No More Gambling: Estimating Non-Survey Regional Input-Output Tables by Averaging Location Quotients

Esteban Fernández Vázquez,
REGIO*lab*, University of Oviedo, Spain evazquez@uniovi.es

Geoffrey J.D. Hewings,
REGIOlab, University of Oviedo, Spain hewwings@uniovi.es
University of Illinois, USA hewings@illinois.edu

Keywords: regional nonsurvey input-output estimation, parameter choice, averaging estimates

Abstract:
Datasets containing regional input–output (IO) tables are not frequently produced by surveys but estimated using regionalization techniques. When the required information is available, an adjusting procedure of national IO tables based on RAS algorithms and its variants are widely applied. But if the required information is not at hand, a regionalization procedure based on Location Quotients (LQ) is the most popular technique. Broadly speaking, this class of methods consist in multiplying the national matrix of IO coefficients by a regional LQ, with the solutions suggested by Flegg *et al*. (1995) and Flegg and Webber (1997, 2000) extensively adopted and applied. However, some authors (Lehtonen and Tykkyläinen, 2012; Kowaleski, 2015) have criticized these solutions, arguing that they critically depend on the specification of an unknown parameter value: there is a multiplicity of estimated regional IO tables depending on this particular choice. In order to respond to these critiques, regression-based techniques have been proposed (see Buendía *et al*. 2022, for a recent example) to find "optimal" values for the parameter of interest.
This paper proposes an easy-to-apply way of producing non-survey regional and interregional IO tables from national information and data in the form of LQs. Instead of relying on a specific parameter value that sets the values of the LQ applied, the resulting regional IO table is calculated as the simple average of several solutions produced by setting a range of plausible values for our LQs. The technique is illustrated and evaluated by conducting a numerical simulation similar to the exercise presented in Bonfiglio and Chelli (2009). Our results show how this simple approach largely alleviates the arbitrariness present in the choice of the parameter values for specifying the LQs, while reducing remarkably the risk of producing regional IO tables that deviates from the true ones.
A similar problem arises in a completely different field within economic modeling: in time-series analysis and particularly in the field of forecasting. In this branch of the literature, the use of forecast combinations has been extensively studied, and it has been empirically that forecast combinations produce a superior performance when compared with individual predictions. The initial work by Newbold and Granger (1974), Makridakis *et al*. (1982), Makridakis and Winkler (1983), have been continued by more recent studies, such as in Stock and Watson (2004), Smith and Wallis (2009) or Genre *et al*. (2013). All provide examples of the empirical success of forecasting strategies consisting of computing a simple mean of individual forecasts. Through this averaging, the variance of the resulting predictions is significantly reduced and there are remarkable gains in terms of accuracy. The methodology is illustrated for a single input-output table but can be readily extended to multiregional systems. To be updated/shortened. [I will re-work this once we agree on the rest of the paper]

1. Introduction

Economic analysis has not only sought to develop readily implementable models of national and regional economies but also to uncover empirical regularities that can be used to create and enhance future modeling efforts. Some of these regularities have been appropriated from other sciences (the gravity formulations, for example) while others have been estimated based on extensive empirical testing. A good example here would be the estimation of the wage curve, initially developed by Blanchflower and Oswald (1995); extensive empirical analysis (now numbering well over 100 studies) demonstrates a downward-sloping curve linking the level of a worker's pay to the unemployment rate in the worker's region. The empirical evidence suggests that the elasticity of wages with respect to unemployment is -0.1. In building more extensive models of regional economies, such as econometric-input-output or computable general equilibrium, the wage curve findings have been embedded into the formulations. However, when it comes to the interindustry components of the model, survey data are lacking and a variety of nonsurvey techniques has been proposed and tested, but to a lesser extent than the wage-curve dynamics.

Renewed interest in the dynamics of income inequality and the increased attention that has been focused on “left-behind” places have together generated demand for subnational analysis that can be effectively linked with macroeconomic models. Beenstock (2023), Kratena (2023) and Bhattacharjee *et al*. (2024) have recently explored alternative ways in which regional dimensions can be integrated into macroeconomic models. Rokicki *et al*. (2021) evaluted the impacts of using regional/multiregional models derived from survey as opposed to nonsurvey methods. These two issues are clearly related since without assurances about the quality of the regional modeling components, conducting macro-to-regional analyses would not engender confidence in the results.

Datasets containing regional input–output (IO) tables are not frequently produced by surveys but estimated using regionalization techniques. When the required information is available, an adjusting procedure of national IO tables based on RAS algorithms and its variants have been widely applied. However, if the required information is not at hand, a regionalization procedure based on Location Quotients (LQ) remains one of the most the most popular technique. Broadly speaking, this class of methods consist in multiplying the national matrix of IO coefficients by a regional LQ, with the solutions and modifications suggested by Flegg *et al*. (1995) and Flegg and Webber (1997, 2000) extensively adopted and applied. However, some authors (Lehtonen and Tykkyläinen, 2012; Kowaleski, 2015) have criticized these modifications, arguing that they critically depend on the specification of an unknown parameter value: there is a multiplicity of estimated regional IO tables depending on this particular choice. In order to respond to these critiques, regression-based techniques have been proposed (see Buendía *et al*. 2022, for a recent example) to find “optimal” values for the parameter of interest.

A similar problem arises in a completely different field within economic modeling: in time-series analysis and particularly in the field of forecasting. In this branch of the literature, the use of forecast combinations has been extensively studied, and it has been empirically tested that forecast combinations produce a superior performance when compared with individual predictions. The initial work by Newbold and Granger (1974), Makridakis *et al*. (1982), Makridakis and Winkler (1983), have been continued by more recent studies, such as in Stock and Watson (2004), Smith and Wallis (2009) or Genre *et al*. (2013). All provide examples of the empirical success of forecasting strategies consisting of computing a simple mean of individual forecasts. Through this averaging, the variance of the resulting predictions is significantly reduced and there are remarkable gains in terms of accuracy.

Based on this same idea, this paper proposes an easy-to-apply method of producing non-survey regional and interregional IO tables from national information and data in the form of LQs. Instead of relying on a specific parameter value that sets the values of the adjusted LQ that is applied, the resulting regional IO table is calculated as the simple average of several solutions produced by setting a range of plausible values for the adjusted LQs. The technique is illustrated and evaluated by conducting a numerical simulation similar to the exercise presented in Bonfiglio and Chelli (2009). The results show how this simple approach alleviates largely the arbitrariness present in the choice of specific parameter values for specifying the LQs, while reducing remarkably the risk of producing regional IO tables that deviates from the true ones. The methodology is illustrated for a single input-output table but can be readily extended to multiregional systems.

The paper is organized as follows; the next section briefly reviews approaches to regional input-output model estimation. Section 3 explores the search for empirical regularities while section 4 focuses on location quotient-based adjustments, with a final section reviewing the current state of the art. An alternative testing strategy is proposed om section 5 and the results are presented in section 6 together with a discussion of the findings. A concluding section suggests some needed extensions and modifications.

2. Non-survey Estimates of Regional Input-Output Models[1]

Input-output analysis, especially at the regional or interregional level, has typically been presented

[1] A recent paper by Sargento *et al*.(2024) provides an extensive review of the state of the art

without consideration of uncertainty/errors. This is in sharp contrast to most econometric modeling where attention to estimation methods and associated errors feature prominently. In between, we have CGE modelers who appropriate parameters from the literature often without regard to their provenance (i.e., how they were estimated) or the appropriateness of the re-location from one economy to another.

2.1 Matrix balancing techniques

There is a long tradition, dating back to the 1960s, in the development of non-survey regional models. Schaffer and Chu (1969) represent some early evaluations, mainly centered on simple and cross-location quotient adjustments of national input-output tables. Subsequent research explored a two-economy approach (Round 1978) in which the region of interest and the rest of the national economy comprised the two economies; the estimation technique ensured that the sum of the two economies was consistent with the national total. Stevens *et al*. (1973) initially proposed a regional purchase coefficient approach to adjustment of national tables; the authors claimed this was a more sophisticated differentiator of a region's economy than one generated by the use of location quotients. The RAS technique was popular in the 1980s as was the one proposed by Wilson (1970) who introduced an entropy approach, especially in the context of multiregional modeling.

2.2 Quotient-Based Approaches[2]

While these methodologies attracted considerable attention, the data requirements were somewhat more onerous than the use of location quotients.A recent line in this literature are the works by Flegg and colleagues, who have proposed a modified location quotient approach that features a parameter that captures some measure of agglomeration/regional size. Given a standard input-output system, the national coefficients (a_{ij}) are modified to generate regional coefficients (a_{ij}^r) as $a_{ij}^r = a_{ij} \times lq_{ij}$, where lq_{ij} stands for some type of LQ based on some observable variable X (such as value added or employment) that is available at both the regional and national level. By assumption, if $lq_{ij} > 1$, then $a_{ij}^r = a_{ij}$.

Subsequently, variations to this simple (SLQ) approach were proposed as follows:

[2] This section draws on Flegg *et al*. (2021)

Simple LQ (SLQ): $SLQ_{ij} = SLQ_i = \frac{X_i^R/X^R}{X_i^N/X^N}$ (1)

Cross-Industry LQ (CILQ): $CILQ_{ij} = \frac{X_i^R/X^R}{X_j^N/X^N} = \frac{SLQ_i}{SLQ_j}$ (2)

Flegg Proposal (FLQ): $FLQ_{ij} = \begin{cases} (CILQ_{ij}) \times (\lambda^*); if\ i \neq j \\ (SLQ_{ij}) \times (\lambda^*); if\ i = j \end{cases}$ (3)

Augmented (AFLQ): $AFLQ_i = \begin{cases} (SILQ_{ij}) \times log_2\left[1 + SLQ_j\right]; if\ SLQ_j > 1 \\ FLQ_{ij}; if\ SLQ_j < 1 \end{cases}$ (4)

Flegg's (reference) adjustment relied on the specification of the parameter λ^*. This was defined as follows:

$\lambda^* = [log_2(1 + X^R/X^N)]^{\delta}; 0 \leq \delta < 1$ (5)

In turn this adjustment involved what has come to be known as an agglomeration effect (note the ration of the region's size to that of the nation (X^R/X^N) that was augmented by the parameter δ. Essentially, this could be considered an adjustment factor that modified the input coefficients is response to the region's size. The empirical literature shows that FLQs and AFLQs perform better than other techniques if the correct value of δ is chosen. These results are based on numerical simulation and have been presented at an aggregate and sectoral level.

3. An Alternative procedure for deriving regional tables of IO coefficients

How much reliance we should place on the "preferred" value of the parameter δ (see Flegg *et al.*, 2021)? As Flegg and Tohmo (2013, 2016) have pointed out, there is often considerable variation in the values of δ depending on the specific application of this technique; this compounds the problem about the optimal choice especially if the analyst is focusing on specific sectors within an economy. Since the likelihood of assembling enough survey-based regional input-output tables to conduct a meta analysis is not very high, some other options need to be developed.

Lehtonen and Tykkyläinen (2012) raise the question as to whether the choice of non-survey technique is a gamble; they tested some alternative non-survey techniques using four alternatives including those of Flegg. They employed a variety of tests for the effectiveness of the measures using survey based regional IO tables for Finland. No clear winner was evident and hence the suggestion was made that the choice may be a gamble.

Commented [EF1]: Insert references of the papers that label this Flegg LQ as a "gamble" and the response from Flegg and colleagues

Buendía Azorin *et al*. (2022) proposed some empirical evaluation, based on Korean data, for the estimation of δ from non I-O data – interregional road freight transport and imports from the Rest of the World. They claim better results than using methods based on LQ.

<<insert tables 1 and 2 here>>

Tables 1 and 2, taken from Buendía Azorín *et al*. (2021), provide some insights into their findings. Table 1 provides the results for the FLQ that minimizes the MAPE (Mean Absolute Percentage Error); Table 2 provides comparable information for Spanish regions. Simple correlation analysis revealed poor relationships between the value of δ and regional size for both samples, so it is not clear whether δ is really a measure of size or agglomeration. They also experimented with a more formal econometric approach for the estimation of δ - using exogenous variables such as the share of imports from the rest of the world and the weight of road freight transport from other regions. However, even though these regressors provided a better relationship with δ than regional size, their availability would be significantly limited for most countries.

The underlying problem is very similar to problems in the field of (economic) forecasting where there are different possible forecasts produced by different models and, *a priori*, it is difficult if not impossible to choose the preferred (most accurate) model. In turns out that combining forecasts, instead of relying on one in particular, is common practice since the pioneering works of Bates and Granger (1969) or Newbold and Granger (1974). This would most certainly be the preferred option when one does not have much information for discriminating between competing alternatives

How best to combine the forecasts? If a combination of forecasts is better than relying in one single forecast, how should they be combined? It turns out that a simple mean of forecasts outperforms more sophisticated combination schemes. Several studies, including Makridakis et al. (1982), Makridakis and Winkler (1983) or, more recently, Genre et al. (2013), provide examples of the empirical results of an equally weighted combination. In the current context, the issue centers on the chosen value of δ. One might be tempted to gamble that one value if more likely or reasonable than another; however, there is the further problem than there is considerable heterogeneity of values across sectors for the same region. This should not be surprising since size or agglomeration effects are likely to be different across sectors; the alternative might be just to take an average across a set of feasible solutions. This is the proposed strategy adopted here that will be evaluated by conducting numerical experiments.

4. Adjustment Based on Location Quotients

Bonfiglio and Chelli (2008) evaluated by numerical simulation the performance of several alternative LQs for comparative purposes. They generated and aggregated an MRIO system with $R = 20$ regions and $K = 20$ industries to create a "national" IO table (A^N). Then FLQ and AFLQ (among other LQ options) were used to produce regional matrices $A^P_{r,r}$ as follows:

$A^P_{r,r} = FLQ \times A^N$ or $A^P_{r,r} = AFLQ \times A^N$, with $\delta = 0.1,\ 0.2, \ldots 0.9$. 1,000 simulation draws (6)

Evaluation was made based on Mean Absolute Percentage Errors (MAPEs) between the simulated A_{rr} and predicted $A^P_{r,r}$. FLQ and AFLQ formulations with values of the parameter δ between 0.3-0.4 produced lower deviations on average: if $\delta \leq 0.3$, AFLQ did better than FLQ and it was the preferred coefficient, but for $\delta \geq 0.4$, FLQ did better than AFLQ. However, higher values of parameter δ reduced the variability of the errors.

5. Numerical Experiments

The proposed strategy is an extension of the work of Bonfiglio and Chelli (2008). Assume an MRIO system with **R** regions and **K** industries where:

z^{rs}_{ij} is the trade flow from industry *i* in region *r* to industry *j* in region *s*, with following properties:

- Generally: $z^{rs}_{ij} \sim U(0,100)$

– $z_{ij}^{rs} \sim U(0{,}100) \times$ U(0,0.05) $\times int$U(0,1) if $r \neq s$; to account for (i) much larger intra-regional than inter-regional flows and (ii) potential zero inter-regional trade

– $z_{ii}^{rs} \sim U(0{,}100) \times$ U(1,2); to account for larger intra-sectorial flows

IC_i^r is the intermediate input of industry *i* in region *r*; $IC_i^r = \sum_i^K z_{ij}^{rr}$

v_i^r is the value added of industry *i* in region *r*

$v_i^r = IC_i^r \times$ U(0.5,2); value added in one industry ranges between 50% of the intermediate consumption to double the value

m_i^r are the imports of industry *i* in region *r*; $m_i^r = \sum_i^K z_{ij}^{rs}$, $r \neq s$

x_i^r is the total input of industry *i* in region *r*; $x_i^r = \sum_i^K z_{ij}^{rr} + m_i^r + v_i^r$

The approach may be considered as an extension of Bonfiglio and Chelli (2008):

The Technical (they are really spatial supply coefficients?) coefficients calculated as $a_{ij}^{rs} = z_{ij}^{rs}/x_j^r$

where A_{rr} represents the matrix of technical coefficients in region r. In this paper, only the entries in the principal diagonal are estimated; the sectoral and regional distribution of value added (total output?) is used to calculate the LQs.

Following procedures used by Bonfiglio and Chelli (2008) and others, the comparison of results is made with reference to the MAPEs (Mean Absolute Percentage Error) estimates as follows:

$$MAPE = 100\frac{\sum_{r=1}^{R}\sum_{j=1}^{K}\left|A_{r,r}^{P} - A_{r,r}\right|}{\sum_{r=1}^{R}\sum_{j=1}^{K}\left|A_{r,r}\right|}$$

<<insert figures 1 and 2 here>>

The system has **R**=20 regions and **K**=20 industries; the competing techniques are FLQ and AFLQ with δ = 0.1, 0.2,…, 0.9. The mean of the $A_{r,r}^{P}$ ($\bar{A}_{rr}$) is calculated from the estimates produced by FLQ and AFLQ using the $9 \times 2 = 18$ different values for δ. The results are presented in figures 1 and 2. Figure 1 shows the frequency distribution of the two methods and the mean; the mean provides the highest frequency of smaller MAPEs and a prominent leptokurtic distribution. Figure 2 provides a complementary perspective, showing the frequency of maximum and minimum

MAPEs. Once again, the mean estimator has the lowest frequencies of ,maximum MAPEs and the highest frequency of minimum values.

<<*insert figure 3 here*>>

Figure 3 explores the distributions when the choice of δ is restricted to the set of "optimal values" identified by Bonfiglio and Chelli (2008); these values were 0.3. 0.4 and 0.5. However, the mean was calculated using the nine values of δ used in the first experiments again using FLQ and AFLQ. The reason is that one assumes no a priori information about "optimal values." The results shows a smaller spread of values for the MAPEs for the FLQ and AFLQ distributions but the mean continues to reveal a more leptokurtic distribution. Figure 4 is the comparable presentation to that in figure 2 but with the same restrictions as for figure 3. Figure 5 assumes that there is some consensus around values of δ between 0.3 and 0.5 and this the mean estimates are restricted to these values only. Again, the mean estimator dominates the distribution and achieves a similar dominance in terms of the frequency of minimum (highest) and maximum (lowest) MAPEs, As shown in figure 6.

Esteban: if the only change in figures 7 and 8 the smaller/larger (10/50) number of regions – but the experiment is comparable to figures 1?

Figures 7 and 8 provide the results of some sensitivity analysis by reducing or increasing the number of regions. Reducing the number of regions, figure 7, generates a greater spread in the MAPEs for both of the estimators (FLQ and AFLQ) but appears to have little impact on the mean estimator. When the number of regions is increased, figure 8, the MAPEs for the FLQ and AFLQ cluster in four separate distributions while the mean appears to cluster even more.

6. Evaluation

Although using LQs to produce regional IO matrices from national tables seems useful, the use of FLQ and AFLQ variants is affected by the problem of not knowing, *a priori*, the value of δ to be applied. A solution is to produce alternative A^{P}_{rr} matrices for different values of δ and average the results in the form of a simple mean $\bar{A}_{rr}$. Relying on a small sample of estimates based on a very limited set of survey models may be more difficult to justify.

The preliminary results of the experiments presented in this paper suggest that this strategy yields lower deviations (in terms of MAPEs) on average between real and estimated matrices. The risk of producing large errors derived of a "bad choice" of δ is considerably reduced. In essence, the preferred strategy might be one of generating a variety of estimates and then averaging them rather than trying to choose the "best" option with limited information to guide this choice,

However, there are still issues to be resolved. For example, MAPEs may not be the best measure of evaluation; they do not take into account the analytical importance of the coefficients (see Ramos Cavajal *et al*., 2024 and Sonis and Hewings, 1992. When Flegg-style adjustments have been applied to individual sectors, a range of values of δ has been found suggesting that when, to use Jensen's (1980) distinction, more partitive accuracy is required, some appropriate modification of the averaging approach might be required, perhaps at a second stage of adjustment.

Further, this paper has focused on the intra-regional matrix estimation in a multiregional system. There would seem to be no a priori reason why a similar approach could not be applied to the off-diagonal matrices A_{rs}, by averaging results of gravity-based or entropy-maximizing models that also depends on unknown parameters but with some reasonable expectation about their distributions. Sargento *et al*. (2024) try to place some balance in MRIO estimation by highlighting that the size of the intra-regional flows is likely to be significantly greater than the interregional flows. In this case, it may make sense to adopt an estimation strategy that separates out the two sets of flows.

References

Bates, J.M.; Granger, C.W.J. (1969): "The Combination of Forecasts", Operational Research Quarterly, 20: 451-468.

Beenstock, M. (2023) "Integrating macroeconomics and economic geography: the neoclassical growth model in spatial general equilibrium," *Spatial Economic Analysis*, DOI: 10.1080/17421772.2023.2209598

Bhattacharjee, A., A. Pabst, T. Szendrei and G.J.D. Hewings, (2024) "NiReMS: A regional model at household level combining spatial econometrics with dynamic microsimulation," *Spatial Economic Analysis* (forthcoming)

Blanchflower, D.G. and A.J. Oswald (1995) "An introduction to the wage curve," *Journal of Economic Perspectives*, 9(3), 153-157.

Bonfiglio, A. and F. Chelli (2008) "Assessing the behaviour of non-survey methods for constructing regional input-output tables through a Monte Carlo simulation," *Economic. Systems Research*, 20(3), 243-258.

Buendía-Azorím, J.D., R.M. Alpañez and M. del Mar Sánchez de la Vega (2022) "A new proposal to model regional input–output structures using location quotients. An application to Korean and Spanish regions," *Papers in Regional Science*, 101(5), 1219-1237 DOI: 10.1111/pirs.12692

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

Flegg A.T. and T. Tohmo (2016) “Estimating regional input coefficients and multipliers: the use of FLQ is not a gamble,” Regional Studies, 50, 310–325. https://doi.org/10.1080/00343404.2014.901499

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

Flegg A.T. and C.D. Webber (1997) “On the appropriate use of location quotients in generating regional input–output tables: reply,” *Regional Studies*, 31, 795–805. https://doi.org/10.1080/713693401.

Flegg A.T. and C.D. Webber (2000) “Regional size, regional specialization and the FLQ formula,” *Regional Studies*, 34, 563–569. https://doi.org/10.1080/00343400050085675.

Flegg A.T., C.D. Webber and M.V. Elliott (1995) “On the appropriate use of location quotients in generating regional input–output tables,” *Regional Studies* 29, 547–561. https://doi.org/10.1080/00343409512331349173.

Flegg A.T., L.J. Mastronardi and C.A. Romero (2016) “Evaluating the FLQ and AFLQ formulae for estimating regional input coefficients: empirical evidence for the province of Córdoba, Argentina,” Economic Systems Research, 28, 21–37. https://doi.org/10.1080/09535314.2015.1103703

Flegg, A.T, G.R. Lamonica, F.M. Chelli, M.C. Recchioni and T. Tohmo (2021) “A new approach to modelling the input–output structure of regional economies using non-survey methods,” *Journal of Economic Structures*, https://doi.org/10.1186/s40008-021-00242-8

Hewings, G.J.D., M. Sonis, J. Guo, P.R. Israilevich and G.R. Schindler, (1998) “The hollowing out process in the Chicago economy, 1975-2015,” *Geographical Analysis*, 30, 217-233.

Hewings, G.J.D. and J.B. Parr (2009) “The Changing Structure of Trade and Interdependence in a Mature Economy: The US Midwest,” In P. McCann (ed.) *Technological Change and Mature Industrial Regions: Firms, Knowledge, and Policy*, Cheltenham, UK, Elgar, pp. 64-84.

Genre, V.; Kenny, G.; Meyler, A.; Timmermann, A. (2013): “Combining expert forecasts: Can anything beat the simple average?” International Journal of Forecasting, 29: 108-121.

Jensen, R.C. (1980) “The concept of accuracy on regional input-output models,” *International Regional Science Review*, 5, 139-154.

Jones R.W., Kierzkowski, H. (2001) A framework for fragmentation. In: Arndt SW, Kierzkovski H (eds) *Fragmentation: New Production Patterns in the World Economy*. New York, Oxford University Press.

Kratena, K. (2023) “Effective demand, wages and prices, and the multiplier,” *Economic Systems Research*, DOI: 10.1080/09535314.2023.2204393

Krugman P.R. (1996) *Rethinking International Trade*. Cambridge, MA, The MIT Press.

Krugman P.R. (1981) “Intraindustry specialization and the gains from trade,” *Journal of Political Economy*, 89, 959–974.

Lehtonen, O. and M. Tykkyläinen (2012) “Estimating regional input coefficients and multipliers: is the choice of a non-survey technique a gamble,” *Regional Studies*, 48(2), 382–399. https://doi.org/10.1080/00343404.2012.657619

Makridakis, S. A.; Andersen, R.; Carbone, R.; Fildes, M.; Hibon, R.; Lewandowski, J.; Newton, E.; Parsen, and R. Winkler (1982) “The accuracy of extrapolation (time series) methods: results of a forecasting competition”, Journal of Forecasting, 1: 111- 153.

Makridakis, S.; Winkler, R.L. (1983): “Averages of Forecasts: some empirical results”, Management Science, 29: 987-996.

Newbold, P.; Granger, C.W.J. (1974): “Experience with Forecasting Univariate Time Series and the Combination of Forecasts”, Journal of the Royal Statistical Society, Series A 137: 131-165.

Nijkamp, P. and J. Poot (2002) “The Last Word on the Wage Curve?,” *Discussion Paper*, No. 02-029/3, Tinbergen Institute, Amsterdam and Rotterdam.

Okazaki, F. (1989) “The 'Hollowing-Out' Phenomenon in Economic Development,” Paper presented at the Pacific Regional Science Conference, Singapore.

Ramos Cavajal, C., E. Lasarte Navamuel and G.J.D. Hewings, (2024) “Some Considerations on Assessing the Importance of a Coefficient,” *Socio-Economic Planning Sciences*, 91, 101765. https://doi.org/10.1016/j.seps.2023.101765

Romero, I., E. Dietzenbacher and G.J.D. Hewings (2009) “Fragmentation and complexity: analyzing structural change in the Chicago regional economy,” *Revista de Economía Mundial*, 23 263-282.

Rokicki, B., O. Fritz, J.M. Horridge and G.J.D. Hewings (2021) “Survey-based versus Algorithm-based Multi-Regional Input-Output Tables within the CGE framework – the case of Austria” *Economic Systems Research,* 33(4), 470-491. DOI: 10.1080/09535314.2020.1839385

Sargento, A.L.M., M.L. Lahr, J.P. Ferreira, F. de la Torre Cuevas (2024) “Revisiting Methods for Estimating Interregional Input-Output Accounts: It’s Not Just About Trade Flows,” *Structural Change and Economic Dynamics,* 69, 664-677.

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

Sonis, M.and G.J.D. Hewings, (1992) “Coefficient change in input-output models: theory and applications,” *Economic Systems Research* 4, 143-157

Stevens, B.H., G.I. Treyz, D.J. Ehrlich and J.R. Bower (1983) “A new technique for the construction of non-survey regional input-output models and comparisons with two survey-based models,” *International Regional Science Review*, 8, 271-286.

Wilson, A.G. (1970) *Entropy in urban and Regional Modelling*, London, Pion.

Table 1: Estimation of optimal δ and regional size for Korean regions

Regions	Regional size[a] (%)	Optimal δ^*
1. Gyeonggi-do	22.85	0.545
2. Seoul	18.97	0.186
3. Gyeongsangbuk-do	7.00	0.353
4. Chungcheongnam-do	6.96	0.642
5. Gyeongsangnam-do	6.93	0.293
6. Ulsan	6.32	0.594
7. Incheon	4.96	0.469
8. Jeollanam-do	4.89	0.288
9. Busan	4.73	0.281
10. Chungcheongbuk-do	3.47	0.487
11. Jeollabuk-do	2.82	0.335
12. Daegu	2.82	0.26
13. Gwangju	2.07	0.345
14. Gangwon-do	1.97	0.284
15. Daejeon	1.92	0.451
16. Jeju-do	0.81	0.208
17. Sejong	0.50	0.605
Mean	5.88	0.390
Weighted mean		0.40
Coefficient of variation	1.03	0.37

[a]significant at 1%.
Authors' calculations for Korea MRIO 2015.

Source: Buendía-Azorím *et al*. (2022)

Table 2: Estimation of optimal δ and regional size for Spanish regions

	Regional size[a] (%)	Optimal δ^*
Catalonia 2011	20.53	0.883
Madrid, Community of 2010	18.87	0.798
Andalusia 2010	13.20	0.211
Basque Country 2015	6.62	0.251
Galicia 2011	5.44	0.233
Canary Islands 2005	3.50	0.240
Castilla-La Mancha 2005	3.48	0.259
Aragon 2005	3.25	0.640
Balearic Islands 2004	2.19	0.091
Asturias, Principality of 2015	1.89	0.150
Navarra, Community of 2010	1.88	0.132
Cantabria 2015	1.09	0.129
Rioja, La 2008	0.77	0.389
Mean	6.36	0.339
Weighted mean		0.53
Coefficient of variation	1.06	0.78

[a]Share of gross output.
Authors' calculations for Spain regional IOTs

Source: Buendía-Azorím *et al*. (2022)

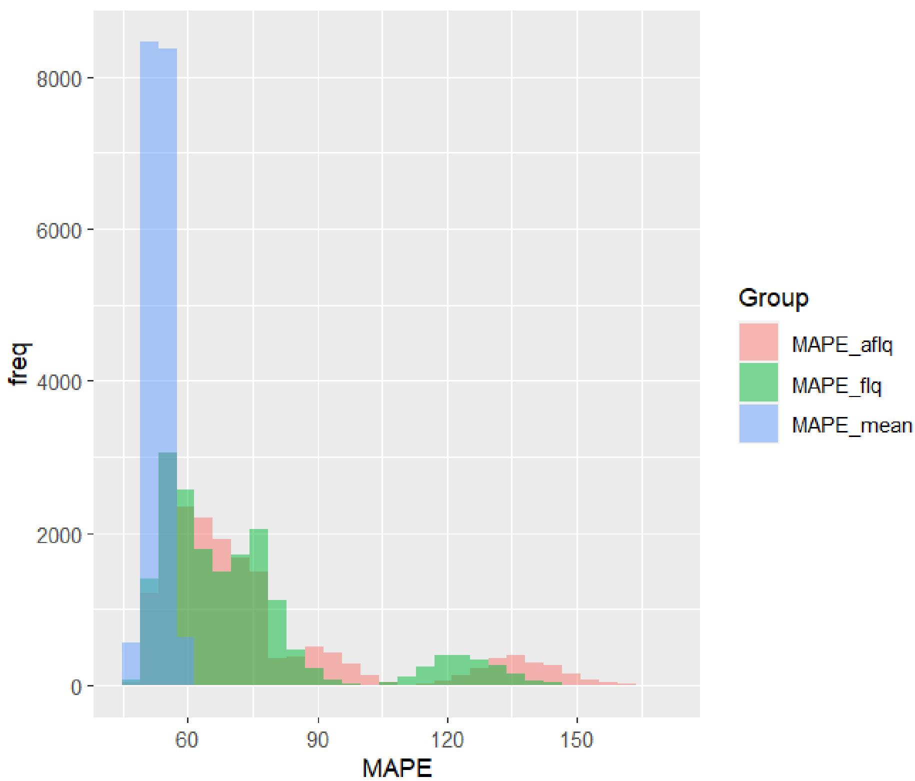

Figure 1: MAPE Frequency Distributions for the Three Alternative Estimators

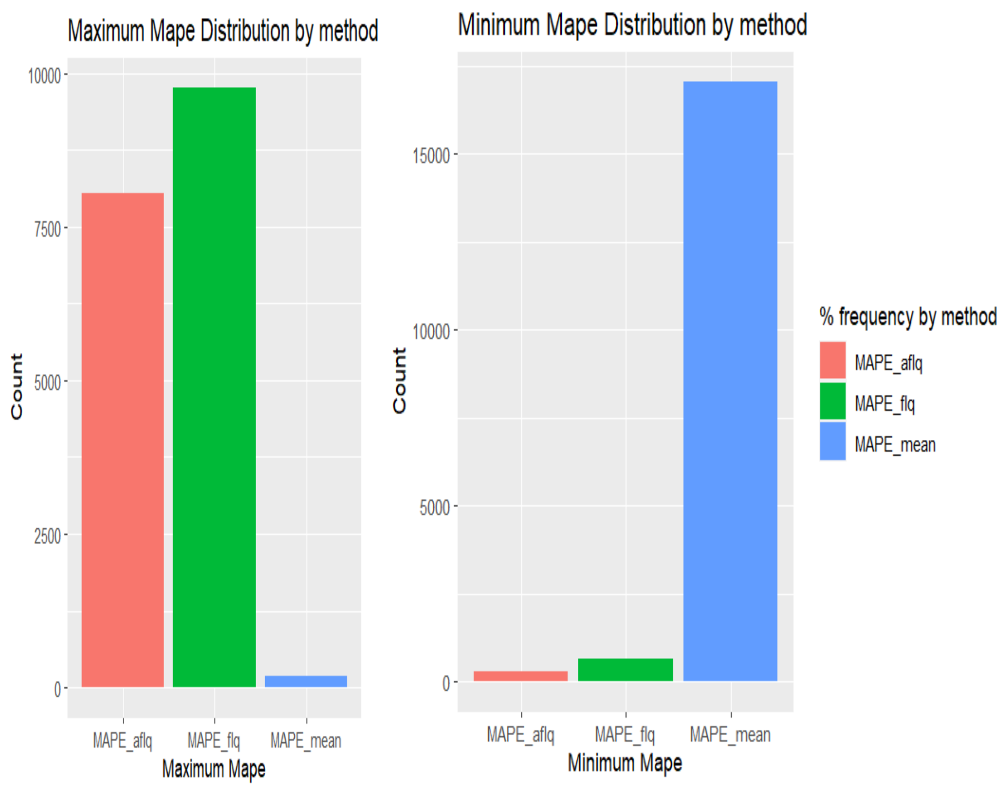

Figure 2: Maximum and Minimum Frequency Distributions for the Three Alternative Estimators

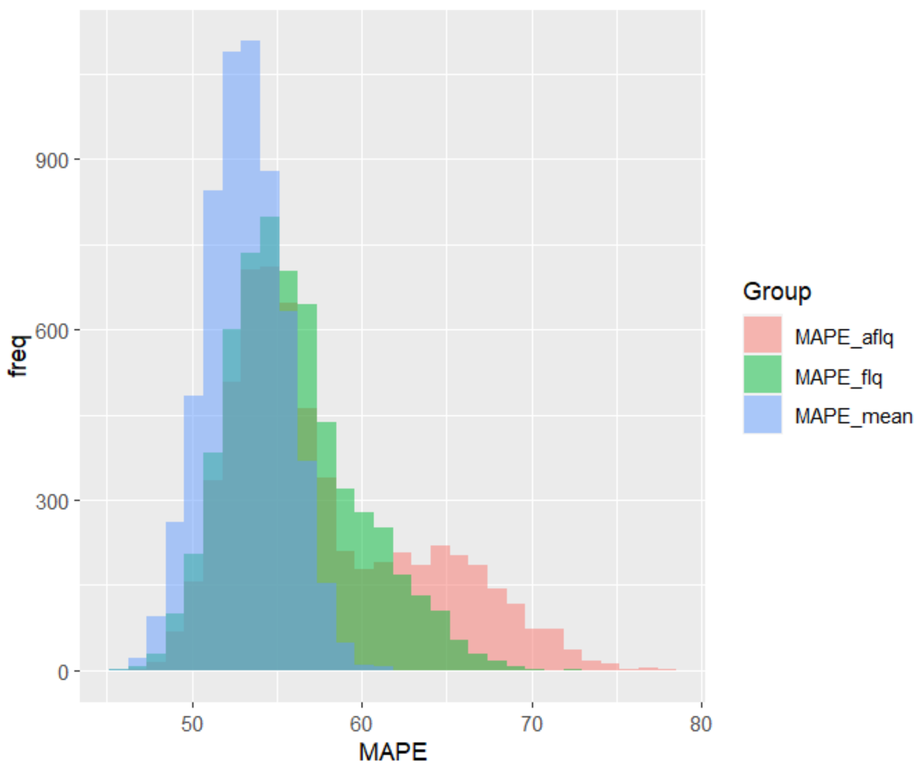

Figure 3: MAPE Frequency Distributions for the Three Alternative Estimators with only three choices for δ following Bonfiglio and Chelli (2008) for FLQ and AFLQ

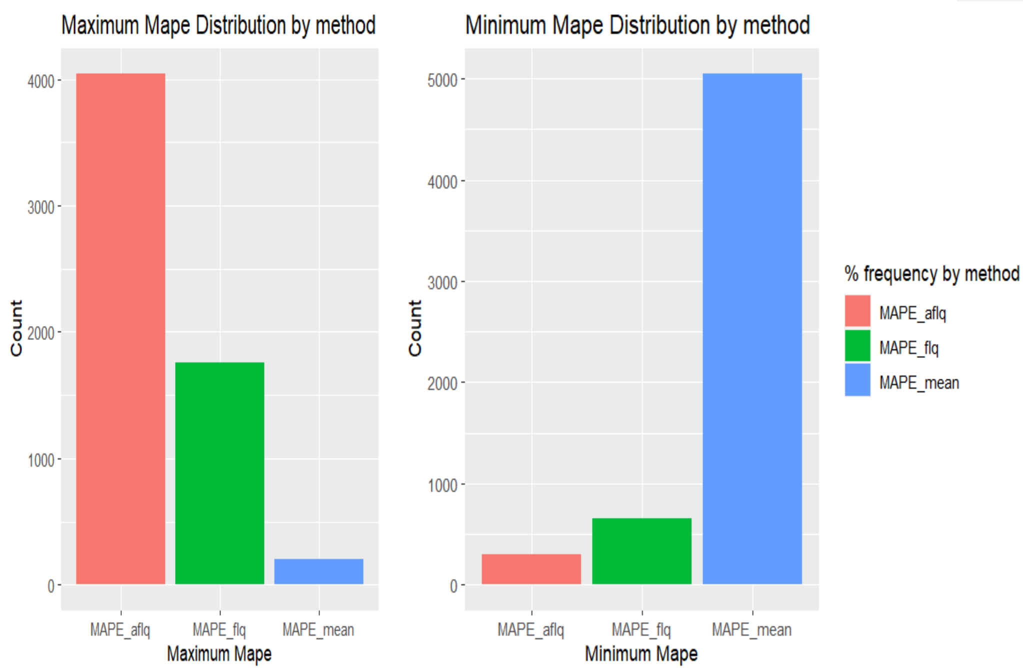

Figure 4: Maximum and Minimum Frequency Distributions for the Three Alternative Estimators with only three choices for δ for FKLQ and AFLQ following Bonfiglio and Chelli (2008)

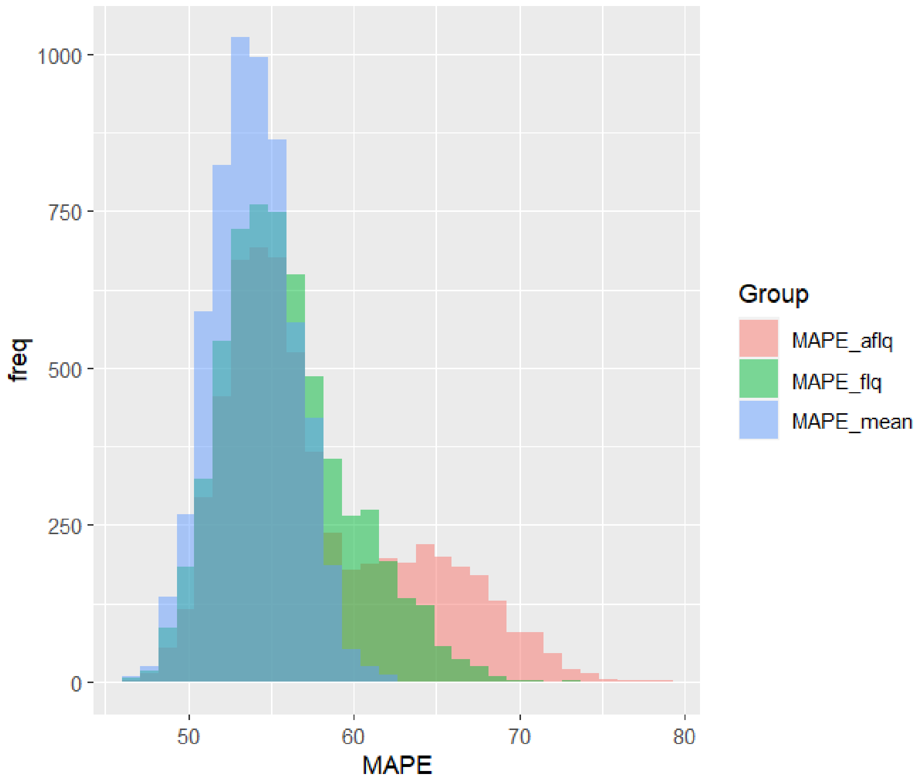

Figure 5: MAPE Frequency Distributions for the Three Alternative Estimators with only three choices for δ following Bonfiglio and Chelli (2008) for all estimators

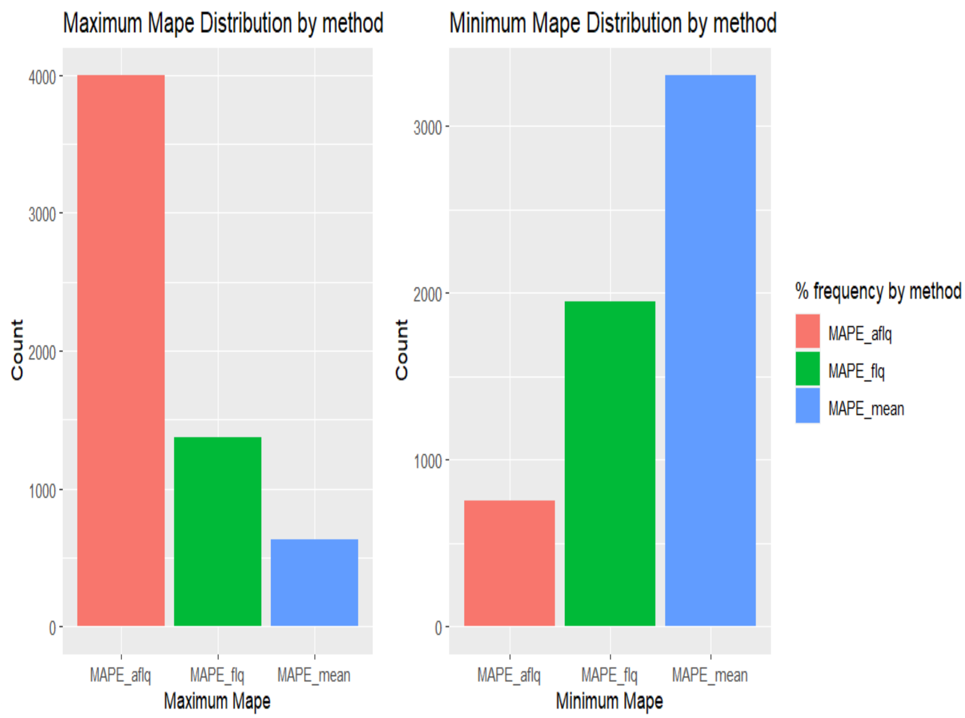

Figure 6: Maximum and Minimum Frequency Distributions for the Three Alternative Estimators with only three choices for δ for all estimators (FLQ, AFLQ and Mean) following Bonfiglio and Chelli (2008)

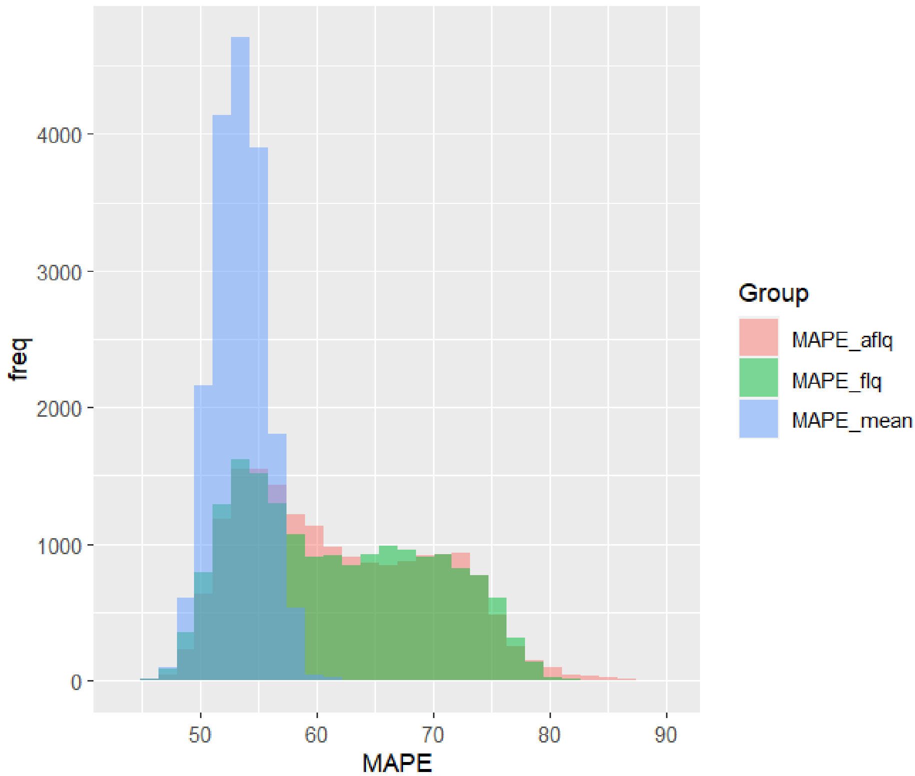

Figure 7: MAPE Frequency Distributions for the Three Alternative Estimators with only 10 regions

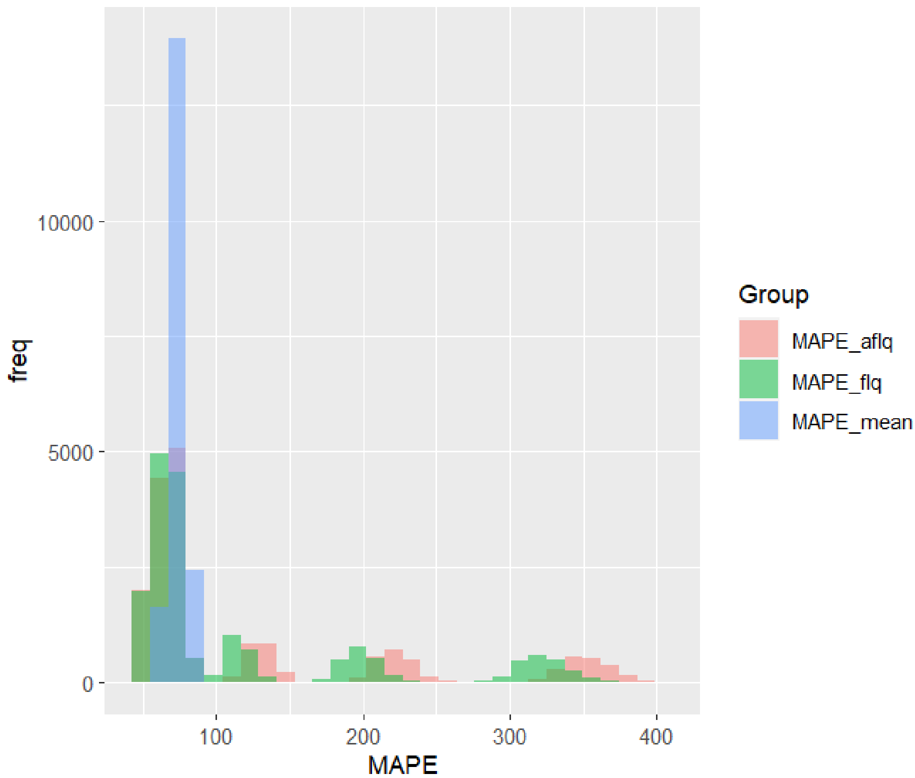

Figure 8: MAPE Frequency Distributions for the Three Alternative Estimators with 50 regions