLQ. Our aim is, therefore, to propose a strategy to estimate the value of δ in such a way as to optimize the regionalization of the NIOT. This strategy combines constrained matrix-balancing procedures with the FLQ. The expectation is that this hybrid approach should yield better results than would be attainable by applying each approach alone.

Of the possible matrix-balancing procedures, the CE method was chosen rather than RAS for several reasons. In particular, the standard RAS method can only handle non-negative matrices, which would limit the application proposed in this paper. Although its generalization GRAS (Junius and Oosterhaven 2003) enables matrices with negative elements to be updated, its objective function has been questioned by Lemelin (2009), Huang et al. (2008) and Temurshoev et al. (2013). Moreover, the results obtained by Golan et al. (1994) and Robinson et al. (2001) appear to favour the CE method, while Lamonica et al. (2020) demonstrate that this method performs well when it is applied to real data, especially for small economies. Indeed, we find that the *modified* CE (MCE) method, as implemented in this study, outperforms GRAS.

The proposed procedure involves three steps. The first applies the MCE method to regionalize the NIOT. This is designed to account for negative or zero input coefficients. The second step uses the derived regional matrix, along with the national table, to estimate the optimal δ for each region via a simple regression model. In the third step, this estimated δ is used to apply the FLQ formula, thereby computing the final estimates of the regional input coefficients. It is worth noting that this hybrid

approach can easily be adapted to enhance the performance of other pure LQ-based methods that depend on one or more unknown parameters.

Our approach is tested on the South Korean interregional input–output table (KIRIOT) for the year 2005. It was built by the Bank of Korea for all 16 South Korean regions, with a classification of 78 economic sectors. A smaller version with 28 economic sectors is also available, but we opted to use the larger version, so as to minimize aggregation bias. The KRIOT is one of the very few survey-based full interregional I–O tables. It has data on the volume of all intersectoral transactions, both within and across regions, so it is ideal for our purposes. Indeed, Zhao and Choi (2015, p. 909) remark:

> *[In 2009, the Bank of Korea] divided the entire country into 16 regions and issued regional Leontief inverse tables and type I output multipliers for each of these regions. These data can be considered as benchmark tables since they are entirely based on surveys.*

Since the KIRIOT contains negative and null flows, the MCE method is employed. The empirical analysis demonstrates that the enhanced version of the FLQ approach developed here outperforms the MCE method, when it is applied separately, and has a similar performance to the 'optimal' FLQ approach, where δ is selected by using the observed regional coefficient matrix. Of course, in reality, such a matrix would seldom be available to regional analysts; its use here is merely to generate optimal values of δ that can be used as a benchmark in our analysis. Our aim is to offer a new way of regionalizing a NIOT in situations where an analyst has access only to the total output of each regional sector. Where such data are unavailable, sectoral employment could be used as a proxy. This proposed hybrid method, hereafter referred to as the FLQ+ method, is an improvement on the present state of the art, in which analysts need to select values of δ on the basis of a priori considerations, e.g. values found in earlier studies, or by taking regional characteristics into account, as is suggested by Jahn et al. (2020). In order to demonstrate the practical advantages of using the FLQ+ method, we compare its performance with the results from the MCE method, GRAS and the FLQ with a single assumed value of δ for every region.

2 Review of pure LQ-based methods

Here, we review the pure LQ methods most often employed to construct regional input–output tables (RIOTs). Some alternative approaches are examined thereafter.

Consider a national economic system consisting of k sectors. Let $\mathbf{X}^n = \left[x_{ij}^n\right]$ and $\mathbf{X}^r = \left[x_{ij}^r\right]$ be matrices whose elements are the flows for intermediate use from sector i to sector j at the national and regional levels, respectively, while $\mathbf{x}^n$ and $\mathbf{x}^r$ are vectors of national and regional total sectoral output. Also, let $\mathbf{A}^n = \left[a_{ij}^n = \frac{x_{ij}^n}{x_j^n}\right]$ and $\mathbf{A}^r = \left[a_{ij}^r = \frac{x_{ij}^r}{x_j^r}\right]$ be the matrices whose elements are the respective national and regional input coefficients.

Now suppose that only $\mathbf{A}^n$ and the vector of regional total sectoral output, $\mathbf{x}^r$, are known. The pure LQ methods are used to estimate the matrix of regional input coefficients, $\mathbf{A}^r$, by adjusting the national input coefficient as follows:

$$\hat{a}_{ij}^r = a_{ij}^n q_{ij}, \tag{1}$$

where q_{ij} represents a scalar applied to the national coefficient.

With the *simple* LQ, a_{ij}^r is estimated via the following formula:

$$\hat{a}_{ij}^r = \mathrm{SLQ}_i a_{ij}^n, \tag{2}$$

where

$$\mathrm{SLQ}_i = \frac{x_i^r/x^r}{x_i^n/x^n} = \frac{x_i^r}{x_i^n} \times \frac{x^n}{x^r}. \tag{3}$$

Here, x_i^r and x_i^n are the total output (production) of the ith regional and national sector, respectively, while x^r and x^n are the corresponding regional and national aggregates. SLQ_i measures the degree of specialization of region r in sector i relative to the nation. The regional input coefficients are derived according the following rule:

$$\hat{a}_{ij}^r = \begin{cases} a_{ij}^n \mathrm{SLQ}_i & \text{if } \mathrm{SLQ}_i < 1 \\ a_{ij}^n & \text{if } \mathrm{SLQ}_i \geq 1 \end{cases}. \tag{4}$$

However, it has long been known that the SLQ tends to underestimate a region's imports from other regions; this understatement occurs as the SLQ rules out any 'cross-hauling' (Stevens et al. 1989). Cross-hauling takes place when a region simultaneously imports and exports a given commodity.

The *cross-industry* LQ was one of the first refinements of the SLQ, as it considers the relative size of both supplying sector i and purchasing sector j. The formula is as follows:

$$\mathrm{CILQ}_{ij} = \frac{\mathrm{SLQ}_i}{\mathrm{SLQ}_j} = \frac{x_i^r/x_i^n}{x_j^r/x_j^n}, \tag{5}$$

where the constraints are applied as in (4). Unlike the SLQ, however, the CILQ applies a cell-by-cell adjustment. This means that it does, in principle at least, deal with the problem of cross-hauling.[1] What it does not do is to consider the relative size of a region, x^r/x^n, which cancels out in formula (5). By contrast, this ratio remains a component of the SLQ formula (3).

Round (1978) argues that any adjustment formula should incorporate three elements: (i) the relative size of the supplying sector i; (ii) the relative size of the purchasing sector j and (iii) the relative size of the region. The CILQ satisfies (i) and (ii) but not (iii), whereas the SLQ satisfies (i) and (iii) but not (ii). Round therefore suggests the following formula, which simultaneously satisfies all three requirements:

[1] Consider a region where $\mathrm{SLQ}_1 = 0.8$, $\mathrm{SLQ}_2 = 1.2$, $\mathrm{SLQ}_3 = 0.6$ and $\mathrm{SLQ}_4 = 1.5$, so that $\mathrm{CILQ}_{1,1} = 1$, $\mathrm{CILQ}_{1,2} = 0.\dot{6}$, $\mathrm{CILQ}_{1,3} = 1.\dot{3}$, $\mathrm{CILQ}_{1,4} = 0.5\dot{3}$, etc. For the SLQ to be valid, this region would need to be an importer but not an exporter of commodities 1 and 3, and vice versa for commodities 2 and 4. The CILQ would encompass a wider set of possibilities. For instance, industries 2 and 4 could import but not export commodity 1, yet this commodity could be exported but not imported by industry 3; consequently, cross-hauling of commodity 1 could occur. In contrast, only exporting of commodity 4 would be possible because $\mathrm{CILQ}_{4j} \geq 1$ for all j.

$$\mathrm{RLQ}_{ij} = \frac{\mathrm{SLQ}_i}{\log_2\left(1 + \mathrm{SLQ}_j\right)}. \tag{6}$$

Nonetheless, Flegg et al. (1995) criticize the SLQ and RLQ on the grounds that they would both tend to underestimate the imports of relatively small regions owing to the way in which the ratio x^r/x^n is implicitly incorporated in each formula. To overcome this drawback, the FLQ was introduced.

The crucial hypothesis underpinning the FLQ is that a region's propensity to import from other domestic regions is inversely and nonlinearly related to its relative size. By incorporating explicit adjustments for regional size, the FLQ should yield more precise estimates of regional input coefficients and hence multipliers.[2] Along with other non-survey methods, the FLQ aims to offer regional analysts a means by which they can build regional tables that reflect, as closely as possible, each region's economic structure. It is defined as follows (cf. Flegg and Webber 1997):

$$\mathrm{FLQ}_{ij} = \begin{cases} \mathrm{CILQ}_{ij}\lambda & \text{for } i \neq j \\ \mathrm{SLQ}_{ij}\lambda & \text{for } i = j \end{cases}, \tag{7}$$

where λ captures a region's relative size. This scalar is defined as follows:

$$\lambda = \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]^{\delta}. \tag{8}$$

Here $0 \leq \delta < 1$ is a parameter that controls the degree of convexity in Eq. (8). The larger the value of δ, the lower the value of λ, and the greater the allowance for extra regional imports. The FLQ formula is implemented just like other LQ methods, so:

$$\hat{a}^r_{ij} = \begin{cases} a^n_{ij}\mathrm{FLQ}_{ij} & \text{if } \mathrm{FLQ}_{ij} < 1 \\ a^n_{ij} & \text{if } \mathrm{FLQ}_{ij} \geq 1 \end{cases}. \tag{9}$$

Many case studies, including those mentioned earlier, have demonstrated that the FLQ can yield more accurate results than the SLQ and CILQ. This evidence is corroborated by the Monte Carlo study of Bonfiglio and Chelli (2008). Nonetheless, some conflicting evidence is presented by Lamonica and Chelli (2017), who find initially that the SLQ gives slightly better results than the FLQ.

Lamonica and Chelli's unusual study employs data from the World Input–Output Database, whereas other studies have analysed data for individual countries or used Monte Carlo methods (Bonfiglio and Chelli 2008). Lamonica and Chelli examined data for the period 1995–2011, classified into 35 economic sectors. Their sample comprised 27 European countries and 13 other major countries, with the rest of the world as a composite 'country'. However, when this sample was disaggregated by size of economy, an interesting divergence appeared. For the smaller economies, characterized by a high

[2] Based on an analysis of data for all Chinese provinces, Okamoto (2014) casts some doubt on this hypothesis. However, the sheer size and diversity of China make it difficult to apply non-survey methods successfully. Furthermore, the intermediate inputs in Chinese regional tables include imports, whereas the FLQ requires such imports to be excluded. Okamoto attempted to circumvent this problem by calculating regional self-sufficiency rates, based on the assumption that imports were determined by regional demand. While this assumption is reasonable, its use is bound to introduce some inaccuracy.

percentage of near-zero input coefficients, the FLQ (with $\delta = 0.2$) was found to be the best method, whereas the SLQ performed the best in the larger economies.

McCann and Dewhurst (1998) criticize the FLQ on the grounds that regional coefficients may surpass national coefficients where there is regional specialization, a possibility that is precluded by the FLQ formula. Flegg and Webber (2000) therefore proposed the following *augmented* FLQ:

$$\mathrm{AFLQ}_{ij} = \begin{cases} \mathrm{FLQ}_{ij}\left[\log_2\left(1+\mathrm{SLQ}_j\right)\right] & \text{for } \mathrm{SLQ}_j > 1 \\ \mathrm{FLQ}_{ij} & \text{for } \mathrm{SLQ}_j \leq 1 \end{cases}, \tag{10}$$

where $\log_2\left(1+\mathrm{SLQ}_j\right)$ captures the regional specialization of sector j. If $\mathrm{SLQ}_j > 1$ and $\mathrm{FLQ}_{ij} \geq 1$, the national coefficients are scaled upwards. However, to avoid excessive upward adjustments, the restriction $\mathrm{FLQ}_{ij} \leq 1$ is retained. Hence the regionalization proceeds as follows:

$$\hat{a}^r_{ij} = \begin{cases} a^n_{ij}\mathrm{AFLQ}_{ij} & \text{if } \mathrm{SLQ}_j > 1 \\ a^n_{ij}\mathrm{FLQ}_{ij} & \text{if } \mathrm{SLQ}_j \leq 1 \end{cases}, \tag{11}$$

subject to $\mathrm{FLQ}_{ij} \leq 1$.

While the AFLQ has the theoretical merit of incorporating a measure of regional specialization, it tends to produce similar outcomes to the FLQ (Bonfiglio and Chelli 2008; Flegg and Tohmo 2013; Flegg and Webber 2000; Flegg et al. 2016; Kowalewksi 2015). Indeed, the analysis in Appendix demonstrates that the AFLQ does not yield more accurate estimates of input coefficients than the FLQ, although it does perform better in terms of multipliers.[3]

Another variant of the FLQ is put forward by Kowalewksi (2015). Using output rather than employment to measure regional size, her *industry-specific* FLQ can be defined as:

$$\mathrm{SFLQ}_{ij} \equiv \mathrm{CILQ}_{ij} \times \left[\log_2(1 + x^r/x^n)\right]^{\delta j}. \tag{12}$$

The novel feature of this formula is that δ is allowed to vary across industries. This greater realism is undoubtedly an attractive feature, yet it does introduce much greater complexity into the modelling process (Flegg and Tohmo 2019). For that reason, we do not consider the SFLQ further.

To complete this review of pure LQ-based methods, we should note that the FLQ's focus is on the output and employment generated within a given region. Consequently, as Flegg and Tohmo (2019) emphasize, it should only be applied to NIOTs where imports are excluded from the inter-industry transactions (type B tables). By contrast, where the focus is on the total supply of commodities, Kronenberg's Cross-Hauling Adjusted Regionalization Method (CHARM) is an appropriate technique (Többen and Kronenberg 2015; Flegg and Tohmo 2018). Unlike the FLQ, however, CHARM requires type A tables, where the national transactions include imports.[4]

[3] Lampiris et al. (2020) use Eurostat data for 2010 and 2014 to assess the performance of various pure LQ-based formulae in a sample of 18 countries. For this data set, the AFLQ generally outperforms the FLQ, although the differences in outcomes are typically very small. δ in the range 0.1 to 0.3 gives the best results.

[4] See Fujimoto (2019, p. 108–111) for a very clear explanation of how the Japanese type A tables can be converted into type B tables for purposes of estimation.

3 Review of some alternative approaches

Here, we review three recent studies that offer alternatives to the pure LQ-based methods discussed previously.[5] The first is an innovative study by Fujimoto (2019), who examines official survey-based data for nine Japanese regions in 2005, the most recent year available in a series of official tables published quinquennially since 1960. This study's focus is on cross-hauling and its primary aim is to determine which of four alternative assumptions is most appropriate. Each assumption is associated with a particular modelling approach, as follows:

1. There is no cross-hauling (LQ approach);
2. Cross-hauling depends on regional size (FLQ approach);
3. Cross-hauling is proportional to its potential, as measured by output or demand (RCHARM);
4. Cross-hauling is proportional to its potential, as measured by the volume of trade (MCHARM).

To represent the 'LQ approach', Fujimoto rejects the SLQ in favour of the scaling formula

$$\mathrm{DSLQ}_i^r = \frac{X_i^r}{D_i^r} \Big/ \frac{X_i^n}{D_i^n}, \tag{13}$$

where X denotes output and D denotes demand. The rationale for using this alternative formula is to overcome aggregation bias (Fujimoto 2019, p. 113).

For the second approach, Fujimoto employs the formula:

$$\mathrm{FLQ}_i^r \equiv \mathrm{DSLQ}_i^r \times \left[\log_2(1 + v^r/v^n)\right]^{\delta}, \tag{14}$$

where v^r/v^n is the ratio of total regional to total national value-added payments. However, the author does not explain why DSLQ_i^r is used instead of CILQ_{ij}^r nor why value added is used as a proxy for regional size rather than superior measures such as output or employment. It is, therefore, misleading to refer to this approach as the 'FLQ approach'. The third approach is the refined version of CHARM developed by Többen and Kronenberg (2015), while the fourth is the modified version of CHARM proposed by Fujimoto.

Fujimoto (2019, p. 115) remarks that '[t]he FLQ approach has a problem in addition to the difficulty of [specifying] a value for δ: the cross-hauling caused in interregional trade depends not only on regional size.' To demonstrate this, he derives the following equation for $\mathrm{DSLQ}_i^r < 1$:

$$\Delta E_i^r = \Delta M_i^r = \mathrm{DSLQ}_i^r(1-\lambda)\left(1 - m_i^n\right)D_i^r, \tag{15}$$

[5] Boero et al. (2018) use US county-level data on demand and supply, along with measures of transport costs, to estimate trade flows. They develop a way of estimating regional tables and trade flows simultaneously, thereby making it possible to obtain more precise estimates. Although this interesting new procedure seems to yield reasonably accurate results, the authors note (p. 236) that it is computationally burdensome, especially where the focus is on a single county. For that reason, along with the expectation that it would be more useful in the USA than elsewhere, we opted not to discuss it further.

where E and M denote exports and imports, respectively, m_i^n is the national propensity to import from abroad and D is demand. Fujimoto adds that there is 'no reason why cross-hauling [should] depend on DSLQ_i^r and m_i^n [and] this dependence causes serious bias …' (p. 116).

However, a straightforward interpretation can be given to the role of both DSLQ_i^r and m_i^n in Eq. (15). Since it is assumed that $m_i^r = m_i^n$, the term $(1 - m_i^n)$ captures the proportion of regional demand D_i^r that is met by domestic suppliers, some of which are located in region r and the rest in other regions. As expected, there is a negative relationship between m_i^n and ΔM_i^r, *ceteris paribus*.

To explain the positive relationship between DSLQ_i^r and ΔM_i^r, we note that ΔM_i^r represents the difference between the extra imports generated by using FLQ_i^r and those generated by applying DSLQ_i^r. Although this difference is invariably positive, its magnitude increases as DSLQ_i^r rises. This is due to the inclusion of λ in FLQ_i^r. The upshot is that there is a positive relationship between DSLQ_i^r and ΔM_i^r.[6]

The above discussion suggests that Fujimoto has failed to identify a genuine problem with the FLQ approach. Nonetheless, after performing various statistical tests, using data for 106 sectors and nine regions, he rejects this approach on the grounds (i) that it yields biased estimates of regional imports and (ii) that the appropriate value of δ is unknown and depends on each case.

However, since Fujimoto does not use the correct FLQ formula, his findings do not constitute a valid test of the FLQ approach. It is also likely that superior estimates could have been obtained by using a different δ for each Japanese region. In particular, the islands of Hokkaido and Okinawa may well require different values of δ from the mainland regions. The approach discussed later in this paper affords a way of generating such region-specific values.

As regards the other approaches, the scatter diagrams of estimated and survey-based import propensities (Fujimoto 2019, figure 2) reveal an almost identical pattern for the DSLQ and RCHARM methods, with evidence of substantial and widespread underestimation. By contrast, the diagram for MCHARM suggests a more random distribution, albeit with a greater variance and some heteroscedasticity. Even so, using the mean absolute error as the criterion, RCHARM invariably outperforms MCHARM (Fujimoto 2019, table 2). The DSLQ is clearly in third place.[7]

In another recent study, Pereira-López et al. (2020) focus on the AFLQ rather than the FLQ. They argue persuasively that regional specialization can have different effects on the columns (cost structure) and rows (selling structure) of a regional coefficient matrix. For instance, a region that is specialized in the extraction of mining products may sell most of its output to processing sectors such as metal industries located in other regions.

[6] To illustrate, let $\lambda = 0.8$ and $m_i^n = 0.2$. Also, let $\mathrm{DSLQ}_i^r = 0.6$ initially (case A) but rise to 0.8 (case B).
Case A. Using FLQ_i^r, $M_i^r = [1 - 0.6 \times 0.8 \times 0.8]D_i^r = 0.616D_i^r$. By contrast, with DSLQ_i^r, we get $M_i^r = [1 - 0.6 \times 0.8]D_i^r = 0.52D_i^r$. Hence $\Delta M_i^r = (0.616 - 0.52)D_i^r = 0.096D_i^r$, so the extra regional imports due to using FLQ_i^r rather than DSLQ_i^r are 9.6% of regional demand.
Case B. Using FLQ_i^r, $M_i^r = [1 - 0.8 \times 0.8 \times 0.8]D_i^r = 0.488D_i^r$. By contrast, with $DSLQ_i^r$, we get$M_i^r = [1 - 0.8 \times 0.8]D_i^r = 0.36\,D_i^r$. Hence $\Delta M_i^r = (0.488 - 0.36)D_i^r = 0.128\,D_i^r$, so the extra regional imports are now 12.8% of regional demand.

[7] This analysis could be enhanced by weighting the import propensities by regional size. In addition, the root mean squared error could be used instead of the mean absolute error; this would capture some very large errors evident in the scatter diagram for MCHARM, and allow the overall error to be decomposed into bias, variance and covariance components (Stevens et al. 1989).

The AFLQ has the limitation that it presumes that specialization only affects purchasing sectors (columns). The authors thus propose a bidimensional procedure, the 2DLQ, whereby a parameter α is used to adjust the rows, while another parameter β is applied to the columns. The regionalization employs the following relationships:

$$\hat{a}_{ij}^r = \begin{cases} (SLQ_{\mathrm{i}})^{\alpha} a_{ij}^n (x_j^r / x_j^n)^{\beta} & \text{if } SLQ_i \leq 1 \\ [0.5 \tan \mathrm{h}(SLQ_{\mathrm{i}} - 1) + 1]^{\alpha} a_{ij}^n (x_j^r / x_j^n)^{\beta} & \text{if } SLQ_{\mathrm{i}} > 1 \end{cases}, \tag{16}$$

where (x_j^r / x_j^n) measures the relative size of purchasing sector j. The hyperbolic tangent function (tan$h$) allows the estimated regional coefficients to be 'slightly higher' than the corresponding national coefficients if $\mathrm{SLQ}_i > 1$ (Pereira-López et al. 2020, p. 480). They make the interesting observation that the CILQ, FLQ and AFLQ are all nested within the 2DLQ: $\mathrm{FLQ}_{ij} = \mathrm{2DLQ}_{ij}$ if $\alpha = (\ln \lambda\ \mathrm{SLQ}_i / \ln \mathrm{SLQ}_i)$ and $\beta = 1$, $\mathrm{FLQ}_{ij} = \mathrm{AFLQ}_{ij}$ if $\mathrm{SLQ}_j \leq 1$, and $\mathrm{FLQ}_{ij} = \mathrm{CILQ}_{ij}$ if $\delta = 0$.

Equation (16) has the same data requirements as the AFLQ but the authors claim that it makes more efficient use of such data and consequently yields more accurate results. To test this claim, they use the Eurostat IO database for 2005 to extract symmetric 59×59 domestic coefficient matrices for:

1. Austria, Belgium, France, Germany, Italy and Spain (the 'observed' matrices);
2. the European Area 17 (EA17) (the parent table).

Two estimated coefficient matrices are derived for each country by applying the 2DLQ and AFLQ formulae to the parent table. These estimates are then compared by using the weighted absolute percentage error (WAPE) and two other statistics, U and U^*, which consider the number of cells (n^2) and the number of non-empty cells ($n^2 - z$), respectively. WAPE is defined as follows:[8]

$$\mathrm{WAPE} = 100 \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \hat{a}_{ij}^r - a_{ij}^r \right| \Big/ \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^r. \tag{17}$$

The results for coefficients show that using the 2DLQ method reduces the WAPE for all countries, most noticeably for Austria and Italy (Pereira-López et al. 2020, table 2). However, there is little change in the outcomes for Belgium, France and Germany. Spain is an intermediate case. On average, the WAPE is lowered by 4.5%. The U and U^* measures yield similar results. The authors also assess the performance of the 2DLQ in terms of multipliers. Here the 2DLQ method gives better results than the AFLQ for all countries apart from Belgium. The average improvement is 3.55%.[9]

Of the two studies reviewed thus far, the 2DLQ method seems the most promising, when evaluated in terms of its theoretical foundations, empirical performance and ease of application. Even so, some caveats should be borne in mind. The first concerns the

[8] The U and U^* statistics are wrongly attributed to Zhao and Choi (2015), who do not use them. See Flegg and Tohmo (2013, p. 715) for a range of statistics where $(n^2 - z)$ is used as a divisor.

[9] In a subsequent study, again using Eurostat data but with a larger sample, Pereira-López et al. (2021) obtain encouraging results for the 2DLQ *vis-à-vis* the FLQ, AFLQ and CILQ. However, it should be noted that the 2DLQ formula involves two unknown parameters, which introduces greater complexity into the analysis compared with the other three formulae.

need to determine suitable values of α and β. Here the authors provide some reassurance that the range of suitable values of these parameters is relatively small and that analysts would not go far wrong by choosing an α of 0.1 or 0.15 and a β in the range 0.8 to 1.2 (Pereira-López et al. 2020, table 5).

The second caveat concerns the fact that countries rather than regions are used in the testing process, which poses some potential problems. While the authors are right to stress the quality and consistency of the Eurostat data they employ, it is not true to say that using countries instead of regions is the only possible way to proceed. For instance, suitable survey-based official regional and national data are available for Finland (Flegg and Tohmo 2013), South Korea (Jahn et al. 2020) and Japan (Fujimoto 2019). It would be instructive to re-examine the 2DLQ method by using one or more of such data sets to perform a sensitivity analysis.

The study by Lahr et al. (2020) represents a radical departure from those discussed hitherto. The authors examine data for 2014 from the World Input–Output Database, pertaining to 23 manufacturing sectors in 28 EU countries. A novel feature of this study is its use of quasi-binomial econometric models, in which the regional purchase coefficient (RPC) is regressed on various variables.[10] Each country is treated as a 'region' and the aggregate of all 28 countries as the 'nation'. Accordingly, RPC_i^c is the proportion of national requirements of industry i supplied by firms located within country c.

The following regressors were found to be statistically significant ($p = 0.001$, two-tailed test):

- SLQ
- ln (land area)
- ln (hotel room-nights/area)
- ln (weight/value)
- supply/demand ratio (SDR).

In addition, six industry-specific binary variables were included. $R^2 = 0.660$. By contrast, a model with the SLQ alone gave $R^2 = 0.142$. This worse fit is unsurprising, since the SLQ rules out the key factor of cross-hauling and cannot allow for the peculiarities of specific industries.

Lahr et al. also carry out a test of what is described as the FLQ method. This is done by regressing RPC_i^c on SLQ_i^c and the 'employment share'. $R^2 = 0.195$. However, this test is inconsistent in several respects with the FLQ approach. Most importantly, the FLQ has a cross-industry foundation, which cannot be captured in a rows-only estimation. Consequently, no account is taken of the likelihood that purchasing industries would differ in their use of particular inputs. This aspect is captured in $CILQ_{ij}^c$ and hence in FLQ_{ij}^c but not in SLQ_i^c. We should note too that the FLQ formula is multiplicative, whereas the regression model is additive. It is also unclear how the regressor 'employment share' was measured.[11] Finally, whereas output was used to calculate SLQ_i^c, employment was used

[10] These variables are similar to those used in the pioneering studies of Stevens et al. (1983) and Treyz and Stevens (1985).

[11] It is unclear whether this regressor is simply each country's share of overall EU employment, e^c/e^{eu}, or, more appropriately, the nonlinear scalar λ used in the FLQ formula. However, we note that λ is incorrectly defined as $\delta \log_2(1 + e^c/e^{eu})$ rather than as $[\log_2(1 + e^c/e^{eu})]^\delta$ (cf. Lahr et al. 2020, p. 1590).

to measure regional size. That would affect the results to the extent that productivity differed across EU countries.

The authors compare the performance of their econometric approach with that of the SLQ, SDR and CHARM. The procedures are judged in terms of their ability to estimate RPCs and to replicate each country's coefficient matrix, Leontief inverse and output multipliers. As expected, the SLQ and SDR yield similar results and, on average, both methods greatly overstate input coefficients (Lahr et al. 2020, table 5). By comparison, the regression-based approach necessarily yields a mean error of zero. The authors find that RCHARM performs 'somewhat better' than the SLQ and SDR, yet it still systematically overstates RPCs, with a mean error of 0.240 (Lahr et al. 2020, p. 1594). By contrast, MCHARM yields negative RPCs for 337 of 1568 national industries. Therefore, both variants of CHARM have serious demerits as a means of estimating RPCs.

A crucial consideration when selecting a non-survey method is its ability to yield unbiased estimates of input coefficients. A regression-based approach can be relied upon to perform very well in that respect but it is also possible to obtain unbiased estimates via the FLQ approach, so long as an appropriate value of the parameter δ is used.

A drawback of the RPC approach vis-à-vis the FLQ is its more demanding data requirements.[12] In the model discussed above, for instance, it would be challenging to find data for some of the regressors, whereas the FLQ only requires figures for output (or employment) in each regional and national industry.

On the other hand, the FLQ has often been criticized on the basis that the results obtained from one country or region are not necessarily transferable elsewhere, since the optimal δ would differ. This problem is addressed in the present paper via a procedure whereby country-specific and region-specific values of δ can be derived. Lahr et al. (2020, p. 1591) note that econometric approaches face a similar challenge in terms of transferability of results.

While the model constructed by Lahr et al. sheds some helpful light on the determinants of RPCs in EU countries, it would be interesting to see how well it would perform when constructing a RIOT for, say, Catalonia from a Spanish NIOT.

4 The CE approach to regionalizing NIOTs

Here, we explain our MCE approach to regionalizing NIOTs. Various mathematical programming methods based on a constrained optimization framework exist. These typically minimize a penalty function, which measures the deviation of the balanced matrix from the initial matrix, subject to a set of balancing conditions. We limit our attention to the CE method, which is one of the pioneering methods. It is widely used and Lamonica et al. (2020) have shown that it performs very well when applied to countries with small economies. Moreover, in contrast to other constrained optimization methods, it is stable in the sense that the MAD index, i.e. the sum of the absolute differences between the observed and estimated input coefficients, does not change abruptly from one economy to the next. Furthermore, unlike the standard RAS method, which can only handle non-negative matrices, the CE method can easily

[12] Szabó (2015, p. 50) remarks that '[d]espite its theoretical advantages, the [RPC] approach did not gain popularity [owing] to its high data requirements, which usually cannot be satisfied at the regional level.'

be adapted to deal with matrices with negative entries. Although we could have used GRAS to deal with this issue, we wished to avoid the problem noted in the Introduction regarding this technique.

Referring to Lamonica et al. (2020) for details and treating the primary input $\mathbf{v}^n$ and the final demand $\mathbf{f}^n$ as 'additional' intermediate input and output, the method starts with the following augmented input coefficients matrix $\mathbf{A}^n$:

$$\mathbf{A}^n = \begin{bmatrix} a^n_{1,1} = x^n_{1,1}/x^n_1 & \cdots & a^n_{1,k} = x^n_{1,k}/x^n_k & a^n_{1,k+1} = f^n_1/x^n_{k+1} \\ \vdots & & \vdots & \vdots \\ a^n_{k,1} = x^n_{k,1}/x^n_1 & \cdots & a^n_{k,k} = x^n_{k,k}/x^n_k & a^n_{k,k+1} = f^n_k/x^n_{k+1} \\ a^n_{k+1,1} = v^n_1/x^n_1 & \cdots & a^n_{k+1,k} = v^n_k/x^n_k & a^n_{k+1,k+1} = 0 \end{bmatrix}. \tag{18}$$

It turns out that:

$$\mathbf{A}^n\mathbf{x}^n = \mathbf{x}^n, \tag{19}$$

and

$$\mathbf{u}'\mathbf{A}^n = \mathbf{u}, \tag{20}$$

where $\mathbf{u}$ is a unitary vector and $\mathbf{u}'$ is its transpose.

The task at hand is to generate a new matrix $\mathbf{A}^r$ from the existing $\mathbf{A}^n$, with the same dimension as before, but respecting new row and column totals $\mathbf{x}^r$. More specifically, we seek a matrix $\mathbf{A}^r$ that satisfies the following consistency and additivity conditions:

$$\mathbf{A}^r\mathbf{x}^r = \mathbf{x}^r, \tag{21}$$

$$\mathbf{u}'\mathbf{A}^r = \mathbf{u}. \tag{22}$$

Formally, the problem involves minimizing the following function for $a^n_{ij} > 0$:

$$\underset{a^r_{ij}}{\text{Min}}\, H = \sum_{i=1}^{k+1}\sum_{j=1}^{k+1} a^r_{ij} \ln \frac{a^r_{ij}}{a^n_{ij}}; \tag{23}$$

subject to

$$\sum_{j=1}^{k+1} a^r_{ij}x^r_j = x^r_i, \tag{24}$$

and

$$\sum_{i=1}^{k+1} a^r_{ij} = 1. \tag{25}$$

The solution is obtained by solving the following Lagrangian function for problem (23):

$$L = \sum_{i=1}^{k+1}\sum_{j=1}^{k+1} a_{ij}^r \ln \frac{a_{ij}^r}{a_{ij}^n} + \sum_{i=1}^{k+1} \lambda_i \left(x_i^r - \sum_{j=1}^{k+1} a_{ij}^r x_j^r \right) + \sum_{j=1}^{k+1} \mu_j \left(1 - \sum_{i=1}^{k+1} a_{ij}^r \right). \tag{26}$$

The first-order optimal conditions are:

$$\frac{\partial L}{\partial a_{ij}^r} = \ln \frac{a_{ij}^r}{a_{ij}^n} + 1 - \lambda_i x_j^r - \mu_j = 0,\ i,j = 1, 2, \ldots, k+1; \tag{27}$$

$$\frac{\partial L}{\partial \lambda_i} = x_i^r - \sum_{j=1}^{k+1} a_{ij}^r x_j^r = 0,\ i = 1, 2, \ldots, k+1; \tag{28}$$

$$\frac{\partial L}{\partial \mu_j} = 1 - \sum_{i=1}^{k+1} a_{ij}^r = 0,\ j = 1, 2, \ldots, k+1. \tag{29}$$

Solving this system of equations yields:

$$a_{ij}^r = a_{ij}^n e^{\left(-1+\lambda_i x_j^r + \mu_j\right)},\ i,j = 1, 2, \ldots, k+1; \tag{30}$$

$$\sum_{j=1}^{k+1} x_j^r a_{ij}^n e^{\left(-1+\lambda_i x_j^r + \mu_j\right)} = x_i^r,\ i = 1, 2, \ldots, k+1; \tag{31}$$

$$\sum_{i=1}^{k+1} a_{ij}^n e^{\left(-1+\lambda_i x_j^r + \mu_j\right)} = 1,\ j = 1, 2, \ldots, k+1. \tag{32}$$

We should note that the estimates of a_{ij}^r depend on the values of the Lagrangian multipliers, which must be determined by solving the nonlinear system (31) and (32) for the unknowns λ_1, λ_2, ..., λ_{k+1}. As no closed-form solution exists, this system is solved by using numerical algorithms.

However, any standard solver will fail to find a solution to Eqs. (31) and (32) when a_{ij}^n and x_i^r, $i, j = 1, 2, \ldots, k+1$, are values from real IOTs. In fact, the solution to these equations requires some attention for two main reasons: the evaluation of the exponential functions for large values of their arguments and the solution to a high-dimension system of nonlinear equations. These difficulties are addressed by adopting the solution proposed by Lamonica et al. (2020). Since the KIRIOT shows null and negative entries, we adapt the CE method to account for negative entries in the NIOT.

To this end, we revise problem (23) by assuming that the sign of the national matrix is preserved in the regional matrix (i.e. $a_{ij}^n < 0$ implies $a_{ij}^r < 0$, and $a_{ij}^n > 0$ implies $a_{ij}^r > 0$), while a null entry in the national matrix implies a null entry in the regional matrix (i.e. $a_{ij}^n = 0$ implies $a_{ij}^r = 0$). In the following, we refer to this assumption as the 'sign-preserving assumption'. We therefore formulate the following optimization problem:

$$\underset{a_{ij}^n}{\text{Min}}\, H = \sum_{i=1}^{k+1}\sum_{j=1}^{k+1} |a_{ij}^r| \ln \frac{a_{ij}^r}{a_{ij}^n}; \tag{33}$$

subject to

$$\sum_{j=1}^{k+1} a_{ij}^r x_j^r = x_i^r, \tag{34}$$

and

$$\sum_{i=1}^{k+1} a_{ij}^r = 1. \tag{35}$$

The solution is obtained by solving the following Lagrangian function for problem (33):

$$L = \sum_{i=1}^{k+1}\sum_{j=1}^{k+1} |a_{ij}^r| \ln \frac{a_{ij}^r}{a_{ij}^n} + \sum_{i=1}^{k+1} \lambda_i \left(x_i^r - \sum_{j=1}^{k+1} a_{ij}^r x_j^r \right) + \sum_{j=1}^{k+1} \mu_j \left(1 - \sum_{i=1}^{k+1} a_{ij}^r \right). \tag{36}$$

We consider the set of pairs $(i, j) \in \Gamma = \{1, 2, \ldots, k+1\} \times \{1, 2, \ldots, k+1\}$ and its partition $\Gamma = \Gamma^+ \cup \Gamma^-$, where the sets Γ^+ and Γ^- are defined as: $\Gamma^+ = \{(i,j) | a_{ij}^r > 0\}$ and $\Gamma^- = \{(i,j) | a_{ij}^r < 0\}$. The Lagrangian can be rewritten as

$$L = \sum_{(i,j)\in\Gamma^+}^{k+1} a_{ij}^r \ln \frac{a_{ij}^r}{a_{ij}^n} - \sum_{(I,j)\in\Gamma^-}^{k+1} a_{ij}^r \ln \frac{a_{ij}^r}{a_{ij}^n} + \sum_{i=1}^{k+1} \lambda_i \left(x_i^r - \sum_{j=1}^{k+1} a_{ij}^r x_j^r \right) + \sum_{j=1}^{k+1} \mu_j \left(1 - \sum_{i=1}^{k+1} a_{ij}^r \right). \tag{37}$$

The first-order optimal conditions are:

$$\frac{\partial L}{\partial a_{ij}^R} = \ln \frac{a_{ij}^r}{a_{ij}^n} + 1 - \lambda_i x_j^r - \mu_j = 0, \ (i,j) \in \Gamma^+; \tag{38}$$

$$\frac{\partial L}{\partial a_{ij}^R} = -\ln \frac{a_{ij}^r}{a_{ij}^n} + 1 - \lambda_i x_j^r - \mu_j = 0, \ (i,j) \in \Gamma^-; \tag{39}$$

$$\frac{\partial L}{\partial \lambda_i} = x_i^r - \sum_{j=1}^{k+1} a_{ij}^r x_j^r = 0, \ i = 1, 2, \ldots, k+1; \tag{40}$$

$$\frac{\partial L}{\partial \mu_j} = 1 - \sum_{i=1}^{k+1} a_{ij}^r = 0, \ j = 1, 2, \ldots, k+1. \tag{41}$$

Solving this system of equations yields:

$$a_{ij}^r = a_{ij}^n e^{\left(-1+\lambda_i x_j^r + \mu_j\right)}, \ (i,j) \in \Gamma^+; \tag{42}$$

Table 1 Simplified pattern of the KIRIOT

$\mathbf{Z}^{1;1}$	$\mathbf{Z}^{1;2}$	…	$\mathbf{Z}^{1;16}$	$\mathbf{C}^{1;1}$	…	$\mathbf{E}^{1;16}$	$\mathbf{x}^1$
$\mathbf{Z}^{2;1}$	$\mathbf{Z}^{2;2}$	…	$\mathbf{Z}^{2;16}$	$\mathbf{E}^{2;1}$	…	$\mathbf{E}^{2;16}$	$\mathbf{x}^2$
…	…	…	…	…	…	…	…
$\mathbf{Z}^{16;1}$	$\mathbf{Z}^{16;2}$	…	$\mathbf{Z}^{16;16}$	$\mathbf{E}^{16;1}$	…	$\mathbf{C}^{16;16}$	$\mathbf{x}^{16}$
$(\mathbf{v}^1)'$	$(\mathbf{v}^2)'$	…	$(\mathbf{v}^{16})'$				
$(\mathbf{x}^1)'$	$(\mathbf{x}^2)'$	…	$(\mathbf{x}^{16})'$				

$$a_{ij}^r = a_{ij}^n e^{\left(+1-\lambda_i x_j^r - \mu_j\right)},\ (i,j) \in \Gamma^-; \tag{43}$$

$$\sum_{(i,j)\in\Gamma^+}^{k+1} x_j^r a_{ij}^n e^{\left(-1+\lambda_i x_j^r + \mu_j\right)} + \sum_{(i,j)\in\Gamma^-}^{k+1} x_j^r a_{ij}^n e^{\left(+1-\lambda_i x_j^r - \mu_j\right)} = x_i^r,\ i = 1,2,\ldots,k+1; \tag{44}$$

$$\sum_{(i,j)\in\Gamma^+}^{k+1} a_{ij}^n e^{\left(-1+\lambda_i x_j^r + \mu_j\right)} + \sum_{(i,j)\in\Gamma^-}^{k+1} a_{ij}^n e^{\left(+1-\lambda_i x_j^r - \mu_j\right)} = 1,\ j = 1,2,\ldots,k+1. \tag{45}$$

We solve Eqs. (44) and (45) with the Matlab solver *fsolve*, by using the scaling procedure as in Lamonica et al. (2020) to overcome the numerical challenges of these equations.

5 Estimation of δ and the FLQ+ approach

We now explain the method employed to estimate δ. We consider the more detailed version of the South Korean interregional input–output table (KIRIOT) for the year 2005, which was constructed by the Bank of Korea for 16 regions and 78 sectors.

Table 1 illustrates a simplified form of the KIRIOT, where:

- $\mathbf{Z}^{r;r}$ ($r=1, 2, \ldots, 16$) is a 78×78 matrix whose elements $\left(z_{ij}^{r;r}\right)$ are the flows for intermediate use from sector i to sector j of region r;
- $\mathbf{Z}^{r;k}$ ($r, k=1, 2, \ldots, 16$ and $r \neq k$) is a 78×78 matrix whose elements $\left(z_{ij}^{r;k}\right)$ are the exports for intermediate use from sector i of region r to sector j of region k;
- $\mathbf{C}^{r;r}$ ($r=1, 2, \ldots, 16$) is a 78×5 matrix of the domestic final demand in region r;
- $\mathbf{E}^{r;k}$ ($r, k=1, 2, \ldots, 16$ and $r \neq k$) is a 78×5 matrix of the exports of region r for final demand in region k;
- $\mathbf{x}^r$ ($r=1, 2, \ldots, 16$) is a 78×1 vector whose elements are the sectoral output of region r;
- $(\mathbf{v}^r)'$ ($r=1, 2, \ldots, 16$) is a 1×78 vector whose elements are the sectoral value added plus the primary sectoral input of region r, inclusive of imports from abroad.

Using the South Korean NIOT, the national augmented matrix of input coefficients, $\mathbf{A}^n = [a_{ij}^n]$, was determined as follows. First, let:

- $\mathbf{X}^n = \sum_{r=1}^{16}\sum_{r=1}^{16}\mathbf{Z}^{r;r}$;
- $\mathbf{f}^n = \left(\sum_{r=1}^{16}\mathbf{C}^{r;r} + \sum_{r=1}^{16}\sum_{k=1}^{16}\mathbf{E}^{r;k}\right)$;
- $\mathbf{x}^n = \sum_{r=1}^{16}\mathbf{x}^r$;
- $(\mathbf{v}^n)' = \sum_{r=1}^{16}(\mathbf{v}^r)'$.

Next, consider the following national table, $\mathrm{IOT}^{\mathrm{n}}$, in block-matrix notation:

$$\mathbf{IOT}^n = \begin{bmatrix} \mathbf{X}^n & \mathbf{f}^n \\ (\mathbf{v}^n)' & 0 \end{bmatrix}. \tag{46}$$

Hence, we can determine the following national augmented matrix of input coefficients:

$$\mathbf{A}^n = \mathbf{IOT}^n\mathbf{D}(\mathbf{x}^n)^{-1}, \tag{47}$$

where $\mathbf{D}(\mathbf{x}^{\mathrm{n}})$ is the diagonal matrix whose elements are $d_{ii}(\mathbf{x}^n) = x_i$, $i=1, 2, \ldots, k+1$.

We are now able to define the three steps of our hybrid method:

Step 1: From $\mathbf{A}^{\mathrm{n}}$ and the regional sectoral total output $\mathbf{x}^{\mathrm{r}}$, estimate the regional input coefficient matrices $\hat{\mathbf{A}}^r$ ($r=1, 2, \ldots, 16$), using the MCE method described in the previous section.

Step 2: Using $\hat{\mathbf{A}}^r$, estimate the parameter δ in the FLQ formula as follows:

Let $\alpha_{ij}=\mathrm{CILQ}_{ij}$ if $i\neq j$ and $\alpha_{ij}=\mathrm{SLQ}_i$ if $i=j$. Also, define $\beta\equiv\log_2\left(1+\frac{x^r}{x^n}\right)$. Hence, using Eqs. (7) and (8), we get:

$$\hat{a}_{ij}^r = a_{ij}^n\alpha_{ij}\beta^{\delta}. \tag{48}$$

Thus, for any pair $(\hat{a}_{ij}^r, a_{ij}^n)$ that satisfies the sign-preserving assumption, we set:

$$\log\left(\frac{\hat{a}_{ij}^r}{a_{ij}^n\alpha_{ij}}\right) = \delta(\beta) + \varepsilon ij, \tag{49}$$

where ε_{ij} is a random number with zero mean. Taking expectations, we can obtain an *estimate* of the optimal δ for each region simply by dividing the mean of the regressand by $\log(\beta)$, which is a given value for each region. See Table 2.

Step 3. Using $\mathbf{A}^{\mathrm{n}}$, estimate the entries of the regional matrix $\mathbf{A}^{\mathrm{r}}$, the a_{ij}^r, using the FLQ method with the estimated optimal parameter $\hat{\delta}$. We refer to this hybrid approach as the FLQ+ method. As before, the estimated input coefficients are denoted by $\hat{a}_{ij}^r$.

6 Analysis of input coefficients

To validate the proposed FLQ+ method, the estimated matrices of regional input coefficients are compared with the true ones, $\mathbf{A}^r = \left[a_{ij}^r = \frac{z_{ij}^{r;r}}{x_j}\right]$. Since the RIOT being considered contains negative and null flows, we assess its performance by using the mean

Table 2 Computation of δ from model (49). Source: authors' calculations

Region	Mean of regressand	log (β)	Estimated δ	MSE
Gyeonggi	− 0.6482	− 1.2003	0.54	0.388
Seoul	− 0.4804	− 1.2983	0.37	0.166
N Gyeongsang	− 0.5264	− 2.0246	0.26	0.095
S Gyeongsang	− 0.6640	− 2.1419	0.31	0.091
Ulsan	− 0.9298	− 2.1623	0.43	0.325
South Jeolla	− 0.5295	− 2.3021	0.23	0.072
S Chungcheong	− 0.8869	− 2.2742	0.39	0.269
Incheon	− 1.1510	− 2.3980	0.48	0.187
Busan	− 0.7473	− 2.4909	0.30	0.083
N Chungcheong	− 0.9054	− 3.0180	0.30	0.095
Daegu	− 0.7892	− 3.0353	0.26	0.064
North Jeolla	− 0.8951	− 3.0864	0.29	0.119
Gangwon	− 0.7976	− 3.3234	0.24	0.107
Gwangju	− 1.0590	− 3.3095	0.32	0.103
Daejeon	− 1.1905	− 3.4014	0.35	0.094
Jeju	− 0.8049	− 4.4716	0.18	0.018
Mean	− 0.8128	− 2.6211	0.31	0.140
V	0.32	0.32	0.30	0.100

MSE mean square error, *V* coefficient of variation

absolute difference index (MAD), as suggested by Wiebe and Lenzen (2016). It is defined as:

$$\mathrm{MAD} = \sum_{i=1}^{78}\sum_{j=1}^{78}\frac{\left|a_{ij}-\hat{a}_{ij}\right|}{78\times 78}. \tag{50}$$

The FLQ, as implemented here, is the optimal FLQ in the sense that δ is chosen on the basis of minimizing the MAD. Specifically, for each of the sixteen regions, we compute the MAD of the FLQ for 99 different values of δ in the interval [0, 1], in increments of 0.01. The optimal δ is the value yielding minimum MAD. Determined in this way, the regional δ values vary from 0.22 to 0.48, with a mean of 0.35 (see Table 3). We refer to these values as the true or optimal values of δ. The characteristics of these regions are illustrated in Table 3 and their locations are shown in Fig. 1. For more details, see Flegg and Tohmo (2019) and Jahn et al. (2020).

Before examining the results, it may be helpful to explore why the optimal values of δ in Table 3 vary noticeably across regions. The role of δ is to adjust for any differences in regional propensities to import from other regions or from abroad that cannot be explained solely by differences in regional size. Now consider the following regression model fitted using data from Table 3:

$$\ln\delta = -3.154 + 0.582\ln R + 1.040\ln F + 2.515\ln V + e, \tag{51}$$

where R is the share of gross output imported from other regions; F is the share imported from abroad; V is the share of value added in gross output; e is a residual. All regressors are statistically significant at the 1% level (one-tailed tests) and the model easily passes all χ^2 diagnostic tests. $R^2 = 0.677$.

Table 3 Characteristics of South Korean regions in 2005. Source: adapted from Flegg and Tohmo (2019, table 9)

Region	Share of national output	Share of national employment	Share of intraregional inputs	Share of inputs from other regions	Share of inputs from abroad	Share of value added	Optimal value of δ
Gyeonggi	0.201	0.202	0.226	0.245	0.120	0.410	0.48
Seoul	0.182	0.254	0.237	0.173	0.060	0.529	0.32
North Gyeong-sang	0.084	0.054	0.247	0.254	0.163	0.336	0.31
South Gyeong-sang	0.073	0.067	0.223	0.284	0.125	0.369	0.32
Ulsan	0.071	0.025	0.202	0.240	0.283	0.275	0.41
South Jeolla	0.065	0.033	0.288	0.163	0.219	0.331	0.28
South Chun-gcheong	0.063	0.039	0.201	0.274	0.177	0.348	0.43
Incheon	0.055	0.048	0.175	0.288	0.171	0.366	0.45
Busan	0.051	0.074	0.200	0.266	0.077	0.457	0.32
North Chun-gcheong	0.029	0.030	0.181	0.307	0.104	0.408	0.30
Daegu	0.029	0.047	0.189	0.279	0.061	0.472	0.29
North Jeolla	0.027	0.032	0.192	0.304	0.074	0.430	0.34
Gangwon	0.022	0.029	0.198	0.230	0.044	0.528	0.22
Gwangju	0.022	0.028	0.165	0.307	0.099	0.430	0.37
Daejeon	0.019	0.027	0.133	0.281	0.065	0.520	0.44
Jeju	0.007	0.011	0.172	0.253	0.039	0.536	0.26
Mean	0.0625	0.0625	0.202	0.259	0.118	0.422	0.35
V	0.89	1.08	0.18	0.16	0.58	0.20	0.22

Shares are expressed as a proportion of gross output. Seoul is classified as a 'special city'; Busan, Daegu, Daejeon, Gwangju, Incheon and Ulsan as 'metropolitan cities'; Jeju as a 'special self-governing province'; and the rest as 'provinces'. The last column displays the optimal values of δ from Table 4

Equation (51) shows that regions with an above-average share of imports from other regions or from abroad require a bigger δ to compensate; and likewise if the share of value added in gross output is above average. Just over two thirds of the interregional variation in ln δ can be explained by these three factors, while almost one third must be ascribed to region-specific factors, differences in regional industrial structure, measurement errors, etc.

Table 4 displays the values of MAD (100×) for the MCE, FLQ with optimal δ, and FLQ+ methods. A key finding is that the MADs for the FLQ (both versions) are typically less than half of those for the MCE method. This result demonstrates the potential gains from employing the FLQ as a regionalization technique rather than applying the MCE method alone.

Another key finding is that the estimates of δ from the proposed FLQ+ method are mostly fairly close to the optimal values; the differences are invariably only in the second decimal place and range from −0.06 to 0.09. It is noticeable that the FLQ+ overestimates δ in the two biggest regions, Gyeonggi and Seoul. In the majority of regions, however, the estimates from this method are a little lower than the optimal values and, on

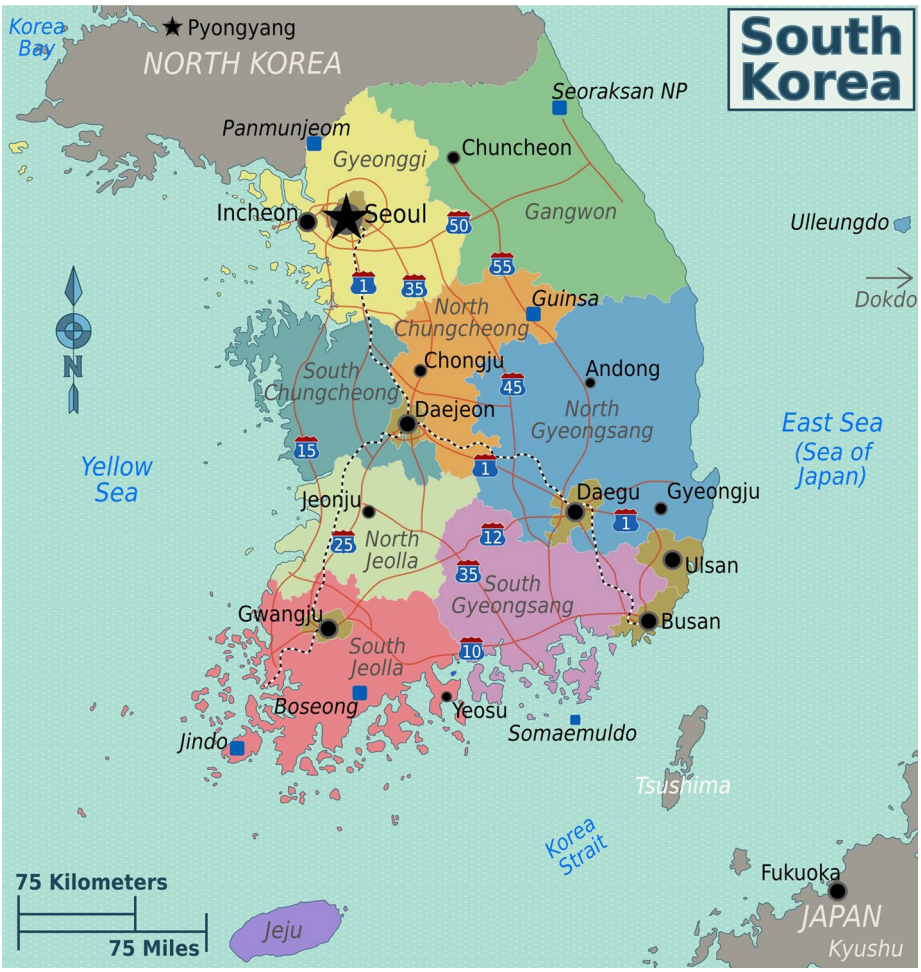

Fig. 1 South Korean regions. Source: South Korea regions map merged.png author: Peter Fitzgerald, NordNordWest; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported; available in Wikimedia Commons

average, δ is understated by 0.02. However, when the results are weighted by the regional share of national output, this outcome changes to a small overstatement on average. The biggest errors occur in the two smallest regions, Daejeon and Jeju. Table 3 reveals that the metropolitan city of Daejeon has an unusually low intraregional share of inputs, which could explain why the estimated δ is noticeably below the optimal value, with a discrepancy of 0.09. As for Jeju, the fact that it is a remote island can probably explain much of the understatement of δ by the FLQ+ method.

Table 5 provides some descriptive statistics of regional input coefficients estimated by the optimal FLQ and FLQ+ methods. Clearly, the two distributions are very close to each other in terms of central tendency and dispersion. Furthermore, they have an identical shape, as measured by the coefficients of kurtosis and skewness (not shown). The uniform behaviour of the two methods lends credence to the FLQ+ method proposed in this work, which is applicable when regional sectoral outputs are the only regional data available. By contrast, the optimal FLQ method requires knowledge of the entire regional matrix, so that an optimal value of δ can be computed for each region.

Table 4 MAD (×100) and δ values by region for the MCE, optimal FLQ and FLQ+ methods. Source: here and in the subsequent tables, all calculations were performed by the authors

Region	MCE	Optimal FLQ		FLQ+		Difference in δ
	MAD	MAD	Optimal δ	MAD	Estimated δ	FLQ − FLQ+
Gyeonggi	0.249	0.120	0.48	0.121	0.54	− 0.06
Seoul	0.330	0.147	0.32	0.148	0.37	− 0.05
N Gyeongsang	0.324	0.157	0.31	0.158	0.26	0.05
S Gyeongsang	0.254	0.143	0.32	0.143	0.31	0.01
Ulsan	0.354	0.162	0.41	0.162	0.43	− 0.02
South Jeolla	0.347	0.169	0.28	0.170	0.23	0.05
S Chungcheong	0.289	0.146	0.43	0.147	0.39	0.04
Incheon	0.305	0.132	0.45	0.132	0.48	− 0.03
Busan	0.295	0.148	0.32	0.148	0.30	0.02
N Chungcheong	0.270	0.148	0.30	0.148	0.30	0.00
Daegu	0.295	0.139	0.29	0.139	0.26	0.03
North Jeolla	0.305	0.136	0.34	0.137	0.29	0.05
Gangwon	0.393	0.203	0.22	0.203	0.24	− 0.02
Gwangju	0.339	0.134	0.37	0.134	0.32	0.05
Daejeon	0.338	0.134	0.44	0.138	0.35	0.09
Jeju	0.372	0.152	0.26	0.160	0.18	0.08
Mean	0.316	0.148	0.35	0.149	0.33	0.02
Weighted mean	0.303	0.144	0.37	0.145	0.38	− 0.01
V	0.13	0.13	0.22	0.13	0.29	2.47

Table 6 displays descriptive statistics for the absolute differences between the observed input coefficients and those estimated via the FLQ+ and GRAS+ methods. FLQ+ combines the MCE method, our preferred matrix-balancing approach, with the FLQ, whereas GRAS+ combines GRAS with the FLQ. The two methods produce similar results on the whole, although it is interesting that GRAS+ gives larger mean absolute differences than FLQ+ for the biggest eight regions, yet almost identical values for five of the six smallest regions. A possible explanation of this outcome is that the FLQ+ is apt to perform better when the percentage of zero flows is relatively low, as is the case for the larger regions. The two methods concur that Busan and Gangwon have, respectively, the smallest and biggest maximum absolute differences.[13] Taking a holistic view, the MAD is increased by 7.3% on average by using GRAS+ rather than FLQ+ and by 12.3% when the results are weighted by regional size. This finding, along with the theoretical criticisms noted earlier regarding the objective function employed in GRAS, lends support to our use of a modified CE method as the foundation of our proposed FLQ+ approach.

[13] We can confirm that the unusual figure of 124.6 for Gangwon in the second column of Table 6 is correct. It exceeds 100 because the observed intermediate flow from sector 6 to sector 22 is 14882.38, while its total output is 11664, which is less than 14,882.38 because the taxes on net production are 8820.29. Hence the input coefficient = 14,882.38/11664 = 1.276. The corresponding coefficient estimated via the FLQ+ method is 0.030, so the maximum difference is 100 (1.276 − 0.030) = 124.6. Sector 6 is Mining of coal, crude petroleum and natural gas, while sector 22 is Coke and hard coal.

Table 5 Descriptive indices (×100) for the estimated input coefficients from the optimal FLQ and FLQ+ methods

	Optimal FLQ				**FLQ+**			
	Mean	**Median**	**Q3**	**SD**	**Mean**	**Median**	**Q3**	**SD**
Gyeonggi	0.2517	0.0097	0.1066	1.1559	0.2346	0.0091	0.0994	1.0774
Seoul	0.2400	0.0052	0.0918	0.9174	0.2244	0.0049	0.0858	0.8578
N Gyeongsang	0.2301	0.0089	0.0872	1.1129	0.2537	0.0098	0.0961	1.2274
S Gyeongsang	0.2225	0.0081	0.0855	1.0509	0.2289	0.0083	0.0880	1.0808
Ulsan	0.1826	0.0043	0.0713	0.8149	0.1768	0.0042	0.0690	0.7890
S Jeolla	0.2293	0.0042	0.0769	1.1567	0.2590	0.0047	0.0868	1.3066
S Chungcheong	0.1800	0.0066	0.0687	0.8771	0.1967	0.0072	0.0750	0.9586
Incheon	0.1674	0.0065	0.0742	0.6868	0.1533	0.0059	0.0680	0.6289
Busan	0.1985	0.0063	0.0856	0.7915	0.2094	0.0066	0.0903	0.8349
N Chungcheong	0.1709	0.0067	0.0659	0.8113	0.1683	0.0066	0.0649	0.7995
Daegu	0.1782	0.0056	0.0695	0.7944	0.1970	0.0061	0.0768	0.8784
N Jeolla	0.1504	0.0045	0.0557	0.7488	0.1768	0.0053	0.0655	0.8804
Gangwon	0.1918	0.0052	0.0707	0.9931	0.1774	0.0048	0.0653	0.9182
Gwangju	0.1208	0.0025	0.0412	0.5875	0.1501	0.0031	0.0512	0.7301
Daejeon	0.0939	0.0026	0.0349	0.4469	0.1298	0.0036	0.0482	0.6174
Jeju	0.1253	0.0007	0.0381	0.6687	0.1798	0.0009	0.0546	0.9595
Mean	0.1833	0.0055	0.0702	0.8509	0.1948	0.0057	0.0741	0.9091
Weighted mean	0.2121	0.0066	0.0834	0.9507	0.2132	0.0066	0.0834	0.9603
V	0.247	0.435	0.284	0.242	0.193	0.401	0.214	0.211

Q3: third quartile (Q1 = 0 always), SD: standard deviation

Table 6 Statistics of the absolute differences (×100) between the observed and estimated input coefficients from the FLQ+ and GRAS+ methods

	$\left\|a_{ij}^r - \hat{a}_{ij}^r\right\|$ **FLQ+**				$\left\|a_{ij}^r - \hat{a}_{ij}^r\right\|$ **GRAS+**			
	Mean	**Max**	**Median**	**SD**	**Mean**	**Max**	**Median**	**SD**
Gyeonggi	0.1207	40.9	0.0033	0.8375	0.1379	51.3	0.0029	0.9830
Seoul	0.1478	27.9	0.0030	0.7102	0.1748	28.2	0.0020	0.8378
N Gyeongsang	0.1584	40.5	0.0043	0.8583	0.1876	68.7	0.0034	1.3135
S Gyeongsang	0.1426	34.7	0.0038	0.7409	0.1610	51.9	0.0029	1.0010
Ulsan	0.1620	48.3	0.0028	1.0850	0.1950	61.7	0.0017	1.3991
South Jeolla	0.1702	34.7	0.0030	0.9026	0.2044	67.6	0.0019	1.5013
S Chungcheong	0.1472	47.0	0.0033	0.9046	0.1556	60.2	0.0024	1.1178
Incheon	0.1323	25.5	0.0026	0.7887	0.1386	27.4	0.0022	0.8462
Busan	0.1484	18.7	0.0033	0.6654	0.1496	23.1	0.0028	0.7192
N Chungcheong	0.1476	49.6	0.0032	1.0025	0.1529	58.5	0.0028	1.1151
Daegu	0.1393	42.1	0.0032	0.7675	0.1398	49.3	0.0027	0.8530
North Jeolla	0.1373	52.0	0.0027	0.9149	0.1377	62.1	0.0023	1.0641
Gangwon	0.2027	124.6	0.0038	1.9126	0.2044	125.0	0.0035	1.9534
Gwangju	0.1357	48.6	0.0023	0.8746	0.1340	52.3	0.0021	0.9069
Daejeon	0.1382	40.8	0.0022	0.8513	0.1390	39.8	0.0023	0.8444
Jeju	0.1599	51.1	0.0010	0.9909	0.1522	51.1	0.0008	1.0117
Mean	0.1494	45.4	0.0030	0.9255	0.1603	54.9	0.0024	1.0917
Weighted mean	0.1448	39.2	0.0032	0.8542	0.1626	50.1	0.0025	1.0559
V	0.127	0.51	0.267	0.309	0.156	0.43	0.279	0.290

6.1 Analysis of the important input coefficients

Here we analyse the ability of the FLQ+ and GRAS+ methods to reproduce what Hewings and Romanos (1981) call 'inverse important coefficients', whereby an input coefficient a_{ij} is said to be *inverse important* if an error of α% in that coefficient produces a corresponding error of β% in one or more entries of the Leontief inverse.

In this analysis, α and β were set equal to 30% and 20%, respectively, as suggested by Hewings and Romanos. This means that an input coefficient is held to be inverse important if a perturbation of 30% generates a change of at least 20% in one or more entries in the Leontief inverse. More formally, we assume that only one coefficient, a_{ij}, of the matrix $\mathbf{A}$ is perturbed. Let a_{ij}^p denote this perturbed value of $a_{ij,}$ i.e. $a_{ij}^p = a_{ij}(1+\alpha/100)$. This shock generates a perturbed matrix $\mathbf{L^p} = [L_{ks}^p]$ of the Leontief matrix $\mathbf{L} = (\mathbf{I}-\mathbf{A})^{-1} = [L_{ks}]$, $k, s = 1, 2, \ldots, 78$, whose elements are determined as follows:

$$L_{ks}^p = L_{ks} + [L_{ki}L_{sj}a_{ij}(/100)]/[1 - L_{ji}a_{ij}(/100)], \tag{52}$$

where the indices i and j relate to the perturbed coefficient in $\mathbf{A}$, while k and s pertain to the Leontief inverse $\mathbf{L}$ and the corresponding perturbed matrix $\mathbf{L^p}$.

The relative error of each coefficient of the Leontief matrix reads as

$$\frac{L_{ks}^P - L_{ks}}{L_{ks}} = \frac{L_{ki}L_{sj}a_{ij}(/100)}{L_{ks}[1L_{ji}a_{ij}(/100)]}. \tag{53}$$

We then say that the coefficient a_{ij} is inverse important if there exists at least one pair (k, s) such that the following inequality holds:

$$\left|\frac{L_{ki}L_{sj}a_{ij}(/100)}{L_{ks}[1 - L_{ji}a_{ij}(/100)]}\right| \geq \beta/100. \tag{54}$$

Table 7 reveals a sharper distinction between the two methods, with the superiority of FLQ+ over GRAS+ now much more apparent, especially when the results are weighted by regional size. The MAD is now raised by 11.5% on average by using GRAS+ rather than FLQ+ and by 18.2% when regional weights are applied. It is clear that GRAS+ yields relatively poor estimates of the most important input coefficients. It is also apparent that the absolute differences increase when the focus is placed on the most important coefficients. This outcome is to be expected, given the increased size of the coefficients being estimated. Of the $78^2 = 6084$ coefficients, those deemed to be important ranged from 1651 (27.1%) for Jeju to 2147 (35.3%) for Daejeon, with a median of 1869 (30.7%).

6.2 Comparison of alternative methods

The analysis thus far has demonstrated the superiority of the proposed FLQ+ approach over the MCE and GRAS+ methods. However, what has not yet been explored is whether the FLQ can be expected to yield more accurate estimates of input coefficients than more straightforward methods such as the SLQ. For the purposes of this comparison, it is assumed that the analyst does not wish to employ matrix-balancing methods. It is further assumed that the choice is between the SLQ and the FLQ with a fixed value of $\delta = 0.3$ for all regions under consideration.

Table 7 Statistics of the absolute differences (×100) between the observed and estimated 'inverse important' input coefficients from the FLQ+ and GRAS+ methods

	$\left\lvert a_{ij}^r - \hat{a}_{ij}^r \right\rvert$ **FLQ+**				$\left\lvert a_{ij}^r - \hat{a}_{ij}^r \right\rvert$ **GRAS+**			
	Mean	**Max**	**Median**	**SD**	**Mean**	**Max**	**Median**	**SD**
Gyeonggi	0.3008	40.9	0.0743	1.2803	0.3554	51.3	0.0825	1.5706
Seoul	0.3374	18.4	0.0724	0.8849	0.4479	28.2	0.0654	1.2231
N Gyeongsang	0.4178	40.5	0.0908	1.3990	0.5306	68.7	0.1110	2.2601
S Gyeongsang	0.3771	34.7	0.0849	1.1992	0.4422	51.9	0.0851	1.6980
Ulsan	0.4026	48.3	0.0713	1.7161	0.5020	61.7	0.0801	2.2664
South Jeolla	0.4398	34.7	0.0923	1.4292	0.5537	67.6	0.1006	2.5508
S Chungcheong	0.3500	47.0	0.0805	1.4193	0.3791	60.2	0.0734	1.8007
Incheon	0.3236	25.5	0.0577	1.2792	0.3482	27.4	0.0618	1.3841
Busan	0.3747	18.7	0.0880	1.0106	0.3893	23.1	0.0926	1.1152
N Chungcheong	0.3476	49.6	0.0634	1.5083	0.3670	58.5	0.0625	1.7175
Daegu	0.3559	42.1	0.0710	1.2597	0.3655	49.3	0.0725	1.4249
North Jeolla	0.3293	52.0	0.0693	1.4640	0.3361	62.1	0.0649	1.7321
Gangwon	0.5654	124.6	0.1004	3.4241	0.5765	125.0	0.0993	3.4986
Gwangju	0.3344	48.6	0.0542	1.4388	0.3349	52.3	0.0529	1.5045
Daejeon	0.3101	40.8	0.0552	1.2330	0.3108	39.8	0.0558	1.2182
Jeju	0.3927	51.1	0.0747	1.6197	0.4077	51.1	0.0689	1.7564
Mean	0.3725	44.8	0.0750	1.4728	0.4154	54.9	0.0768	1.7951
Weighted mean	0.3583	37.5	0.0765	1.3111	0.4235	50.1	0.0796	1.7087
V	0.173	0.53	0.183	0.381	0.203	0.43	0.223	0.337

As noted in Sect. 2, the SLQ has the theoretical drawback that it precludes cross-hauling and hence tends to understate a region's imports from other regions. It is unsurprising, therefore, to observe in Table 8 that the SLQ generates less accurate results than the FLQ for all regions apart from Seoul and (marginally) South Jeolla. On average, the MAD is raised by 25.8% by using the SLQ rather than the FLQ and by 20.7% if regional weights are applied. In addition, the MAD for the SLQ exhibits greater interregional variation. The superior performance of the SLQ in Seoul is probably due to the relatively high variation in the size of the input coefficients in this region. In such situations, the SLQ is apt to produce more accurate estimates (Lamonica and Chelli 2017).

Given the uncertainty regarding the appropriate value of δ, it is reassuring that Table 9 shows that the outcomes from the FLQ are little affected by variation in the value of this parameter in the range 0.2 to 0.4. The SLQ still outperforms the FLQ in Seoul and South Jeolla but yields inferior results in the other fourteen regions.

7 Analysis of multipliers

One of the most useful features of RIOTs is the fact that they yield estimates of sectoral output multipliers, so Tables 10 and 11 examine the relative performance in this regard of eight alternative procedures. The following formulae are used in this evaluation:

$$\text{MAPE} = \frac{100}{78} \sum_{j=1}^{78} \frac{\left| \widehat{L_j} - L_j \right|}{L_j}; \tag{55}$$

Table 8 Statistics of the absolute differences (×100) between the observed and estimated input coefficients from the FLQ (with a fixed $\delta = 0.3$) and SLQ

	$\left\lvert a_{ij}^r - \hat{a}_{ij}^r \right\rvert$ **FLQ (with** $\delta = 0.3$**)**				$\left\lvert a_{ij}^r - \hat{a}_{ij}^r \right\rvert$ **SLQ**			
	Mean	**Max**	**Median**	**SD**	**Mean**	**Max**	**Median**	**SD**
Gyeonggi	0.1306	28.0	0.0044	0.7456	0.1883	45.6	0.0067	1.0584
Seoul	0.1474	27.9	0.0032	0.7219	0.1093	27.8	0.0017	0.6091
N Gyeongsang	0.1573	44.1	0.0042	0.8980	0.1744	20.7	0.0052	0.7840
S Gyeongsang	0.1427	42.3	0.0038	0.7345	0.1831	20.4	0.0055	0.8640
Ulsan	0.1653	41.8	0.0034	0.9905	0.1935	64.5	0.0023	1.2788
South Jeolla	0.1688	42.3	0.0028	1.0038	0.1593	29.8	0.0023	0.7967
S Chungcheong	0.1574	39.2	0.0043	0.8342	0.2411	24.8	0.0071	1.1199
Incheon	0.1435	22.9	0.0042	0.7423	0.2358	27.5	0.0070	1.0377
Busan	0.1483	18.8	0.0033	0.6658	0.1986	18.3	0.0055	0.8757
N Chungcheong	0.1476	49.1	0.0032	0.9973	0.2058	27.1	0.0064	1.0600
Daegu	0.1388	46.7	0.0029	0.8194	0.2055	63.3	0.0048	1.1814
North Jeolla	0.1367	53.3	0.0027	0.9311	0.1877	22.9	0.0052	0.9586
Gangwon	0.2050	125.1	0.0034	1.9620	0.1723	120.9	0.0040	1.7209
Gwangju	0.1360	48.2	0.0023	0.8713	0.1921	54.0	0.0027	1.1145
Daejeon	0.1419	36.5	0.0026	0.8237	0.2268	55.4	0.0043	1.3287
Jeju	0.1527	51.1	0.0007	1.0352	0.1709	55.6	0.0007	1.2610
Mean	0.1513	44.8	0.0032	0.9235	0.1903	41.4	0.0045	1.0656
Weighted mean	0.1481	37.0	0.0036	0.8363	0.1787	36.5	0.0046	0.9547
V	0.118	0.53	0.290	0.324	0.167	0.63	0.448	0.248

Table 9 Statistics illustrating how the absolute differences (×100) between the observed and estimated input coefficients vary with changes in the assumed value of δ

	$\left\lvert a_{ij}^r - \hat{a}_{ij}^r \right\rvert$ **FLQ (with** $\delta = 0.2$**)**				$\left\lvert a_{ij}^r - \hat{a}_{ij}^r \right\rvert$ **FLQ (with** $\delta = 0.4$**)**			
	Mean	**Max**	**Median**	**SD**	**Mean**	**Max**	**Median**	**SD**
Gyeonggi	0.1502	27.9	0.0054	0.7620	0.1215	33.0	0.0039	0.7681
Seoul	0.1512	27.8	0.0035	0.7590	0.1484	27.9	0.0030	0.7083
N Gyeongsang	0.1644	34.0	0.0047	0.8089	0.1596	52.4	0.0038	1.0142
S Gyeongsang	0.1504	23.6	0.0047	0.6777	0.1461	42.4	0.0033	0.8403
Ulsan	0.1806	35.1	0.0044	0.9522	0.1619	47.1	0.0028	1.0645
South Jeolla	0.1723	31.6	0.0032	0.8710	0.1728	50.8	0.0024	1.1507
S Chungcheong	0.1818	28.5	0.0058	0.8295	0.1468	47.7	0.0032	0.9132
Incheon	0.1705	21.0	0.0053	0.8168	0.1327	24.4	0.0033	0.7495
Busan	0.1583	18.2	0.0041	0.6743	0.1493	22.9	0.0029	0.7166
N Chungcheong	0.1554	37.3	0.0043	0.9125	0.1520	57.8	0.0027	1.1050
Daegu	0.1452	34.9	0.0041	0.7128	0.1450	55.4	0.0023	0.9432
North Jeolla	0.1493	41.4	0.0036	0.8213	0.1376	62.0	0.0023	1.0630
Gangwon	0.2030	124.2	0.0041	1.8780	0.2130	125.8	0.0029	2.0477
Gwangju	0.1505	36.2	0.0032	0.8309	0.1343	56.9	0.0017	0.9564
Daejeon	0.1600	25.4	0.0036	0.8160	0.1349	45.1	0.0019	0.8906
Jeju	0.1564	51.1	0.0010	0.9861	0.1586	51.1	0.0006	1.1137
Mean	0.1625	37.4	0.0041	0.8818	0.1509	50.2	0.0027	1.0028
Weighted mean	0.6007	31.3	0.0044	0.8158	0.1460	41.6	0.0031	0.8944
V	0.096	0.65	0.278	0.317	0.139	0.47	0.307	0.315

Table 10 Statistics of the mean absolute percentage differences (MAPEs) between the estimated and observed type I output multipliers from alternative approaches

	MCE	FLQ+	FLQ0.3	GRAS	GRAS+	AFLQ	AFLQ+	SLQ
Gyeonggi	11.408	1.913	3.355	16.875	5.682	0.605	1.748	4.940
Seoul	15.252	1.947	0.707	2023.701	8.554	0.108	1.145	2.320
N Gyeongsang	14.979	2.120	3.155	8.967	9.253	0.415	2.235	7.788
S Gyeongsang	1.697	4.489	4.293	9.531	8.050	3.039	2.799	5.232
Ulsan	23.051	3.868	0.186	315.443	10.203	2.905	3.440	3.860
South Jeolla	15.573	1.184	3.263	2.784	9.754	8.091	11.756	9.666
S Chungcheong	0.327	2.077	0.463	3.149	6.177	7.046	2.575	17.014
Incheon	22.071	2.638	2.415	284.593	4.704	1.882	2.557	13.515
Busan	12.517	2.825	2.825	261.020	5.314	3.191	2.729	2.436
N Chungcheong	4.436	4.190	4.190	12.482	6.827	0.247	0.247	12.200
Daegu	13.808	3.204	4.513	199.788	5.636	3.979	2.922	9.937
North Jeolla	6.728	2.046	2.333	4.315	4.871	5.298	8.479	14.849
Gangwon	5.919	6.580	7.671	377.345	7.859	1.692	2.505	2.589
Gwangju	11.166	4.110	3.730	503.281	4.803	4.924	4.097	3.717
Daejeon	18.257	2.850	1.493	575.148	2.762	4.563	2.612	13.339
Jeju	7.077	1.933	5.860	2.827	5.187	7.743	17.545	13.453
Mean	11.517	2.998	3.153	287.578	6.602	3.483	4.337	8.553
Weighted mean	12.456	2.589	2.548	461.489	7.201	2.392	3.041	6.822
V	6.540	1.321	1.911	486.558	2.056	2.559	4.367	4.881

FLQ0.3 denotes the FLQ with a fixed $\delta = 0.3$

$$\mathrm{MPE} = \frac{100}{78} \sum_{j=1}^{78} \frac{\widehat{L_j} - L_j}{L_j}, \tag{56}$$

where $\widehat{L_j}$ and L_j denote the estimated and observed column totals of the Leontief inverse matrices.

Table 10 displays the outcomes in terms of MAPEs when estimating type I output multipliers for the sixteen regions. The first two columns illustrate the benefits of pursuing the FLQ+ approach rather than the MCE method alone. In particular, there is a sharp fall in both unweighted and weighted means, along with a large decrease in dispersion. It is evident that the MCE method by itself does not yield satisfactory estimates of multipliers.

Although the FLQ+ and FLQ with $\delta = 0.3$ have similar means, there are marked differences in the MAPEs for several regions. Busan and North Chungcheong are exceptions where the two methods yield identical results; this is because $\delta = 0.3$ happens to be the estimated value from the FLQ+.

By itself, GRAS generates highly inaccurate results, yet combining it with the FLQ in the form of GRAS+ is clearly helpful in producing a more acceptable set of results. This better outcome is due to the fact that, when estimating δ in the GRAS+ procedure, we consider only those pairs $(\hat{a}_{ij}^r, a_{ij}^n)$ satisfying the 'sign-preserving assumption', whereas GRAS accounts for all possible pairs. Nonetheless, GRAS+ still yields substantially worse results than FLQ+.

It is interesting that the AFLQ has the lowest weighted mean of all eight approaches, whereas FLQ+ is the best in terms of the unweighted mean. This outcome is

Table 11 Statistics of the mean percentage differences (MPEs) between the estimated and observed type I output multipliers from alternative approaches

	MCE	FLQ+	FLQ0.3	GRAS	GRAS+	AFLQ	AFLQ+	SLQ
Gyeonggi	11.408	−1.913	3.355	−16.875	−5.682	−0.605	−1.748	4.940
Seoul	15.252	−1.947	−0.707	−2023.701	−8.554	−0.108	−1.145	−2.320
N Gyeongsang	14.979	−2.120	−3.155	−8.967	−9.253	0.415	2.235	7.788
S Gyeongsang	1.697	−4.489	−4.293	−9.531	−8.050	−3.039	−2.799	5.232
Ulsan	23.051	−3.868	−0.186	−315.443	−10.203	−2.905	−3.440	3.860
South Jeolla	15.573	−1.184	−3.263	2.784	−9.754	8.091	11.756	9.666
S Chungcheong	0.327	−2.077	0.463	3.149	−6.177	7.046	2.575	17.014
Incheon	22.071	−2.638	2.415	−284.593	−4.704	−1.882	−2.557	13.515
Busan	12.517	−2.825	−2.825	−261.020	−5.314	−3.191	−2.729	2.436
N Chungcheong	−4.436	−4.190	−4.190	−12.482	−6.827	0.247	0.247	12.200
Daegu	13.808	−3.204	−4.513	−199.788	−5.636	−3.979	−2.922	9.937
North Jeolla	−6.728	−2.046	−2.333	−4.315	−4.871	5.298	8.479	14.849
Gangwon	5.919	−6.580	−7.671	377.345	−7.859	−1.692	−2.505	2.589
Gwangju	11.166	−4.110	−3.730	−503.281	−4.803	−4.924	−4.097	3.717
Daejeon	18.257	−2.850	−1.493	−575.148	−2.762	−4.563	−2.612	13.339
Jeju	−7.077	−1.933	−5.860	2.827	−5.187	7.743	17.545	13.453
Mean	9.237	−2.998	−2.374	−239.315	−6.602	0.122	1.018	8.263
Weighted mean	11.736	−2.589	−0.876	−444.087	−7.201	0.026	−0.095	5.978
V	9.492	−1.321	−2.821	−512.024	−2.056	4.321	6.070	5.357

FLQ0.3 denotes the FLQ with a fixed $\delta = 0.3$. Negative values indicate underestimates

attributable to the superior performance of the AFLQ in several larger regions, most noticeably in Seoul. Nonetheless, it is odd that AFLQ+ often generates worse results than the AFLQ, especially in South Jeolla and Jeju. As regards the remaining approach, the SLQ, Table 10 reveals that its performance is clearly inferior to that of the FLQ and AFLQ.

While the MAPE index is a good way of assessing the relative accuracy of alternative methods, it tells us nothing about the direction or extent of bias. For that reason, Table 11 displays outcomes for actual rather than absolute percentage deviations.

A key finding from Table 11 is that the FLQ+ and GRAS+ procedures invariably generate underestimates of multipliers. Taking regional size into account, the FLQ+ understates these multipliers by 2.6% on average, whereas GRAS+ does so by 7.2%. Apart from the sign, the results from the FLQ+ and GRAS+ methods are identical in the two tables. By contrast, the SLQ produces a positive bias of 6.8% on average. Seoul is the only region where this bias is negative.

Another important outcome from Table 11 is that the AFLQ shows minimal bias on average. However, the estimated multipliers from the AFLQ display greater dispersion than those from FLQ+, which would tend to boost the MAPEs in Table 10. Even so, the AFLQ still emerges with a weighted mean MAPE that is marginally lower than that for FLQ+. What is more, of the eight procedures examined in Table 10, the AFLQ produces the best results for the problematic Seoul region.

8 Conclusion

This paper has proposed a new approach to the regionalization of national input–output tables where very limited regional data exist and analysts are considering employing methods based on location quotients. We focus on the FLQ, which often yields the most accurate results of such methods. The FLQ formula involves an unknown parameter, δ, which plays a crucial role in the regionalization process. However, the difficulty of selecting an appropriate value of δ has been an obstacle to the successful application of the FLQ approach. Hitherto, analysts have had to choose a value of δ on the basis of a priori considerations and the results of previous case studies, which often conflict. Our aim has been to develop an enhanced and more objective way of implementing the FLQ approach.

Our proposed procedure, the FLQ+ method, is a hybrid approach that uses the results from a modified cross-entropy (MCE) method and a simple regression model to estimate δ. This value, which is specific to a given country and region, is used to estimate regional input coefficients and hence multipliers via the FLQ method. Our empirical analysis, based on a rare survey-based South Korean interregional input–output table for 2005 with 16 regions and 78 sectors, gave credence to the proposed method, in that the FLQ+ behaved much like the FLQ with an optimal δ.

The FLQ+ approach generated substantially more accurate estimates of both input coefficients and sectoral output multipliers than those from the MCE approach alone. Moreover, the MCE method clearly outperformed GRAS. In further testing of the FLQ+ approach, we considered the simple LQ (SLQ), the augmented FLQ (AFLQ) and the FLQ with an assumed value of δ. However, the SLQ gave inferior outcomes to the FLQ+, so we would not recommend using this method.

The AFLQ and FLQ+ gave almost identical results when estimating input coefficients. By contrast, in terms of multipliers, the AFLQ gave some good outcomes, characterized by minimal bias, whereas the FLQ+ tended to understate the multipliers. Indeed, the AFLQ generated the most accurate estimates of multipliers for the problematic Seoul region.

Using the FLQ with an assumed value of δ has the merit of ease of application, yet it runs the risk of choosing an inappropriate value. To explore this issue, we used values of δ in the interval 0.3 ± 0.1 to estimate input coefficients. This range is suggested by Jahn et al. (2020). We found that the mean absolute differences were not much affected by variation in δ. However, an analysis of multipliers revealed marked differences in the mean absolute percentage errors for several regions, despite the fact that the FLQ+ and FLQ with $\delta = 0.3$ had similar means. Therefore, we would recommend using the FLQ+ approach, which takes region-specific characteristics into account, along with any country-specific differences.

In concluding, we should note some avenues for further research on this topic. The first would be to re-examine the effectiveness of the AFLQ, in the light of its good performance in estimating multipliers. It would also be interesting to see how well the 2DLQ refinement of the AFLQ proposed by Pereira-López et al. (2020) performs when tested using a national data set such as that for South Korea. Secondly, it would be useful to investigate why the FLQ+ tended to underestimate output multipliers. This outcome may well because the optimal values of δ were derived by minimizing the mean absolute

difference. An alternative might be to employ Theil's inequality coefficient, which captures both bias and variance (Stevens et al. 1989). Thirdly, it would be worthwhile to consider how superior data might be used in practical applications of the FLQ+ approach. Finally, it would be interesting to examine the possible benefits of incorporating the theoretical restrictions regarding the value of δ into the constraint set of the MCE model.

Appendix

This appendix shows how the proposed hybrid approach can be adapted to encompass the AFLQ. All that is required is to modify Eqs. (48) and (49) as follows. First define $\alpha_{ij} = \text{CILQ}_{ij}$ if $i \neq j$ and $\alpha_{ij} = \text{SLQ}_i$ if $i = j$, so that Eq. (48) changes to

$$\hat{a}_{ij}^r = a_{ij}^n w_{ij} \beta^{\delta}, \tag{57}$$

where w_{ij} reads as

$$w_{ij} = \begin{cases} \alpha_{ij}\left[\log_2\left(1 + SLQ_j\right)\right] & \text{for } \text{SLQ}_j > 1 \\ \alpha_{ij} & \text{for } \text{SLQ}_j \leq 1 \end{cases}.$$

Thus, for any pair $(\hat{a}_{ij}^r, a_{ij}^n)$ that satisfies the sign-preserving assumption, we set:

$$\log\left(\frac{\hat{a}_{ij}^r}{w_{ij} a_{ij}^n}\right) = \delta \log(\beta) + \varepsilon ij, \tag{58}$$

Table 12 Comparison of results for the FLQ and AFLQ in terms of input coefficients. Source: authors' calculations

Region	FLQ		FLQ+		AFLQ		AFLQ+	
	MAD	Optimal δ	MAD	Estimated δ	MAD	Optimal δ	MAD	Estimated δ
Gyeonggi	0.12	0.48	0.12	0.54	0.12	0.60	0.12	0.56
Seoul	0.15	0.32	0.15	0.37	0.14	0.58	0.14	0.69
N Gyeongsang	0.16	0.31	0.16	0.26	0.16	0.43	0.16	0.35
S Gyeongsang	0.14	0.32	0.14	0.31	0.16	0.46	0.16	0.39
Ulsan	0.16	0.41	0.16	0.43	0.16	0.49	0.16	0.53
South Jeolla	0.17	0.28	0.17	0.23	0.16	0.45	0.17	0.28
S Chungcheong	0.15	0.43	0.15	0.39	0.14	0.48	0.14	0.42
Incheon	0.13	0.45	0.13	0.48	0.14	0.51	0.14	0.39
Busan	0.15	0.32	0.15	0.30	0.15	0.39	0.15	0.35
N Chungcheong	0.15	0.30	0.15	0.30	0.15	0.41	0.15	0.36
Daegu	0.14	0.29	0.14	0.26	0.14	0.31	0.14	0.29
North Jeolla	0.14	0.34	0.14	0.29	0.13	0.41	0.14	0.31
Gangwon	0.20	0.22	0.20	0.24	0.20	0.32	0.20	0.28
Gwangju	0.13	0.37	0.13	0.32	0.13	0.42	0.14	0.36
Daejeon	0.13	0.44	0.14	0.35	0.13	0.49	0.14	0.40
Jeju	0.15	0.26	0.16	0.18	0.15	0.30	0.16	0.19
Mean	0.15	0.35	0.15	0.31	0.15	0.44	0.15	0.38
V	0.13	0.22	0.12	0.30	0.12	0.20	0.12	0.32

from which a value of δ can be derived as done for Eq. (49). However, it should be noted that the optimal δ for the AFLQ can normally be expected to exceed the corresponding value for the FLQ (e.g. Bonfiglio and Chelli 2008, table 1).

Table 12 reveals that the FLQ and AFLQ give almost identical values of the MAD. Therefore, Ockham's principle suggests that the FLQ should be preferred on the basis of its greater simplicity. However, such a conclusion would not be supported by the outcomes in terms of type I output multipliers presented in Tables 10 and 11, which indicate that the AFLQ can produce superior results to the FLQ. It should be noted, finally, that the values of δ shown in Table 12 for the AFLQ are almost always higher than those for the FLQ but that expected outcome is irrelevant when assessing their relative performance.

Abbreviations

a^n_{ij}: Observed national input coefficient; $\hat{a}^r_{ij}$: Estimated regional input coefficient; NIOT: National input–output table; RIOT: Regional input–output table; KIRIOT: South Korean interregional input–output table; LQ: Location quotient; SLQ: Simple LQ; CILQ: Cross-industry LQ; RLQ: Round's LQ; FLQ: Flegg's LQ; FLQ+: A refined version of the FLQ approach proposed in this paper; AFLQ: Augmented FLQ; SFLQ: Industry-specific FLQ; DSLQ: Scaling formula proposed by Fujimoto (2019); 2DLQ: Bidimensional LQ; CHARM: Cross-hauling adjusted regionalization method; RCHARM: Revised CHARM; MCHARM: Modified CHARM; RPC: Regional purchase coefficient; SDR: Supply/demand ratio; CE: Cross-entropy; MCE: Modified cross-entropy; RAS: A biproportional matrix-balancing method; GRAS: Generalized RAS; MAD: Mean absolute difference.

Acknowledgements

We thank the Editor and the anonymous referees for their scrutiny of our paper. Chris Webber read several earlier versions of the paper and made numerous helpful and insightful suggestions for refining it.

Authors' contributions

This paper reports the results of a collaborative study by the authors, all of whom read and approved the final manuscript.

Funding

None to report.

Availability of data and materials

The basic data were downloaded from the Bank of Korea's website, bok.or.kr.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Accounting, Economics and Finance, University of the West of England, Coldharbour Lane, Bristol BS16 1QY, UK. [2]Faculty of Economics, Università Politecnica delle Marche, Ancona, Italy. [3]School of Business and Economics, University of Jyväskylä, PO Box 35, 40014 Jyvaskyla, Finland.

Received: 17 December 2020 Revised: 2 July 2021 Accepted: 3 July 2021
Published online: 16 July 2021

References

Bacharach M (1970) Biproportional matrices and input output change. Cambridge University Press, England

Boero R, Edwards BK, Rivera MK (2018) Regional input–output tables and trade flows: an integrated and interregional non-survey approach. Reg Stud 52:225–238. https://doi.org/10.1080/00343404.2017.1286009

Bonfiglio A, Chelli F (2008) Assessing the behaviour of non-survey methods for constructing regional input–output tables through a Monte Carlo simulation. Econ Syst Res 20:243–258. https://doi.org/10.1080/09535310802344315

Dávila-Flores A (2015) Modelos Interregionales de Insumo Producto de la Economía Mexicana. Universidad Autónoma de Coahuila, Saltillo

Davis HC, Lofting EM, Sathaye JA (1977) A comparison of alternative methods of updating input–output coefficients. Technol Forecast Soc Change 10:79–87. https://doi.org/10.1016/0040-1625(80)90004-9
Flegg AT, Tohmo T (2013) Regional input–output tables and the FLQ formula: a case study of Finland. Reg Stud 47:703–721. https://doi.org/10.1080/00343404.2011.592138
Flegg AT, Tohmo T (2016) Estimating regional input coefficients and multipliers: the use of FLQ is not a gamble. Reg Stud 50:310–325. https://doi.org/10.1080/00343404.2014.901499
Flegg AT, Tohmo T (2018) The Regionalization of national input–output tables: a review of the performance of two key non-survey methods. In: Mukhopadhyay K (ed) Applications of the input–output framework. Springer, Singapore, pp 347–386
Flegg AT, Tohmo T (2019) The regionalization of national input–output tables: a study of South Korean regions. Pap Reg Sci 98:601–620. https://doi.org/10.1111/pirs.12364
Flegg AT, Webber CD (1997) On the appropriate use of location quotients in generating regional input–output tables: reply. Reg Stud 31:795–805. https://doi.org/10.1080/713693401
Flegg AT, Webber CD (2000) Regional size, regional specialization and the FLQ formula. Reg Stud 34:563–569. https://doi.org/10.1080/00343400050085675
Flegg AT, Webber CD, Elliott MV (1995) On the appropriate use of location quotients in generating regional input–output tables. Reg Stud 29:547–561. https://doi.org/10.1080/00343409512331349173
Flegg AT, Mastronardi LJ, Romero CA (2016) Evaluating the FLQ and AFLQ formulae for estimating regional input coefficients: empirical evidence for the province of Córdoba, Argentina. Econ Syst Res 28:21–37. https://doi.org/10.1080/09535314.2015.1103703
Fujimoto T (2019) Appropriate assumption on cross-hauling national input–output table regionalization. Spat Econ Anal 14:106–128. https://doi.org/10.1080/17421772.2018.1506151
Golan A, Vogel SJ (2000) Estimation of non-stationary social accounting matrix coefficients with supply-side information. Econ Syst Res 12:447–471. https://doi.org/10.1080/09535310020003775
Golan A, Judge G, Robinson S (1994) Recovering information from incomplete or partial multisectorial economic data. Rev Econ Stat 76:541–549. https://doi.org/10.2307/2109978
Hermannsson K (2016) Beyond intermediates: the role of consumption and commuting in the construction of local input–output tables. Spat Econ Anal 11:315–339. https://doi.org/10.1080/17421772.2016.1177194
Hewings GJD, Romanos MC (1981) Simulating less-developed regional economies under conditions of limited information. Geogr Anal 13:373–390
Hosoe N (2014) Estimation errors in input–output tables and prediction errors in computable general equilibrium analysis. Econ Model 42:277–286. https://doi.org/10.1016/j.econmod.2014.07.012
Huang W, Kobayashi S, Tanji H (2008) Updating an input–output matrix with sign-preservation: some improved objective functions and their solutions. Econ Syst Res 20:111–123. https://doi.org/10.1080/09535310801892082
Jahn M (2017) Extending the FLQ formula: a location quotient-based interregional input–output framework. Reg Stud 51:1518–1529. https://doi.org/10.1080/00343404.2016.1198471
Jahn M, Flegg AT, Tohmo T (2020) Testing and implementing a new approach to estimating interregional output multipliers using input–output data for South Korean regions. Spat Econ Anal 15:165–185. https://doi.org/10.1080/17421772.2020.1720918
Junius T, Oosterhaven J (2003) The solution of updating or regionalizing a matrix with both positive and negative entries. Econ Syst Res 15:87–96. https://doi.org/10.1080/0953531032000056954
Kapur J, Kesavan H (1992) Entropy optimization principles with applications. Academic Press, New York
Kowalewksi J (2015) Regionalization of national input–output tables: empirical evidence on the use of the FLQ formula. Reg Stud 49:240–250. https://doi.org/10.1080/00343404.2013.766318
Kronenberg K, Fuchs M (2021) The socio-economic impact of regional tourism: an occupation-based modelling perspective from Sweden. J Sustain Tour. https://doi.org/10.1080/09669582.2021.1924757
Lahr ML, Ferreira JP, Többen JR (2020) Intraregional trade shares for goods-producing industries: RPC estimates using EU data. Pap Reg Sci 99:1583–1605. https://doi.org/10.1111/pirs.12541
Lamonica RG, Chelli FM (2017) The performance of non-survey techniques for constructing sub-territorial input–output tables. Pap Reg Sci 97:1169–1202. https://doi.org/10.1111/pirs.12297
Lamonica RG, Recchioni MC, Chelli FM, Salvati L (2020) The efficiency of the cross-entropy method when estimating the technical coefficients of input–output tables. Spat Econ Anal 15:62–91. https://doi.org/10.1080/17421772.2019.1615634
Lampiris G, Karelakis C, Loizou E (2020) Comparison of non-survey techniques for constructing regional input–output tables. Ann Oper Res 294:225–266. https://doi.org/10.1007/s10479-019-03337-5
Lemelin A (2009) A GRAS variant solving for minimum information loss. Econ Syst Res 21:399–408. https://doi.org/10.1080/09535311003589310
Léony Y, Peeters L, Quinqu M, Surry Y (1999) The use of maximum entropy to estimate input–output coefficients from regional farm accounting data. J Agric Econ 50:425–439. https://doi.org/10.1111/j.1477-9552.1999.tb00891.x
McCann P, Dewhurst JHL (1998) Regional size, industrial location and input–output expenditure coefficients. Reg Stud 32:435–444. https://doi.org/10.1080/00343409850116835
Morrissey K (2016) A location quotient approach to producing regional production multipliers for the Irish economy. Pap Reg Sci 95:491–506. https://doi.org/10.1111/pirs.1214.3
Okamoto N (2014) Does regional size matter in regionalization of national input–output table by the FLQ formula? A case study of China, Discussion Paper 222, Institute of Economic Research, Chuo University
Pavia JM, Cabrer B, Sala R (2009) Updating input–output matrices: assessing alternatives through simulation. J Stat Comput Simul 79:1467–1482. https://doi.org/10.1080/00949650802415154
Pereira-López X, Carrascal-Incera A, Fernández-Fernández M (2020) A bidimensional reformulation of location quotients for generating input–output tables. Spat Econ Anal 15:476–493. https://doi.org/10.1080/17421772.2020.1729996
Pereira-López X, Sanchez-Choez NG, Fernández-Fernández M (2021) Performance of bidimensional location quotients for constructing input–output tables. J Econ Struc 10:7. https://doi.org/10.1186/s40008-021-00237-5

Robinson S, Cattaneo A, El-Said M (2001) Updating and estimating a social accounting matrix using cross entropy methods. Econ Syst Res 13:47–64. https://doi.org/10.1080/09535310120026247
Round JI (1978) An inter-regional input–output approach to the evaluation of nonsurvey methods. J Reg Sci 18:179–194. https://doi.org/10.1111/j.1467-9787.1978.tb00540.x
Singh I, Singh L (2011) Regional input output table for the state of Punjab. https://mpra.ub.uni-muenchen.de/32344/. Accessed 17 May 2021
Stevens BH, Treyz GI, Ehrlich DJ, Bower JR (1983) A new technique for the construction of non-survey regional input–output models. Int Reg Sci Rev 8:271–286
Stevens BH, Treyz GI, Lahr ML (1989) On the comparative accuracy of RPC estimating techniques. In: Miller RE, Polenske KR, Rose AZ (eds) Frontiers of input–output analysis. Oxford University Press, Oxford, pp 245–257
Stone R (1961) Input output and national accounts. OEEC, Paris
Szabó N (2015) Methods for regionalizing input–output tables. Reg Stat 5:44–65. https://doi.org/10.15196/RS05103
Temurshoev U, Miller RE, Bouwmeester MC (2013) A note on the GRAS method. Econ Syst Res 25:361–367. https://doi.org/10.1080/09535314.2012.746645
Többen J, Kronenberg T (2015) Construction of multi-regional input–output tables using the CHARM method. Econ Syst Res 27:487–507. https://doi.org/10.1080/09535314.2015.1091765
Treyz GI, Stevens BH (1985) The TFS regional modelling methodology. Reg Stud 19:547–562. https://doi.org/10.1080/09595238500185531
Vazquez EF, Hewings GJD, Carvajal CR (2015) Adjustment of input–output tables from two initial matrices. Econ Syst Res 27:345–361. https://doi.org/10.1080/09535314.2015.1007839
Wiebe KS, Lenzen M (2016) To RAS or not to RAS? What is the difference in outcomes in multi-regional input–output models? Econ Syst Res 28:383–402. https://doi.org/10.1080/09535314.2016.1192528
Zhao X, Choi S-G (2015) On the regionalization of input–output tables with an industry-specific location quotient. Ann Reg Sci 54:901–926. https://doi.org/10.1007/s00168-015-0693-x

Publisher's Note