National Accounting Review

NAR, 7(1): 125–142.
DOI: 10.3934/NAR.2025006
Received: 11 November 2024
Revised: 17 February 2025
Accepted: 03 March 2025
Published: 10 March 2025

https://www.aimspress.com/journal/NAR

Research article

Estimating SFLQ-based regional input-output tables for South Korean regions

Timo Tohmo*

School of Business and Economics, University of Jyväskylä, Jyvaskyla, Finland

*** Correspondence:** Email: timo.tohmo@jyu.fi.

Abstract: The present study employed official input-output data for 17 regions in South Korea from 2015 to analyze the industry-specific Flegg location quotient (SFLQ) formula as a tool for estimating regional input-output tables. The paper builds on the work of Kowalewski (2015), who proposed an interesting variant of the Flegg location quotient (FLQ) by letting the value of the parameter δ in the FLQ vary across industries. The main aim of the present study was to employ Kowalewski's SFLQ formula to produce estimated regional input-output tables for South Korea and to examine the accuracy of the resulting estimates. We estimated Kowalewski's regression model for all Korean regions in 2015, something that has not previously been performed. Subsequently, we compared the regionalized SFLQ-based estimates with the official regional input-output tables produced by the Bank of Korea. We tested the accuracy of the estimations using different statistics.

Keywords: SFLQ; FLQ; location quotients; South Korea; input-output tables

JEL Codes: C13, C67, O18, R15

1. Introduction

Regional planners and researchers often wish to know the regional economic effects of certain investments or events. Thus, regional input-output (IO) tables are tools commonly used for regional planning. However, survey-based regional input-output tables are typically not available. Furthermore, producing an input-output table based on a survey of establishments in the economy is a complex, time-consuming, and expensive task.

If suitable regional input-output tables are unavailable, one common way to proceed is to regionalize national input-output tables. Non-survey methods enable regional input-output analysis to be performed without employing prohibitively expensive techniques to adjust national coefficients. One very common non-survey approach is to use location quotients (LQs) (see Miller & Blair, 2009). Various LQs have been employed in earlier studies as non-survey techniques (e.g., Schaffer & Chu, 1969; Hewings, 1971; Morrison & Smith, 1974; Flegg et al., 1995 and Flegg & Webber, 2000).

The accuracy of the well-known Flegg location quotient (FLQ) in producing regional input-output tables has been examined in several studies, which demonstrated that the FLQ clearly outperformed previous methods (see, e.g., Flegg & Webber, 2000; Tohmo, 2004; Flegg & Tohmo, 2013a, 2016; Flegg et al., 2016 and Bonfiglio & Chelli, 2008). Kowalewski (2015) introduced an interesting variant of the FLQ by relaxing the supposition that the value of δ in the FLQ is constant across industries. She applied the method to official survey-based data for Baden-Württemberg, Germany.

The Bank of Korea has produced official tables for the year 2015 covering all 17 regions in South Korea. Our study primarily seeks to analyze the use of Kowalewski's industry-specific FLQ, the SFLQ, to regionalize the national input-output table of South Korea and to test the accuracy of the regional input-output tables it can produce. We re-estimate Kowalewski's regression model for all Korean regions in 2015. Consequently, we compare the regionalized SFLQ-based estimates with official input-output tables and test the accuracy of the estimations with different statistics.

In section 2, we discuss the FLQ formula and some other widely used LQs. In section 3, we present some key findings from the results produced by the simple location quotient (SLQ), cross-industry location quotient (CILQ), and FLQ regionalization methods. The main aim of our study is to analyze the accuracy of the SFLQ in producing regionalized input-output tables. In section 4, the industry-specific results produced by SFLQ are analyzed, and several statistics are applied to assess the accuracy of the estimates. Section 5 concludes the paper.

2. The SFLQ and related formulae

Miller & Blair (2009) argued that regional input-output studies attempt to quantify the impacts caused by new final demands for products made in a specific region. The data needed for regional analysis and for the construction of regional input-output tables can be gathered from surveys that are based on inquiries, interviews, and statistics. Surveys are regarded as reliable methods for constructing regional input-output coefficients; however, producing a regional input-output table based on surveys and interviews is time-consuming, and the costs would normally exceed the budget allocated for research.

Typically, researchers face a lack of regional input-output figures, and using surveys to construct regional input-output tables requires too much money and time. One option is to adjust the national coefficients to produce a regional table using techniques that employ available statistical data. One very common approach to estimating regional input-output coefficients from national figures is to use non-survey methods; recently, such methods have been significantly developed.

However, defining exactly what constitutes a non-survey method is difficult because, in practice, in the construction of regional input-output tables, indicators of ad hoc judgment or some kinds of data-smoothing techniques are used (Round, 1983). Consequently, input-output tables are essentially hybrid tables based on semi-survey techniques that, to some extent, use primary and secondary sources.

The Korean interregional input-output tables for the year 2005 are arguably one of the very few survey-based interregional IO tables available. Zhao & Choi (2015) argued that they serve as benchmark tables since they are entirely derived from surveys (see also Flegg et al., 2021).

However, the Bank of Korea (BOK) does not provide details on how the interregional input-output tables for the year 2015 were compiled. We are unaware of any reference indicating whether the BOK collected detailed data directly from businesses and industries within each region including information on production, consumption, and inter-industry transactions. Nevertheless, we think that it is reasonable to suppose that the compilation method has not changed significantly over the years and that the 2015 IO tables should, therefore, serve as good benchmark tables.

One very commonly used category of methods to produce regional input-output tables is the LQ approach. Several alternative variants of LQs have been proposed and tested. The rationale for these alternatives is discussed in, for example, Schaffer & Chu (1969), Morrison & Smith (1974), Round (1978), Harrigan et al. (1980), Flegg et al. (1995), Flegg & Webber (1997, 2000), Bonfiglio & Chelli (2008), Kowalewski (2015), and Flegg & Tohmo (2013a, 2013b).

To apply LQs requires the regional and national transaction matrices to be transformed into input coefficient matrices by “regionalization” of the national coefficient matrix. The regional input coefficients, r_{ij}, are computed from the corresponding national coefficients, a_{ij}, via the following formula:

$$r_{ij} = t_{ij} * a_{ij} \tag{1}$$

where r_{ij} is the regional input coefficient, which measures the amount of regional input *i* needed in one unit of regional output j. Inputs from abroad and from other regions are excluded. a_{ij} is the national coefficient, and it excludes foreign inputs. t_{ij} is the adjustment coefficient, which is assumed to take into account the region’s inputs of i from other regions and can be estimated via LQs (Flegg & Tohmo, 2016). Therefore,

$$r_{ij} = LQ_{ij} * a_{ij} \tag{2}$$

LQs are regarded as a useful method for regionalizing a national input-output table. Researchers have often used either the simple location quotient (SLQ) or the cross-industry location quotient (CILQ) for this purpose. The SLQ is defined as follows:

$$SLQ_i = (\frac{RE_i}{NE_i}) \times (\frac{TNE}{TRE}) \tag{3}$$

where RE_i is the regional output or employment in sector *i* for a given region, while NE_i refers to the corresponding national figure. TRE and TNE are the regional and national totals, respectively. Employment may be affected by differences in regional productivity. Consequently, output is a preferable measure, but employment is probably the most common measure used in research projects because of the lack of output data.

Whenever LQ_i > 1, sector *i* is more concentrated in the region than the nation as a whole. Sectors where LQ < 1 are reduced by multiplying them by the SLQ, which will increase the import coefficients by a corresponding amount. However, no adjustment is made for sectors with an LQ above one (see Isard (1960) for a discussion of the SLQ). In that case, the region is presumed to satisfy its own demand, and the regional input-output coefficients are set equal to the corresponding national coefficients, which leads to the truncation of the SLQ to one (SLQ = 1).

The SLQ is a widely used regionalization method, but it has a drawback: only the size of the selling industry is considered (Morrison & Smith, 1974).

The CILQ for sectors *i* and *j* is defined (see Miernyk (1968) for a description of the CILQ) here as follows:

$$CILQ_{ij} = \left(\frac{SLQ_i}{SLQ_j}\right) = (\frac{RE_i}{NE_i})/(\frac{RE_j}{NE_j}) \tag{4}$$

where RE_i and NE_i denote regional and national output (or employment), respectively, of sector *i* (Flegg & Webber, 1997; Miller & Blair, 2009). In this context, the subscripts *i* and *j* are used to indicate the supplying and purchasing sectors, respectively. No scaling is applied to a_{ij} where $CILQ_{ij}$ ≥ 1, and the CILQ is truncated to $CILQ_i$ = 1 whenever $CILQ_i$ > 1. This truncation indicates that, in cases where the regional supplying sector is relatively small compared to the purchasing sector, some of the inputs must be imported from abroad or from other regions.

Consequently, the national coefficients are adjusted downward by multiplying them by the CILQ. Smith & Morrison (1974) suggested using the SLQs along the principal diagonal of the CILQ because the CILQ does not consider the size of the local industry. Thus, when adjusting the a_{ij} (national coefficients), it is reasonable to use the SLQ along the principal diagonal and to use the CILQ elsewhere. One of the key advantages of the CILQ is its flexibility in allowing different scalings to be applied to individual cells within the national coefficient matrix.

The SLQ and CILQ are commonly used methods in regional analysis. However, they are known to understate regional trade substantially, leading to an overestimation of regional economic impacts.

Flegg et al. (1995) sought to overcome the drawbacks of the SLQ and CILQ methods by introducing their own formula, known as the Flegg location quotient (FLQ). This approach was subsequently refined by Flegg & Webber (1997). The FLQ is defined as follows:

$$FLQ_{ij} = CILQ_{ij} \times \lambda * \quad for\ i \neq j \tag{5}$$

$$FLQ_{ij} = SLQ_i \times \lambda * \quad for\ i = j \tag{6}$$

where

$$\lambda * \equiv [\log_2(1 + \frac{TRE}{TNE})]^{\delta} \tag{7}$$

The parameter δ controls the convexity in Equation (7), and it is constrained to the range 0 ≤ δ < 1. When δ increases, the value of λ^* decreases, which in turn results in a greater allowance for interregional imports. If δ =0, then the FLQ and CILQ formulae coincide ($FLQ_{ij} = CILQ_{ij}$ for $i \neq j$

and $FLQ_{ij} = SLQ_i$ for $i = j$). TRE/TNE is the ratio of regional to national output or employment. No scaling is applied where $FLQ_{ij} \geqslant 1$, and the FLQ is truncated to $FLQ_i = 1$ whenever $FLQ_i > 1$.

Several studies have analyzed the accuracy of various LQs and demonstrated that the FLQ is frequently more accurate than the SLQ and the CILQ approaches. Such studies include Flegg & Webber (2000) (Scotland), Tohmo (2004) (Finland), Flegg & Tohmo (2013a, 2016) (Finland), Kowalewski (2015) (Germany), Flegg et al. (2016) (Argentina), and Bonfiglio & Chelli (2008) (Monte Carlo simulation of 400,000 output multipliers).

Mardones and Silva (2021) contributed to this discussion by applying non-survey methods to estimate regional input coefficients and multipliers for Chile. Their work aligns with previous studies that have found FLQ-based methods to yield more accurate estimates than traditional approaches like the SLQ and CILQ. Furthermore, their 2023 study assessed the performance of different non-survey techniques, including Monte Carlo simulations, to address challenges related to incomplete data. Their 2024 article introduces a methodological framework for approximating the sectoral impacts of a carbon tax at the regional level. By incorporating Monte Carlo simulations, the study provides a robust way to assess uncertainties in regional economic responses to environmental taxation.

The work of Mardones et al. is relevant for policymakers and researchers by dealing with incomplete datasets, allowing for a more nuanced understanding of how different sectors and regions might be affected by policy changes, and by suggesting improvements for regional economic modeling.

Kowalewski (2015) introduced a variant of the FLQ (the SFLQ). She relaxed the supposition that the value of δ is constant across industries (the values of δ vary by industry). Kowalewski's SFLQ is defined as follows:

$$SFLQ_{ij} = CILQ_{ij} \times [\log_2(1 + \frac{TRE}{TNE})]^{\delta_j} \tag{8}$$

where TRE/TNE denotes regional size, which is measured by employment. To derive estimates of the δ_j, Kowalewski defines the following regression model:

$$\delta_j = \alpha + \beta_1 CL_j + \beta_2 SLQ_j + \beta_3 IM_j + \beta_4 VA_j + \varepsilon_j \tag{9}$$

where VA_j is the share of value added in total national output, IM_j is the proportion of foreign imports in the overall inputs used across the entire national economy, CL_j is a coefficient of localization, which gauges the concentration of industry j at the national level, and ε_j is an error term. In turn, CL_j is calculated as follows:

$$CL_j = 0.5 \sum_r \left| \frac{E_j^r}{E_j^n} - \frac{E^r}{E^n} \right| \tag{10}$$

where E_j^r is the regional employment in regional selling industry j, E_j^n is the national employment in selling industry j, and E^r and E^n are total regional and national employment.

When regionalizing national input-output tables, LQ methods exploit employment or output figures. The size of a region can also be measured by the value added produced regionally. However, researchers commonly use employment figures because output or value-added figures are not readily available. One drawback of using employment figures is that they might be distorted by variations in regional productivity. Consequently, in our study, output figures are a preferred measure, especially because they are regionally available for all industries.

3. The SLQ-, CILQ-, and FLQ-based multiplier estimates

In this section, we present a selection of results concerning the accuracy of the SLQ, CILQ, and FLQ in producing South Korean regional estimates. Different LQs are used to regionalize the South Korean national input-output table for 2015. The analysis is based on a 33 × 33 national coefficient matrix for the year 2015 developed by the Bank of Korea. The country was partitioned into 17 regions by the Bank, and its official regional input-output tables are used as a reference point for evaluating the precision of the simulations. As a criterion, the mean absolute percentage error (MAPE) is used as follows:

$$\mathrm{MAPE} = \left(\frac{100}{n}\right)\sum\nolimits_j \left|\widehat{m}_j - m_j\right| / m_j \quad (11)$$

where $\widehat{m}_j - m_j$ is the estimated type I output multiplier for sector j minus the corresponding benchmark value m_j, and n = 33 represents the total number of industries considered in the analysis.

The advantages of the MAPE include the fact that it uses all observations and that its variation is low from one sample to another. In addition, the MAPE is expressed in percentage form, which makes it widely understandable. However, the MAPE may not fully encompass all the relevant factors pertaining to the choice of an accurate measurement approach.

The analysis builds upon earlier studies (Flegg & Tohmo, 2013a, 2016, 2018; Flegg et al., 2016), so a range of criteria with different features is employed. The following supplementary statistics are employed to gauge the accuracy of the estimations (multipliers):

$$\mathrm{MPE} = \left(\frac{100}{28}\right)\sum\nolimits_j \left|\widehat{m}_j - m_j\right| / m_j \quad (12)$$

$$\mathrm{WMPE} = 100\sum\nolimits_j w_j(\widehat{m}_j - m_j)/m_j \quad (13)$$

$$\mathrm{SDSD} = \{\mathrm{sd}(\widehat{m}_j) - \mathrm{sd}(m_j)\}^2 \quad (14)$$

$$\mathrm{U} = 100\sqrt{\frac{\sum_j (\widehat{m}_j - m_j)^2}{\sum_j {m_j}^2}} \quad (15)$$

How much do the approximate values differ from the exact values in percentage (on average) of the actual values? This difference can be measured by the mean percentage error (MPE). The MPE has been used in many previous studies because it measures an appropriate amount of bias (relatively). However, the MAPE and MPE results are found to produce quite similar results. Sometimes, it is reasonable to weigh cases to prevent smaller values from being considered to be equal to higher values.

The weighted mean percentage error (WMPE) criterion utilized in this study accounts for the relative significance of each sector by weighting the absolute percentage errors with the proportion of the total regional output contributed by sector j (weight w_j is used). Additionally, the SDSD (squared difference in standard deviations) is also used (the standard deviation indicates the dispersion of the data spread and how representative the mean is). The squaring in the procedure for calculating the standard deviation emphasizes the extremes (larger differences). However, it always gives positive values, hindering the possibility that opposite values cancel each other, leading to zero-sum results. Thus, the SDSD assesses the extent to which each method can replicate the spread and distribution of the benchmark multipliers, thereby indicating how representative the mean value is.

Theil's index of inequality, denoted by U, is a measure used to assess accuracy. Although the Theil index can be regarded as mathematically complex, it is used as one criterion of accuracy because it has the advantage of encompassing both bias and variance (Theil et al., 1966). A range of the SLQ, CILQ, and FLQ results is presented in Table 1. The results are based on 33 sectors in 17 Korean regions in 2015.

Table 1. Accuracy of the multipliers for South Korean regions in 2015 (drawing upon data from 33 sectors across 17 regions).

Method	Criterion				
	MAPE	MPE	WMPE	SDSD × 10^3	U
SLQ	19.873	22.645	23.311	4.993	22.557
CILQ	20.936	20.385	17.598	3.333	23.377
FLQ (δ = 0.15)	12.176	10.285	7.753	0.452	15.175
FLQ (δ = 0.175)	11.083	8.717	6.271	0.442	13.906
FLQ (δ = 0.2)	10.148	7.204	4.851	0.546	12.739
FLQ (δ = 0.25)	8.763	4.394	2.206	0.796	10.836
FLQ (δ = 0.3)	7.889	1.828	−0.216	1.347	9.568
FLQ (δ = 0.325)	7.676	0.656	−1.316	1.977	9.195
FLQ (δ = 0.35)	7.5764	−0.456	−2.351	2.764	8.973
FLQ (δ = 0.375)	7.5840	−1.503	−3.316	2.764	8.897
FLQ (δ = 0.4)	7.700	−2.496	−4.242	2.782	8.962
FLQ (δ = 0.425)	7.899	−3.427	−5.103	3.644	9.133

Source: The results are based on calculations performed by the authors. The results presented herein are based on an unweighted mean of the findings obtained for each of the 17 regions examined.

The WMPE indicates that the FLQ, with an optimum $\delta \approx 0.3$, gives the most accurate results in terms of the multipliers. However, the MAPE, MPE, and U indicate that the best accuracy is achieved with the FLQ, with an optimum $\delta \approx 0.35$–0.375. The difference is interpreted in a manner that may necessitate a smaller δ for sectors that are comparatively larger. Notably, a δ in the range of 0.4 ± 0.025 will yield a MAPE ≈ 8.0%. Table 1 reveals that the SDSD indicates that the FLQ, with an optimum $\delta \approx 0.175$, gives the best outcome of the regionalization of input-output national tables in terms of multipliers. It is well-known that a low mean may lead to high variance and vice versa. In the case of South Korean data, this seems to be an issue because there are conflicts in minimizing bias and variance.

Table 2 shows the MAPE results for all regions. In broad terms, the FLQ produces more accurate results than the SLQ and CILQ, based on all the different criteria. The SLQ and CILQ methods produce very similar outcomes, and both exhibit very low accuracy levels. For example, the MAPE shows that the SLQ and CILQ overestimate the multipliers by around 20% on average across the 17 South Korean regions.

The study estimated the value of *δ* that produces the minimum MAPE in each region. *δ* was varied incrementally by 0.0001 until the optimum (minimum) was reached. The identified optimum values of *δ*, yielding the minimum MAPE in each region, are highlighted in bold in Table 2. Twelve of the optimum values of δ are in the range of 0.35 ± 0.05, yielding an average MAPE of approximately 8%. This result suggests that in the regionalization of the South Korean national input-output table, the FLQ method using δ in the range of 0.35 ± 0.05 leads to multipliers that differ from those of survey-based regional tables by an average of 8%.

How can the performance of a given method of producing regional input-output tables be assessed? Jensen (1980) argued that one barrier is the failure to agree on acceptable levels of accuracy in the use of the model. Jensen & MacDonald (1982) argued that there is some agreement that the accuracy requirements will vary according to the intended use of the table but that there is no agreement on how to assess them. Miernyk (1976) considered any error higher than 50% to be unacceptable and errors larger than 100% to be large. Accordingly, an error of 8% seems acceptable.

Table 2. Estimating the output multipliers for all regions in South Korea in 2015 using the SLQ, CILQ, and FLQ with varying δ values (with MAPE calculated across 33 sectors).

	Region	Minimum MAPE	δ for minimum MAPE	FLQ 0.20 MAPE	FLQ 0.30 MAPE	FLQ 0.35 MAPE	FLQ 0.40 MAPE	FLQ 0.45 MAPE
1	Gyeonggi-do	4.20	0.514	12.69	8.79	7.20	5.90	4.85*
2	Seoul	9.15	0.328	9.31	9.16	9.15	9.23	9.36
3	Gyeongsangbuk-do	5.04	0.401	9.55	6.03	5.28	5.04	5.40
4	Gyeongsangnam-do	4.87	0.319	7.29	4.93	5.15	6.13	7.11
5	Ulsan	6.54	0.480	12.01	8.74	7.63	6.89	6.57*
6	Jeollanam-do	8.36	0.380	10.56	8.72	8.387	8.394	8.44
7	Chungcheongnam-do	6.99	0.4507	13.41	9.33	7.86	7.20	6.99
8	Incheon	4.69	0.414	11.74	6.64	5.26	4.72	4.87
9	Busan	5.90	0.302	7.62	5.90	6.57	7.35	8.50
10	Chungcheongbuk-do	6.70	0.377	9.78	7.22	6.7883	6.7884	7.35
11	Daegu	7.14	0.269	8.44	7.39	8.16	9.52	10.92
12	Jeollabuk-do	6.72	0.301	8.28	6.72	7.45	8.79	10.23
13	Gangwon-do	8.85	0.249	9.18	9.12	9.22	9.54	10.41
14	Gwangju	8.12	0.262	8.86	8.29	8.80	9.61	10.43
15	Daejeon	7.79	0.379	11.86	8.47	7.89	7.86	8.60
16	Jeju-do	8.87	0.177	9.03	10.09	10.47	10.98	11.59
17	Sejong	6.88	0.426	12.89	8.59	7.54	6.97	6.95*
Mean		6.87	0.355	10.15	7.89	7.58	7.70	8.15
Stdev		1.55	0.090	1.93	1.44	1.47	1.75	2.13
V		22.49	25.330	19.04	18.21	19.40	22.78	26.07

Source: Authors' own calculations. * The optimum occurs at δ > 0.5 for Gyeonggi-do, at δ= 0.480 for Ulsan, and at δ = 0.426 for Sejong; note that for Jeju-do, the optimum is lower than δ = 0.2.

However, there are still variations in the optimum δ values. For region 16, the optimum value is lower than δ = 0.2, which is atypically low. Furthermore, the optimum value for region 1 is δ = 0.514 and for region 5 is 0.480.

Table 2 presents the values of δ that yield the lowest MAPE for sectoral multipliers. The values were averaged across all regions to obtain a MAPE ≈ 6.87%. However, Table 2 shows that using a common δ = 0.35 gives a MAPE ≈ 7.58, employing a uniform δ = 0.4 across all regions gives a MAPE ≈ 7.7%, and using a common δ = 0.3 gives a MAPE ≈ 7.89. Thus, the difference between the common δ = 0.4 and the minimum MAPE for the sectoral multipliers is, on average, approximately 0.83 percentage points.

4. The sector-specific approach (SFLQ)

The primary aim of our study is to test the accuracy of the regionalization of national input-output tables produced by a novel sector-specific FLQ equation, the SFLQ, posited by Kowalewski (2015). Kowalewski divides SFLQ analysis into three steps:

1. the optimum values of δ for each sector are identified,
2. the SFLQ performance/accuracy is tested, and
3. whether industries with low or high δ_j have common characteristics is checked.

The SFLQ exploits the regression model (9) to calculate the industry-specific values of δ unique to each region. We computed SLQ_j using sectoral employment data (Table A.1). Although CL_j, IM_j, and VA_j do not vary across regions (Table A.2), Kowalewski's approach necessitates a distinct regression for each region owing to variations in the SLQ_j values across regions. Kowalewski (2015) gives the rationale for these applied regressors.

CL is expected to have a positive coefficient because it describes industrial concentration, i.e., how different industries are distributed across regions (Kowalewski, 2015). Thus, a higher concentration is associated with a higher import propensity regionally. Furthermore, industries with higher concentrations will require a higher δ.

Kowalewski (2015) did not defend the inclusion of VA in the regression model. However, it should be a meaningful regressor because a larger percentage of value added in the entire national output may indicate a smaller proportion of intermediate inputs, leading to decreased imports.

SLQ_j is expected to have a negative coefficient. Kowalewski (2015) argues that higher regional specialization (SLQ_j) corresponds to a decrease in the value of δ_j, which will additionally lower the share of imports and increase intraregional trade.

Let us first examine Kowalewski's results. The regression results based on Kowalewski's model (9) for all regions are presented in Appendices 3–7. Overall, the results for different regions appear sensible. On average, around 45% of the variation in δ_j is explained by the regression model (9). Furthermore, the regression coefficients of CL, VA, and SLQ_j are mainly of a right sign (Table 3). For IM, eight coefficients are of an expected sign. However, CL exhibits statistical significance for four regions and SLQ for thirteen regions. In sum, the results for both IM and VA cast doubt on their relevance, and they indeed seem to be a redundant variable. Consequently, these regressors may be irrelevant.

Table 3. The anticipated sign of variation of each regression coefficient, the number of times it was significant, and R squares. Kowalewski's regression Equation (9).

	Region	CL	SLQ	IM	VA	R^2
1	Gyeonggi-do	+	−	−	+	0.310
2	Seoul	+	−	+	+	0.589
3	Gyeongsangbuk-do	−	−	+	−	0.229
4	Gyeongsangnam-do	+	−	+	−	0.305
5	Ulsan	−	−	+	−	0.073
6	Jeollanam-do	−	−	+	−	0.517
7	Chungcheongnam-do	−	−	+	−	0.293
8	Incheon	+	−	−	−	0.524
9	Busan	+	−	+	+	0.718
10	Chungcheongbuk-do	−	−	+	−	0.250
11	Daegu	+	−	−	−	0.745
12	Jeollabuk-do	+	−	−	−	0.250
13	Gangwon-do	+	−	−	−	0.705
14	Gwangju	+	−	−	−	0.558
15	Daejeon	+	−	−	−	0.679
16	Jeju-do	+	−	+	+	0.755
17	Sejong	−	−	−	−	0.286
	Statistically significant	4	13	0	2	
	Expected sign	Positive	Negative	Negative	Negative	Mean = 0.458

Source: The results are based on calculations performed by the authors.

Appendices 8–16 show the optimum δ_j values (benchmark δ_j values) that result in the lowest MAPE for the multipliers. The optimum δ_j was calculated for each sector using a step size for δ first, and then linear interpolation was applied. For most sectors, estimated delta values align closely with the optimum delta value, suggesting that the estimation method is reasonably accurate. However, certain sectors exhibit significant differences between optimum deltas and estimated deltas (%), such as *others* (220.4%), *transportation* (170.8%), *mined and quarried goods* (173.9%), and *petroleum and coal products* (107.1%). The lowest difference is in *textile and leather products* (3.2%) and in *basic metal products* (3.6%). Certain industries, like *mined and quarried goods,* may dominate in one region but have limited presence in others, leading to over- or under-estimation of the delta. Some sectors with localized production or consumption patterns, such as *textile and leather products* or *construction* may be harder to estimate using non-survey methods, even if SLQ_j is controlled, leading to larger discrepancies between optimum δ_j and estimated δ_j values. The mean of the sectoral difference averages is 30.0%. Regionally, the biggest differences are in Chungcheongbuk-do (141.8%) and in Gwangju (98.4%). The smallest differences are in Jeju-do (0.6%), Daegu (6.4%), and Busan (7.2%).

To analyze the quality of our estimates, we conducted a correlation analysis between $\hat{\delta}_j$ and δ_j. Most of the resulting correlation coefficients are statistically significant at a 5% risk level, which lends support to Kowalewski's approach. For regions 12 and 2, correlations are statistically significant at a 10% risk level. Only for regions 16 and 6 are correlations not statistically significant. However, our analysis still leaves room for improved precision.

Widely examined, we studied the comparative effectiveness of different measures in terms of their MAPE, as presented in Tables 2 and 4. We analyzed the accuracy of SLQ, CILQ, FLQ, and SFLQ (regression 9) regionalization methods to produce regional multipliers compared with the official multipliers computed by the BOK. FLQ figures with $\delta = 0.2$, $\delta = 0.3$, $\delta = 0.35$, $\delta = 0.4$, and $\delta = 0.45$ are shown in Table 2. A selection of MAPE results produced by the SFLQ with those ~~were~~ estimated through regression (9), and those where optimum values have been employed are shown in Table 4 (see also Table A.17). However, researchers who employ non-survey methods are not aware of the optimum δ_j values. Thus, the results of optimum value outcomes serve to demonstrate the most favorable results that could potentially be attained by employing the FLQ and the SFLQ methods under ideal conditions.

Table 4. Estimating output multipliers for different regions of South Korea in 2015 utilizing alternative versions of Kowalewski's regression Equation (MAPE computed from 33 sectors).

	Region	SFLQ optimum delta	SFLQ linear model (9)	SFLQ semi-log model (16)	SFLQ double-log model (17)	SFLQ model (18)
1	Gyeonggi-do	0.920	3.760	4.283	3.863	7.042
2	Seoul	5.907	7.263	6.874	6.741	9.103
3	Gyeongsangbuk-do	0.945	4.930	5.236	4.902	5.375
4	Gyeongsangnam-do	1.032	4.412	4.542	3.337	5.137
5	Ulsan	1.489	6.401	6.666	5.908	7.523
6	Jeollanam-do	2.029	6.005	5.529	4.955	8.484
7	Chungcheongnam-do	1.153	6.958	7.521	6.049	7.724
8	Incheon	0.752	4.017	4.156	3.067	4.788
9	Busan	1.180	4.194	4.519	4.002	6.146
10	Chungcheongbuk-do	1.264	6.294	7.719	7.451	6.507
11	Daegu	2.183	4.831	4.881	4.507	8.015
12	Jeollabuk-do	1.897	5.215	6.219	5.785	7.571
13	Gangwon-do	2.597	5.927	5.673	5.147	8.921
14	Gwangju	1.547	5.728	5.774	5.127	8.616
15	Daejeon	1.572	6.235	6.054	4.868	7.249
16	Jeju-do	5.075	5.648	5.868	7.076	9.906
17	Sejong	1.484	6.725	7.218	6.543	7.181
Mean		1.943	5.561	5.808	5.255	7.370
Stdev		1.311	0.988	1.096	1.255	1.416
V		0.675	0.178	0.189	0.239	0.192

Source: The results are based on calculations performed by the authors.

The ways to refine regression (9) (see Zhao & Choi, 2015) include i) incorporating additional explanatory variables and ii) using nonlinear models. If new explanatory variables are used, the data required must be easily accessible. However, such new variables are difficult to envisage. Regarding refinement ii), Flegg & Tohmo's (2018) results explored alternative nonlinear models (see Table 4):

$$ln\delta_j = a + b_1 CL_j + b_2 SLQ_j + b_3 IM_j + b_4 VA_j + e_j \tag{16}$$

$$ln\delta_j = c + d_1 \ln CL_j + d_2 SLQ_j + d_3 IM_j + d_4 \ln VA_j + f_j \tag{17}$$

Table 4 shows that the linear model is the best for two regions (1 and 12). The semi-log model stands out as the best only for one region (16). The double-log model (17) is the best for large regions (2, 3, and 4) and for the smaller regions (13, 14, 15, and 17). Additionally, for regions 5, 6, 7, 8, 9, 10, and 11, the double-log model produces the best outcome. The differences in performance of the linear and semi-log models seem quite low.

Flegg & Tohmo (2019) modified Kowalewski's regression model (9) by setting a constraint of $\beta_2 = 0$. Furthermore, the dependent variable is re-expressed as the mean value of δ_j across all regions. The exclusion of SLQ_j can be justified by the fact that it was the only region-specific variable. Applying the revised model to 33 sectors and 17 regions in 2015 yielded the following outcome:

$$\delta_j = 0.346 + 0.298CL_j – 0.066IM_j – 0.138VA_j + e_j \tag{18}$$

where the residual term is e_j, CL_j, VA_j, and IM_j are not statistically significant. The R^2 was 0.166, which can be regarded as quite low. Flegg & Tohmo (2019) argued that a low R^2 can be attributed to the absence of relevant explanatory variables, as well as random variation in δ_j values. The results are shown in Table 4. The MAPE shows that SFLQ (regression 18) overestimates the sectoral multipliers by an average of 7.4% across the 17 regions. The largest overestimations are found in regions 16 (9.9%), 2 (9.10%) and 13 (8.92%).

4.1. Testing if the optimum δ_j is the mean value of δ_j across all regions

So far, our dependent variable (optimum δ_j) has been a regionally based δ_j. Next, we test whether our results explored alternative nonlinear models (19, 20, and 21) where the optimum δ_j is the mean value of δ_j across all sectors. To derive estimates for the values of δ_j, we define the following regression models:

$$\delta_j = \alpha + \beta_1 CL_j + \beta_2 SLQ_j + \beta_3 IM_j + \beta_4 VA_j + \varepsilon_j \tag{19}$$

$$ln\delta_j = \alpha + \beta_1 CL_j + \beta_2 SLQ_j + \beta_3 IM_j + \beta_4 VA_j + \varepsilon_j \tag{20}$$

$$ln\delta_j = \alpha + \beta_1 \ln CL_j + \beta_2 \ln SLQ_j + \beta_3 \ln IM_j + \beta_4 \ln VA_j + \varepsilon_j \tag{21}$$

Table 5 shows that the double-log model stands out as the best for 15 regions, while the semi-log model is best for the other two (11 and 12).

Table 5. Estimating output multipliers for different regions of South Korea in 2015 utilizing alternative versions of Kowalewski's regression equation (MAPE computed from 33 sectors). The dependent variable optimum δ_j is the mean value of δ_j across all regions.

	Region	SFLQ average Optimum delta	SFLQ linear model (19)	SFLQ semi-log model (20)	SFLQ double-log model (21)
1	Gyeonggi-do	5.897	7.024	7.495	6.471
2	Seoul	8.537	8.937	8.910	8.613
3	Gyeongsangbuk-do	4.298	5.796	6.196	5.263
4	Gyeongsangnam-do	4.172	5.555	5.448	5.043
5	Ulsan	6.738	7.504	7.774	6.847
6	Jeollanam-do	7.168	7.848	7.898	7.226
7	Chungcheongnam-do	5.972	7.800	8.242	7.094
8	Incheon	3.782	4.828	5.466	4.154
9	Busan	5.409	6.181	6.008	5.831
10	Chungcheongbuk-do	4.564	6.631	6.920	5.751
11	Daegu	7.699	8.197	7.559	7.740
12	Jeollabuk-do	6.550	7.463	6.902	7.119
13	Gangwon-do	7.038	8.853	8.784	8.035
14	Gwangju	7.856	9.164	8.764	8.608
15	Daejeon	6.074	7.030	7.499	6.151
16	Jeju-do	8.041	9.584	9.656	8.484
17	Sejong	5.556	7.121	7.465	6.393
Mean		6.197	7.383	7.470	6.754
Stdev		1.444	1.332	1.215	1.307
V		0.233	0.180	0.163	0.193

Source: The results are based on calculations performed by the authors.

4.2. Summary of results for ~~of~~ different methods

A comparison of different methods of producing accurate multipliers for South Korean regions is shown in Table 6. The SFLQ double-log model [Equation (17)] produces the best results if accuracy is analyzed regarding the MAPE. The MAPE shows that the SFLQ overestimates the sectoral multipliers by an average of 5.25% across the 17 regions. However, when we consider the number of parameters estimated for each method, the results are different. We can consider the number of parameters via the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). Burnhan & Anderson (2004) suggests penalizing the number of parameters to avoid model overfitting. Consequently, as Burnham & Anderson (2004) stated, fitting a model with too few parameters wastes information. In contrast, if a model is fitted with too many parameters, the estimates may be too imprecise to be useful.

Table 6. Estimating the output multipliers for 17 regions of South Korea in 2015 via different methods (evaluation using the MAPE).

Method	MAPE	AIC	BIC
SLQ	19.873	−80.311	−80.311
CILQ	20.936	−81.750	−81.750
FLQ (optimum δ), Table 5	6.872	−129.379	−127.883
FLQ (δ = 0.30), Table 5	7.889	−126.507	−125.011
FLQ (δ = 0.325), Table 5	7.676	−126.691	−125.195
FLQ (δ = 0.35), Table 5	7.576	−126.235	−124.738
FLQ (δ = 0.40), Table 5	7.700	−123.736	−122.240
SFLQ (optimum δ_j is regional δ)	1.943	−148.881	−99.497
SFLQ (estimated δ_j), Equation (9)	5.561	−75.773	−26.388
SFLQ (estimated δ_j), Equation (16)	5.808	−76.107	−26.722
SFLQ (estimated δ_j), Equation (17)	5.255	−93.526	−44.141
SFLQ (estimated δ_j), Equation (18)	7.370	−62.382	−12.997
SFLQ (optimum δ_j is the mean value of δ_j across all regions)	6.197	−86.398	−37.013
SFLQ (estimated δ_j), Equation (19)	7.383	−62.959	−13.574
SFLQ (estimated δ_j), Equation (20)	7.470	−67.336	−17.951
SFLQ (estimated δ_j), Equation (21)	6.754	−81.513	−32.128

Note: Equation 9 is the regression model defined by Kowalewski, Equation 16 is the semi-log model, equation 17 is the double-log model, and equation 18 is the regression model (9) where the dependent variable is re-expressed as the δ_j value of delta across all regions. For alternative nonlinear models (19, 20, and 21) optimum δ_j is the mean value of δ_j across all sectors.

The AIC designates a smaller penalty for extra parameters than does the BIC. The AIC is defined as follows:

$$AIC = n * \ln(\hat{\sigma}^2) + k * 2 \tag{22}$$

where k is the number of parameters, n is the number of observations, and $\hat{\sigma}^2$ is the variance of the estimated sectoral multipliers, i.e., $\hat{\sigma}^2 = (1/n)\sum_j(\widehat{m}_j - m_j)^2$. If k increases, the AIC and BIC increasingly diverge for a given n, which results from the BIC assigning an increasing penalty for extra parameters. Consequently, for $n > 2$, the AIC designates a smaller penalty for extra parameters. The BIC is defined as follows:

$$BIC = n * \ln(\hat{\sigma}^2) + k * \ln(n) \tag{23}$$

If the BIC or AIC is used instead of the MAPE or $\hat{\sigma}^2$ to compare regionalization methods, and if $\hat{\sigma}^2 < 1$, then the optimum value will be negative. This result means that the most negative AIC or BIC values give the smallest overestimates of sectoral multipliers.

The AIC and BIC indicate that the SFLQ double-log model [Equation (17)] is still the best SFLQ model for producing accurate multipliers for South Korean regions. However, Table 6 reveals that the FLQ outperforms the SFLQ if the number of parameters is considered (the AIC and BIC). Consequently, the best accuracy is achieved by FLQ with δ = 0.35. The AIC and BIC results are far better, and even the MAPE results are only 2.32 percentage points worse compared to those from equation (17) (the SFLQ double-log model).

If no other data are available, a regional analyst might prefer to use $\delta = 0.35$ for the FLQ. Our results show that with the SFLQ (double-log model), we can obtain a lower MAPE for all regions instead of only region 10 compared to the FLQ (see Table A.17). The highest gains brought by the SFLQ would be obtained for regions 13 (4.07%), 14 (3.68%), 11 (3.65%), 6 (3.43%), 16 (3.39%), and 1 (3.34%). In contrast, the lowest gains from the SFLQ would be for regions 3 (0.37%) and 17 (1.00%). The average gain brought by the SFLQ would be 2.3 percentage points in South Korean regions.

5 Conclusions

The Bank of Korea has produced official tables for the year 2015 covering all 17 regions in South Korea, and this study employed this rich ~~South Korean~~ dataset. The main aim of our paper was to analyze the ability of Kowalewski's (2015) sector-specific method to produce accurate estimates in regionalizing national input-output tables. Kowalewski's new sector-specific FLQ formula, the SFLQ, no longer assumes that the value of δ remains constant across industries in a given region but, instead, can vary.

We estimated the optimum sectoral δ and $\hat{\delta}_j$ based on Kowalewski's regression model for all regions of South Korea in 2015. To estimate the values of δ_j, a region-specific SLQ_j and three other regressors (CL_j, IM_j, and VA_j), derived from national data (they are the same for all regions), were used. VA_j is the share of value added in total national output, IM_j is the proportion of foreign imports in the overall inputs used across the entire national economy, and CL_j is a coefficient of localization. Although CL_j, IM_j, and VA_j do not vary across regions, Kowalewski's approach requires a separate regression for each region because SLQ_j values vary between regions.

The SFLQ double-log model yields slightly better estimates than the FLQ. However, the SFLQ is a method that includes 33 parameters in estimating every Korean regional input-output table. In contrast, in the FLQ method, there is only one unknown parameter. Arguably, models with more parameters may perform better. Thus, in terms of the BIC and AIC, our paper offers support for the use of the FLQ in regionalizing national input-output tables. However, if researchers have region-specific knowledge and seek to refine the model, they may also use different delta values for some regions. In particular, higher delta values are expected for regions with above-average foreign inputs and for regions with above-average inputs from other regions.

The FLQ with a delta value of 0.35 produces MAPE estimates that differ from the optimum MAPE by 11.9 percentage points. The SFLQ (double-log model) yields MAPE estimates that differ from the optimum MAPE by an average of 240.2%. Consequently, once the need to estimate the values of theδ_j is recognized, the SFLQ yields far less accurate estimates than does the FLQ.

The Bank of Korea does not disclose how the 2015 interregional input-output tables were compiled. However, we believe it is reasonable to assume that the Bank's methodology has not changed significantly over the years; as such, these 2015 IO tables should therefore serve as good benchmark tables.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

We thank the editor and the anonymous referees for their scrutiny of our paper.

Availability of data and materials

The Bank of Korea produced interregional input-output tables for all 17 South Korean regions, with a classification of 78 economic sectors. Economic Statistics System (https://ecos.bok.or.kr/#/Search/input-output).

Conflict of interest

All author declares no conflicts of interest in this paper.

References

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

Burnham KP, Anderson DR (2004) Multimodel inference: Understanding AIC and BIC in model selection. *Sociol Methods Rese* 33: 261–304. https://doi.org/10.1177/0049124104268644

Flegg AT, Tohmo T (2013a) 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 (2013b) A comment on Tobias Kronenberg's 'Construction of regional input–output tables using nonsurvey methods: The role of cross-hauling'. *Int Reg Sci Rev* 36: 235–257. https://doi.org/10.1177/0160017612446371

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 (2019) The regionalization of national input-output tables: A study of South Korean regions. *Pap Reg Sci* 98: 601–621. https://doi.org/10.1111/pirs.12364

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

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, Lamonica GR, Chelli FM, et al. (2021) A new approach to modelling the input–output structure of regional economies using non-survey methods. *J Econ Struct* 10: 1–31. https://doi.org/10.1186/s40008-021-00242-8

Harrigan FJ, McGilvray JW, McNicoll IH (1980) Simulating the structure of a regional economy. *Environ Plan A* 12: 927–936. https://doi.org/10.1068/a120927 https://doi.org/10.1068/a120927

Hewings GJD (1971) Regional input-output models in the U.K.: Some problems and prospects for the use of nonsurvey techniques. *Reg Stud* 5: 11–22. https://doi.org/10.1080/09595237100185021

Isard W (1960) *Methods of regional analysis: An introduction to regional science*. The regional science studies series 4. The M.I.T. Press, Massachusetts Institute of Technology.

Jensen RC (1980) The concept of accuracy in regional input-output models. *Int Reg Sci Rev* 5: 139–154. https://doi.org/10.1177/016001768000500203

Jensen RC, MacDonald S (1982) Technique and technology in regional input-output. *Ann Reg Sci* 16: 27–45. https://doi.org/10.1007/BF01284245

Kowalewski 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

Mardones C, Silva D (2021) Estimation of regional input coefficients and output multipliers for the regions of Chile. *Pap Reg Sci* 100: 875–889. https://doi.org/10.1111/pirs.12603

Mardones C, Silva D (2023) Evaluation of Non-survey Methods for the Construction of Regional Input–Output Matrices When There is Partial Historical Information. *Comput Econ* 61: 1173–1205. https://doi.org/10.1007/s10614-022-10241-x

Mardones C, Correa M (2024) Methodological proposal to approximate the sectoral impacts of a carbon tax at the regional level–the case of Chile. *Econ Syst Res*, 1–24. https://doi.org/10.1080/09535314.2024.2397974

Miernyk WH (1968) Long-Range forecasting with a regional input-output model. *West Econ J* 6: 165–176. https://doi.org/10.1007/978-1-349-15404-3_13

Miernyk WH (1976) Comments on recent developments in regional input-output analysis. *Int Reg Sci Rev* 1: 47–55. https://doi.org/10.1177/016001767600100204

Miller RE, Blair PD (2009) *Input-output analysis: Foundations and extensions,* Cambridge University Press, New York.

Morrison WI, Smith P (1974) Nonsurvey input-output techniques at the small area level: An evaluation. *J Reg Sci* 14: 1–14. https://doi.org/10.1111/j.1467-9787.1974.tb00425.x

OECD (2017) Gross domestic product (GDP) (indicator). https://doi.org/10.1787/dc2f7aec-en

Richardson HW (1972) *Input-output and regional economics*, London, Weidenfeld and Nicolson.

Round JI (1978) An interregional 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

Round JI (1983) Nonsurvey techniques. A critical review of the theory and the evidence. *Int Reg Sci Rev* 8: 189–212. https://doi.org/10.1177/016001768300800302

Schaffer WA, Chu K (1969) Nonsurvey techniques for constructing regional interindustry models. *Pap Reg Sci Assoc* 23: 83–101. https://doi.org/10.1007/BF01941876

Smith P, Morrison WI (1974) *Simulating the urban economy, input-output techniques*.

Theil H, Beerens GAC, DeLeeuw CG, et al. (1966) *Applied Economic Forecasting*, Amsterdam: North-Holland Publishing Company.

Tohmo T (2004) New developments in the use of location quotients to estimate regional input–output coefficients and multipliers. *Reg Stud* 38: 43–54. https://doi.org/10.1080/0034340031000163226

Zhao X, Choi SG (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