Curve shapes and parameters in FLQ regional modelling: some alternative approaches

Anthony T. Flegg*

University of the West of England, Bristol, UK

E-mail: tony.flegg@uwe.ac.uk

Xesús Pereira-López

Universidade de Santiago de Compostela, Spain

E-mail: xesus.pereira@usc.es

Napoleón Sánchez-Chóez

Escuela Politécnica Nacional, Ecuador and Universidade de Santiago de Compostela, Spain

E-mail: napoleon.sanchez@epn.edu.ec

Fernando de la Torre Cuevas

Universidade de Santiago de Compostela, Spain

E-mail: fernando.delatorre@usc.es

Timo Tohmo

University of Jyväskylä, Finland

E-mail: timo.tohmo@jyu.fi

*corresponding author, who acknowledges the award of a travel grant from the Regional Studies Association.

Curve shapes and parameters in FLQ regional modelling: some alternative approaches

ABSTRACT

In this paper, we propose a novel way of implementing the FLQ (Flegg's location quotient) approach to the regionalization of national input–output tables. Although the FLQ formula often yields the most accurate results of the pure LQ-based methods, the need to specify values of the unknown parameter δ in this formula presents an obstacle to its implementation. A possible solution is to use the FLQ+ method, which employs a modified cross-entropy method, along with a regression model, to estimate values of δ specific to both region and country. Here we develop a fresh approach to the use of the FLQ that substantially simplifies its application, while simultaneously enhancing its performance. As before, sectoral outputs (or employment) are the only regional data required. We focus on how regional size, R, is incorporated in the FLQ formula and simplify the way in which R affects the allowance made for imports from other regions. We call this new formula the reformulated FLQ or RFLQ. We also show how the unknown parameter in this formula can be estimated using readily available data. We test our proposal using the 2005 and 2015 Korean survey-based regional IO datasets. We contrast our estimates with survey-based values and compare results with those from several well-known techniques. Furthermore, we examine two different information scenarios: with and without industry-specific information. The results suggest that one can produce regional IO tables with similar or even better accuracy without using unknown parameters.

Keywords: Regional input–output tables; non-survey methods; FLQ; RFLQ

JEL codes: C67, O18, R15

1. INTRODUCTION

1.1. *Location quotients to produce regional input-output models: still a work in progress*

Location quotients (LQs) are still a widely used non-survey technique to regionalize national input–output (IO) tables and to generate interregional IO (IRIO) models. Their principal attraction is the minimal data requirements, namely regional and national output (or employment) by sector. Many different LQ-based formulae have been developed, which are discussed in detail in Flegg et al. (2021) and other several papers. However, we only need to consider some of these methods here. The most basic formula is the *simple* LQ:

$$SLQ_i = \frac{\frac{x_i^r}{x_\bullet^r}}{\frac{x_i^n}{x_\bullet^n}} = \frac{\frac{x_i^r}{x_i^n}}{\frac{x_\bullet^r}{x_\bullet^n}} = \frac{wx_i^r}{wx^r} \tag{1}$$

Here x_i^r and x_i^n are the total output (production) of the *i*th regional and national sector, respectively, while $x_\bullet^r$ and $x_\bullet^n$ are the corresponding regional and national aggregates. wx_i^r represents the weight of regional sector *i* in the national economy, whereas wx^r represents the weight of region *r* in the national economy, i.e., its relative size. SLQ_i measures the degree of specialization of region *r* in sector *i* relative to the nation. The regional input coefficients are derived according to the following rule:

$$\hat{a}_{ij}^r = \begin{cases} a_{ij}^n SLQ_i & if \quad SLQ_i < 1 \\ a_{ij}^n & if \quad SLQ_i \geq 1 \end{cases} \tag{2}$$

where $\hat{a}_{ij}^r$ is the estimated regional input coefficient and a_{ij}^n is the corresponding observed national input coefficient (excluding inputs purchased from abroad).

However, it has long been known that the SLQ tends to understate a region's imports from other regions; this occurs because 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. For a systematic treatment of this issue, see Többen and Kronenberg (2015).

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:

$$CILQ_{ij} = \frac{SLQ_i}{SLQ_j} = \frac{x_i^r / x_i^n}{x_j^r / x_j^n} \tag{3}$$

where the constraints are applied as in (2). 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. What it does not do is to consider the relative size of a region, $x_\bullet^r/x_\bullet^n$, which cancels out in formula (3). By contrast, this ratio remains a component of the SLQ formula (1).

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:

$$RLQ_{ij} = \frac{SLQ_i}{\log_2(1 + SLQ_j)} \tag{4}$$

Nonetheless, Flegg et al. (1995) criticize the SLQ and RLQ on the grounds that both would tend to understate the imports of relatively small regions owing to the way in which the ratio $x^r_\bullet/x^n_\bullet$ is implicitly incorporated in each formula. The FLQ aims to correct this shortcoming.

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. 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. See, for example, the application to Mexican regions by Dávila-Flores (2015).

The FLQ is defined as follows (cf. Flegg and Webber 1997):

$$FLQ_{ij} = \begin{cases} \lambda CILQ_{ij} & \text{for } i \neq j \\ \lambda SLQ_i & \text{for } i = j \end{cases} \tag{5}$$

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

$$\lambda = [\log_2(1 + \frac{x^r_\bullet}{x^n_\bullet})]^\delta \tag{6}$$

Here $0 \leq \delta < 1$ is a parameter that controls the degree of convexity in equation (6). 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 as in equation (2).

A variant of the FLQ is the *augmented* FLQ (AFLQ), which takes regional specialization into account. It is defined as follows (cf. Flegg and Webber 2000):

$$AFLQ_{ij} = FLQ_{ij}[\log_2(1 + SLQ_j)] \tag{7}$$

which is applicable only when $SLQ_j > 1$.

Several case studies, including Flegg and Tohmo (2016), 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). On the other hand, Lamonica and Chelli (2018) find that the FLQ performs better than the SLQ in smaller regions, yet worse in larger regions. The FLQ is strongly criticized by Fujimoto (2019) on both conceptual and empirical grounds. A response to these criticisms is given in Flegg et al. (2021). Finally, we may note that the FLQ's use in a multiregional context is examined by Jahn (2017), Jahn et al. (2020) and Garcia-Hernandez and Brouwer (2021).

Pereira-López et al. (2020) propose a two-dimensional approach (2D-LQ) to estimate domestic coefficients at the sub-territorial level. This technique can be extrapolated to other contexts; for instance, generating intermediate flow matrices. In this approach, estimates of regional coefficients are calculated according to the following rule:

$$\widetilde{\mathbf{A}}^R = \hat{\mathbf{r}}(\alpha)\mathbf{A}^N\hat{\mathbf{s}}(\beta) \tag{8}$$

where **A** is a matrix of intermediate domestic coefficients, and $\hat{\mathbf{r}}(\alpha)$ and $\hat{\mathbf{s}}(\beta)$ are diagonal matrices whose main diagonal elements work as weighting factors. Scalars α and β are the parameters influencing row and column rectifications, respectively. As in the CILQ, RLQ and FLQ, row and column corrections are addressed differently:

$$\mathbf{r}(\alpha) = \begin{cases} (SLQ_i)^{\alpha} & \text{if} \quad SLQ_i \leq 1 \\ \left[\frac{1}{2}\tanh(SLQ_i - 1) + 1\right]^{\alpha} & \text{if} \quad SLQ_i > 1 \end{cases} \tag{9}$$
$$\mathbf{s}(\beta) = (wx_j^r)^{\beta}$$

One of the main novelties introduced by the 2D-LQ formulation is a modified hyperbolic tangent curve to describe the indirect relationships between input coefficients and location economies.

Papers presenting the 2D-LQ and subsequent variants (Sánchez-Chóez et al. 2022) report better results than previous LQ regionalization methods. Such promising results are obtained at the cost of an additional trouble: providing estimates for the α and β parameters that capture the effect of supply-side location economies. Nevertheless, the 2D-LQ's accuracy appears to be less sensitive to variation in α and β than is true for the FLQ and its sole parameter δ (Pereira-López et al. 2021). Martínez-Alpañez et al. (2023) propose a way to estimate α and β via an econometric model. They report satisfactory results for Korean regions, yet their experiments on different Spanish regions fall far behind.

1.2. *Open debates on curve shapes and parameters*

The way we model the indirect relationships between regional size and input coefficients might also be open to discussion. For instance, McCann and Dewhurst (1998) suggested that the FLQ might not necessarily be the best way to capture relationships between regional size and interindustry structures. Figure 1 depicts the relationship between λ, δ and regional size, *R*, which is measured in terms of a region's share of total national output or employment. The graphs are based on a formulation proposed by Flegg and Webber (1997). A key feature of the graphs is that they pass through the points [0,0] and [1,1]. Furthermore, the gradients decline smoothly and continuously as *R* increases.

However, the authors offer no rationale for the shape of the graphs in between [0,0] and [1,1], which is an aspect that is worth reconsidering. In particular, in the smallest regions, a relatively small rise in *R* would result in a large increment in λ. There is no obvious reason why that should be so.

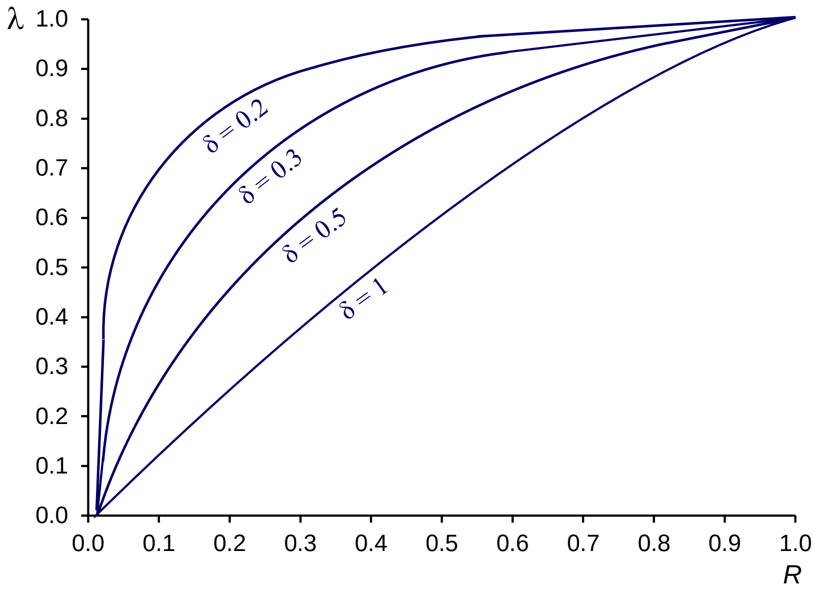

Figure 1. Convexity for different $\boldsymbol{\delta}$ values in the FLQ formula
Source: own elaboration

Perhaps a critical mass of plants and employment is needed before industrial take-off and clustering dynamics appear within a region, which would tend to reduce imports from other regions? The impact of regional heterogeneity should also be considered, especially in the case of big (i.e.: more representative) regions in relatively small nations. In this regard, it might be desirable to have some control over how far the estimated regional input coefficients could diverge from their national counterparts as *R* varies.

According to Jarne et al. (2007), S-shaped curves are capable of describing processes characterized by emergent, inflexion and saturation phases. Such phases can be captured in a logistic function, whose general form is as follows:

$$f(x) = \frac{k}{1 + be^{-ax}}, k, a, b \in \Re \tag{10}$$

For our purposes, a special case of this logistic equation, namely the hyperbolic tangent, is the most appropriate. It can be derived by setting $k = 2$, $a = 2$ and $b = 1$, which gives:

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} \tag{11}$$

Here we use $x - 1$ instead of x, so that:

$$f(x) = \tanh(x - 1) + 1 = \frac{2}{1 + e^{-2(x-1)}} \tag{12}$$

Equation (12) has the following notable properties:

$$\begin{gathered} \lim_{x \to \infty} = \frac{2}{1 + e^{-2(x-1)}} = 2 \\ \lim_{x \to -\infty} \frac{2}{1 + e^{-2(x-1)}} = 0 \end{gathered} \tag{13}$$

Crucially, we have $f(1) = 1$ and $f(0) > 0$.

Figure 2 illustrates the type of logistic function we have in mind. Admittedly, this is one possibility among many. For instance, U-shaped (Duarte et al. 2022; Stöllinger, 2021) and inverted U-shaped curves (Kitsos et al. 2023) are also used to model regional economic dynamics.

Regardless of our choice of curve, the main obstacle in applying the FLQ formula is in determining a value for δ in equation (6). This is crucial because its value might vary across regions, countries and time. This problem has been addressed by applying the FLQ+ approach proposed by Flegg et al. (2021). Another issue is that δ is apt to vary across sectors too (Flegg and Tohmo 2019). The 2D-LQ and non-LQ approaches (e.g.: gravity models) offer no solution to this problem either since they rely on more than one parameter to derive their estimates.

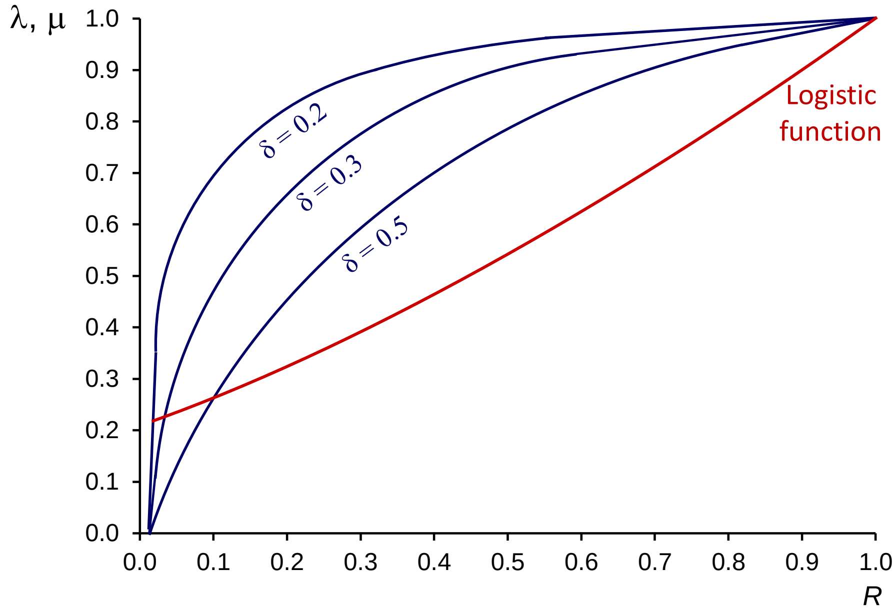

Figure 2. Example of a logistic function (hyperbolic tangent) versus the FLQ for different ***δ*** values
Source: own elaboration

1.3. *The aim of the present paper*

In this paper, we explore an alternative treatment of the indirect relationships between regional attributes (e.g.: regional size) and economic integration across industries. We use a particular form of the logistic function to model such indirect linkages, in the vein of some literature on innovation and technological change. We posit a relationship starting at a non-zero parameter, with a gradient that rises gently as regional supply self-sufficiency increases. Our hypothesis is that, when compared with Figure 1, the logistic curve depicted in Figure 2 can more successfully capture the initial stages, take-off and maturing dynamics of a given industry across both time and space.

We also provide an example of how our method could be implemented without involving parameter estimations. We evaluate if our new method works well with non-parametric rectifications. To this end, we choose a proxy for regional supply self-sufficiency and introduce it directly into our calculations. Moreover, we study the extent to which introducing industry-specific information improves the accuracy of our estimates.

The remainder of this paper is organized as follows. In section 2, we provide a detailed description of our methodological proposal. We first revise Flegg's original method to introduce an alternative S-shaped hyperbolic tangent treatment in its CILQ part. We next present a non-parametric alternative rectification of the CILQ estimates. Section 3 illustrates our findings, with an empirical application using the Korean 2005 and 2015 IRIO models. Finally, section 4 concludes with some final remarks and suggestions for future research.

2. METHODOLOGICAL PROPOSAL

We start by considering a transformation of the CILQ formula using a sigmoidal function. Thus, we account for supply/demand relative sizes. Following the FLQ approach, we retain a rectification parameter μ related to a region's degree of supply self-sufficiency but, unlike the CILQ, we prevent excessive rectifications. We call this formula the *reformulated* FLQ or RFLQ, which is specified as follows:

$$RFLQ_{ij} = \mu\left(\tanh\left(CILQ_{ij} - 1\right) + 1\right) \tag{14}$$

In this equation, we still rely on a parameter μ. Our aim is to capture the effect of using an alternatively shaped curve. In doing so, we employ the same hyperbolic tangent modification, as explained in equations (10) to (13) and depicted in Figure 2. Furthermore, we explore if the value of μ can be determined a priori according to available or estimated data. We apply the same smoothing to our proxy data as we do to the CILQ. Thereafter, we employ equation (14).

Now let **Z** stand for a matrix of intermediate flows. Our first non-parametric scenario, NP1, can be formalized as follows:

$$NP1_{ij} = \bar{\mu}^{r}\left(\tanh\left(CILQ_{ij} - 1\right) + 1\right) \tag{15}$$

where:

$$\bar{\mu}^{r} = \tanh\left(\frac{z_{\bullet\bullet}^{rr}}{z_{\bullet\bullet}^{\bullet r}} - 1\right) + 1 \quad r \subset n \tag{16}$$

In equation (16), a dot (•) stands for summation across a given dimension: in superscripts, denoting origin or destination; in subscripts, denoting supplying or demanding industries. Therefore, $z_{\bullet\bullet}^{rr}$ stands for domestically (i.e. intraregionally) supplied intermediates and $z_{\bullet\bullet}^{\bullet r}$ for nationally supplied intermediates (domestic + interregional imports). $\bar{\mu}^{r}$ denotes our non-parametric estimate of μ for region r. Even though such data might not be available in all countries and regions, literature suggests several ways to estimate these figures. See Thissen et al. (2018) for a recent application to build the EUREGIO dataset.

Finally, we conceive a more information-demanding scenario, NP2, with industry-specific rectifications, as in Zhao and Choi (2015). We proceed as in NP1. Formally:

$$NP2_{ij} = \bar{\mu}_j^r\big(\tanh\big(CILQ_{ij} - 1\big) + 1\big) \tag{17}$$

where:

$$\bar{\mu}^r = \tanh\left(\frac{z_{\bullet j}^{rr}}{z_{\bullet j}^{\bullet r}} - 1\right) + 1 \quad r \subset n \tag{18}$$

In equation (18), $z_{\bullet j}^{rr}$ stands for domestically (i.e. intraregionally) supplied intermediates by industry *j* and $z_{\bullet j}^{\bullet r}$ for nationally supplied intermediates (domestic + interregional imports) by this industry. Because of the amount of data required, this equation only provides an ideal scenario, which we use for counterfactual comparisons.

3. EMPIRICAL APPLICATION

3.1. *Methods and data*

We now present an empirical application to illustrate our approach. We aggregate the Korean IRIO models for 2005 and 2015 up to 31 sectors. Despite possible biases induced by aggregation (Lahr and Stevens 2002), that is the only way we could find to establish straightforward comparisons across time. We subsequently aggregate all trade flows into national IO models. Our aim is to extract a regional model for each region from the national matrices using our methodology.

Our detailed results for the two years are presented in Tables 1 and 2, which display the values of the following statistic for each region and measure:

$$STPE = 100\,\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\tilde{a}_{ij}^r - a_{ij}^r\right|}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^r} \tag{19}$$

where *STPE* is the standardized total percentage error, $\tilde{a}_{ij}^r$ stands for the estimated regional input coefficients and a_{ij}^r for their true counterparts. For the four parametric methods (FLQ, AFLQ, 2D-LQ and RFLQ), we take the parameter value that minimizes the *STPE* for a sequence from 0 to 1, with steps of 0.01.

3.2. *Results*

It is helpful to start by considering the rankings of the parametric methods in terms of the *STPE*, which are shown in Tables 4 and 5 in the appendix. A key finding is the excellent performance of the 2D-LQ approach. It gives the best results for 13 of the 16 regions in 2005 and takes second place in the others. This outstanding performance is mirrored in 2015. By contrast, the

CILQ is clearly the worst approach in both years. We may note too that the FLQ outperforms the AFLQ in 11 out of 16 cases in 2005 and in 13 out of 17 cases in 2015.

The comparative outcomes for the FLQ and RFLQ are also very interesting. Here the RFLQ gives better results than the FLQ for 13 of the 16 regions in 2005. The only exceptions are Seoul, Busan and Gangwon. In 2015, the RFLQ outperforms the FLQ in all regions apart from Seoul, where its performance is very poor. Even so, in terms of rankings, it is evident that the RFLQ is a big improvement on the FLQ. Taking the rankings as a whole, what is striking about the outcomes in the two years is how similar they are, notwithstanding the gap of ten years.

Table 3 provides some useful extra information regarding the parametric methods. One can see that the mean values of α, β and δ are very similar in the two years, while there is a more noticeable difference in the mean values of μ. In terms of dispersion, this is clearly much greater for α than it is for β. It is worth noting here that α and β are the influential parameters in the correction by rows and columns, respectively. It is also evident that α fluctuates substantially more than does δ. Finally, it is interesting to see that parameter μ for the RFLQ exhibits substantially less dispersion than does δ for the FLQ.

The optimal values of μ, shown in parentheses in Tables 1 and 2, are also worth examining. For the initial year 2005, we can discern a tendency for μ to decline as regions get smaller: for the four largest regions, the mean value of μ is 0.530; for the next seven regions, it is 0.394; while, for the four smallest regions, it is 0.298. By contrast, the values of μ for the year 2015 do not vary so greatly across regions: the mean is 0.553 for the four largest regions, whereas it is 0.464 for the next seven and 0.460 for the five smallest regions. We have ignored the new and very small region of Sejong in this comparison. Figures 3 and 4 in the appendix illustrate how sensitive STPE values are to the parameter choice in FLQ, AFLQ and RFLQ.

Notwithstanding the impressive results for the 2D-LQ revealed in Tables 1 and 2, there is a potential problem in the application of this method: it involves two unknown parameters, α and β, whose optimal values vary noticeably across regions, especially so in the case of α.[1] By contrast, the FLQ only involves one unknown parameter. Whilst it is true that the values of δ also vary noticeably across regions, this problem can be overcome by using the FLQ+ method.

[1] Based on an analysis of Eurostat data for six European countries in 2005, Pereira-López et al. (2020, table 5) provide some reassurance that the range of suitable values of α and β 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. However, these values are not consistent with those reported in our Tables 1 and 2.

Tables 1 and 2. STPE results for alternative regionalization techniques, years 2005 and 2015. Source: own elaboration.

	Region	**Regional size (%)**	**CILQ**	**FLQ** (δ)	**AFLQ** (δ)	**2D-LQ** $(\alpha;\ \beta)$	**RFLQ** (μ)	**NP1**	**NP2**
1	Gyeonggi	20.14	69.90	41.56	42.73	37.49	39.16	39.32	35.50
				(0.56)	(0.62)	(0.68; 0.42)	(0.55)		
2	Seoul	18.20	54.36	51.11	57.42	41.80	68.94	68.96	58.07
				(0.19)	(0.71)	(0.72; 0.25)	(0.58)		
3	North Gyeongsang	8.42	82.39	61.06	67.54	53.64	54.43	54.53	49.47
				(0.69)	(0.81)	(0.32; 0.29)	(0.49)		
4	South Gyeongsang	7.35	73.56	55.56	60.53	52.45	51.66	51.67	47.28
				(0.45)	(0.53)	(0.52; 0.30)	(0.50)		
5	Ulsan	7.11	106.77	79.79	83.60	64.28	61.43	64.34	57.01
				(0.56)	(0.89)	(0.16; 0.28)	(0.37)		
6	South Jeolla	6.46	84.98	75.68	75.22	53.80	61.93	69.88	54.24
				(0.38)	(0.95)	(0.40; 0.22)	(0.40)		
7	South Chungcheong	6.33	112.65	69.24	66.41	53.05	53.14	55.41	49.80
				(0.64)	(0.67)	(0.16; 0.32)	(0.39)		
8	Incheon	5.49	116.02	57.61	58.69	50.30	46.68	48.00	46.02
				(0.54)	(0.70)	(0.32; 0.36)	(0.40)		
9	Busan	5.06	85.35	52.99	52.59	49.70	54.66	55.32	49.07
				(0.40)	(0.44)	(1.04; 0.23)	(0.42)		
10	North Chungcheong	2.92	100.61	68.98	70.35	61.99	62.13	63.06	56.39
				(0.50)	(0.52)	(0.40; 0.32)	(0.36)		
11	Daegu	2.88	84.62	57.78	58.34	50.50	55.23	55.45	49.96
				(0.30)	(0.34)	(0.88; 0.25)	(0.42)		
12	North Jeolla	2.74	107.54	69.88	70.15	58.24	62.49	65.41	58.57
				(0.46)	(0.54)	(0.52; 0.29)	(0.32)		
13	Gangwon	2.16	97.97	71.99	72.15	58.16	73.62	76.23	67.83
				(0.18)	(0.33)	(0.64; 0.15)	(0.38)		
14	Gwangju	2.15	112.48	72.81	72.55	60.65	65.47	68.35	61.63
				(0.37)	(0.44)	(0.64; 0.28)	(0.30)		
15	Daejeon	1.93	152.24	87.61	87.27	74.61	74.96	86.06	76.18
				(0.67)	(0.75)	(0.84; 0.36)	(0.19)		
16	Jeju	0.67	94.19	83.92	86.23	53.10	79.24	84.20	72.86
				(0.20)	(0.33)	(0.96; 0.13)	(0.30)		

	Region	Regional size (%)	CILQ	FLQ (δ)	AFLQ (δ)	2D-LQ (α; β)	RFLQ (μ)	NP1	NP2
1	Gyeonggi	22.85	61.53	39.41	40.44	34.23	33.25	33.28	31.04
				(0.55)	(0.58)	(0.44; 0.42)	(0.55)		
2	Seoul	18.97	56.93	56.59	61.66	41.99	70.25	70.37	60.95
				(0.16)	(0.58)	(0.76; 0.17)	(0.62)		
3	North Gyeongsang	7.00	76.31	55.67	61.52	44.65	45.46	45.57	41.89
				(0.35)	(0.41)	(0.28; 0.29)	(0.48)		
4	South Gyeongsang	6.93	67.13	55.82	63.74	46.83	45.19	45.61	41.59
				(0.27)	(0.52)	(0.36; 0.23)	(0.56)		
5	Ulsan	6.32	99.33	73.55	75.35	56.51	52.68	53.37	49.67
				(0.65)	(0.73)	(0.04; 0.30)	(0.43)		
6	South Jeolla	4.89	75.21	65.82	72.03	46.69	54.22	55.16	46.53
				(0.29)	(0.30)	(0.52; 0.19)	(0.49)		
7	South Chungcheong	6.96	94.44	67.93	68.15	58.66	57.42	58.58	53.88
				(0.64)	(0.69)	(0.08; 0.37)	(0.41)		
8	Incheon	4.96	100.01	51.33	52.64	48.56	49.46	49.74	46.41
				(0.48)	(0.55)	(0.80; 0.26)	(0.42)		
9	Busan	4.73	68.68	49.95	49.93	42.74	46.64	47.37	41.78
				(0.28)	(0.33)	(0.92; 0.20)	(0.58)		
10	North Chungcheong	3.47	100.86	72.38	73.68	62.01	64.18	64.39	58.28
				(0.48)	(0.62)	(0.12; 0.34)	(0.41)		
11	Daegu	2.82	77.21	58.77	58.16	44.53	51.39	51.71	46.50
				(0.29)	(0.31)	(1.04; 0.19)	(0.51)		
12	North Jeolla	2.82	82.42	62.71	64.11	50.30	55.35	55.35	49.04
				(0.33)	(0.33)	(0.64; 0.19)	(0.49)		
13	Gangwon	1.97	88.31	67.63	64.74	52.46	67.15	67.17	59.90
				(0.40)	(0.42)	(0.88; 0.19)	(0.46)		
14	Gwangju	2.07	91.12	69.87	72.12	52.27	59.53	59.56	53.60
				(0.35)	(0.41)	(1.08; 0.18)	(0.46)		
15	Daejeon	1.92	124.34	79.82	78.86	59.84	64.35	64.38	58.37
				(0.44)	(0.47)	(1.04; 0.25)	(0.41)		
16	Jeju	0.81	77.16	71.21	71.37	48.13	69.39	69.39	60.85
				(0.19)	(0.27)	(0.68; 0.13)	(0.48)		
17	Sejong	0.50	183.03	87.88	88.30	72.63	76.65	87.41	78.75
				(0.59)	(0.60)	(0.60; 0.29)	(0.19)		

Table 3. Means and coefficients of variation of parameters

Formula	FLQ	2D-LQ		RFLQ
Parameter	δ	α	β	μ
Mean 2005	0.44	0.58	0.28	0.40
Mean 2015	0.40	0.60	0.25	0.47
V (%) 2005	38	47	27	25
V (%) 2015	38	57	32	20

Note: V = coefficient of variation. Source: own elaboration.

Turning now to the choice between 2D-LQ and RFLQ, a considerable advantage of using the RFLQ rather than the 2D-LQ is that one can avoid the complex task of finding suitable values for α and β. Like the FLQ, the RFLQ has only one unknown parameter μ and it is important to consider how one might obtain a value in a practical application. Here our non-parametric scenario NP1, which is much less information-demanding than the alternative scenario NP2, can be of some assistance.

In fact, Table 2 reveals that the outcomes for NP1 in 2015 very closely match those for the RFLQ, with a negligible difference in the *STPE* in all regions except for the smallest, Sejong. This is a very important finding, as it offers a way of obtaining good estimates of μ for use in the RFLQ method. It should be recalled that the RFLQ results displayed in Table 2 were computed using the optimal values of μ obtained from a full set of survey-based data. Such data would rarely be available in a practical application. Looking now at the outcomes for 2005 in Table 1, it is evident that there is a negligible difference in the outcomes for NP1 and RFLQ in half of the regions. Of the other regions, South Joella and Daejeon stand out as having fairly large differences in the outcomes. There is no obvious reason why NP1 should perform so well in 2015 yet produce rather more mixed results in 2005.

Furthermore, Figures 5 and 6 in the appendix show how NP1 outperforms the 2D-LQ for a majority of α, β combinations for 13 regions in 2015 and 11 in 2005. These results also suggest that the 2D-LQ would generally be preferable in small regions and in Seoul. Nevertheless, the problem of estimating α and β in such regions would remain.

As expected, the NP2 scenario yields a substantially more accurate set of results than does the NP1 scenario. However, the sector-specific data underlying NP2 would not normally be available to analysts and the results are presented merely for illustrative purposes.

Finally, it is worth remarking that our results for the RFLQ and NP1 are very encouraging in the sense that they demonstrate that, with reasonable information requirements, we can gain estimates almost as good as those from methods often cited in the literature. What is more, this specification does not rely on any unknown parameter.

4. CONCLUDING REMARKS

In this paper, we have addressed two open and relevant improvement opportunities related to the LQ regionalization literature. First, we have explored if alternative functional shapes can be used to model indirect relationships between regional size, regional self-sufficiency and regional interindustry linkages. Secondly, we have assessed if our alternative functional shapes allow for the substitution of unknown parameters with proxies set a priori.

We draw two tentative conclusions from our empirical work. These should be taken with the caveat that they might be biased owing to the use of peculiar data. First, methods using S-shaped functions appear to be well suited to capture the relations between regional size, regional self-sufficiency and regional technologies. We find that our proposed RFLQ (reformulated FLQ) approach outperforms more conventional methods for most regions in both the 2005 and 2015 Korean datasets. Secondly, our results suggest that combining regional supply self-sufficiency measures with our formula might be a good way to obtain parameters without seriously compromising accuracy in the estimated coefficients.

Various research avenues might be considered to continue the work presented here. Most importantly, a stronger rationale connecting regional characteristics with regional technological structures is needed. Isard's (1960) first channel of synthesis – connecting regional IO with location theory – should be revisited and extended if possible. In this vein, perhaps it would be interesting to evaluate more distinctive functions for modelling regional disparities. Secondly, it might be fruitful to study how sensitive each method is to changes in parameter values. This could allow for more informed choices between parametric and non-parametric techniques. Finally, the combination of our approach with econometric or machine-learning-based estimates (Pakizeh and Kashani 2022) could be considered in the future.

REFERENCES

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

Dávila-Flores, A. (2015) Modelos interregionales de insumo producto de la economía Mexicana. Saltillo, Mexico, Universidad Autónoma de Coahuila.

Duarte, R., Espinosa-Gracia, A., Jiménez, S., and Sánchez-Chóliz, J. (2022). New insights on the relationship between the involvement of countries in global value chains, and intra- and inter-country inequalities. Structural Change and Economic Dynamics, 63:320–329. https://doi.org/10.1016/j.strueco.2022.11.001

Flegg, A. T., Lamonica, G. R., Chelli, F. M., Recchioni, M. C., and Tohmo, T. (2021). A new approach to modelling the input–output structure of regional economies using non-survey methods. Journal of Economic Structures, 10:12. https://doi.org/10.1186/s40008-021-00242-8

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

Flegg, A. T., Webber, C. D., and Elliott, M. V. (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

Fujimoto, T. (2019). Appropriate assumption on cross-hauling national input–output table regionalization. Spatial Economic Analysis, 14:106–128. https://doi.org/10.1080/17421772.2018.1506151

Garcia-Hernandez, J. A., and Brouwer, R. (2021). A multiregional input–output optimization model to assess impacts of water supply disruptions under climate change on the Great Lakes economy. Economic Systems Research, 33: 509–535. https://doi.org/10.1080/09535314.2020.1805414

Isard, W. (1960). Methods of regional analysis: an introduction to regional science. MIT Press, Cambridge MA.

Jahn, M. (2017). Extending the FLQ formula: a location quotient-based interregional input–output framework. Regional Studies, 51:1518–1529. https://doi.org/10.1080/00343404.2016.1198471

Jahn, M., Flegg, A. T., and Tohmo, T. (2020). Testing and implementing a new approach to estimating interregional output multipliers using input–output data for South Korean regions. Spatial Economic Analysis, 15:165–185. https://doi.org/10.1080/17421772.2020.1720918

Jarne, G., Sanchez-Choliz, J., and Fatas-Villafranca, F. (2007). "S-shaped" curves in economic growth. A theoretical contribution and an application. Evolutionary and Institutional Economics Review, 3:239–259. https://doi.org/10.14441/eier.3.239

Kitsos, T., Grabner, S. M., and Carrascal-Incera, A. (2023). Industrial embeddedness and regional economic resistance in Europe. Economic Geography, 90:1–26. https://doi.org/10.1080/00130095.2023.2174514

Lahr, M. L., and Stevens, B. H. (2002). A study of the role of regionalization in the generation of aggregation error in regional input-output models. Journal of Regional Science, 42:477–507. https://doi.org/https://doi.org/10.1111/1467-9787.00268

Lamonica, G. R., and Chelli, F. M. (2018). The performance of non-survey techniques for constructing sub-territorial input-output tables. Papers in Regional Science, 97:1169–1202. https://doi.org/10.1111/pirs.12297

Martínez-Alpañez, R., Buendía-Azorín, J. D., and Sánchez-de-la-Vega, M. del M. (2023). A new improvement proposal to estimate regional input–output structure using the 2D-LQ approach. Economies, 11:20. https://doi.org/10.3390/economies11010020

McCann, P., and Dewhurst, J. H. LL. (1998). Regional size, industrial location and input-output expenditure coefficients. Regional Studies, 32:435–444. https://doi.org/10.1080/00343409850116835

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

Pereira López, X., Carrascal Incera, A., and Fernández-Fernández, M. (2020). A bidimensional reformulation of location quotients for generating input–output tables. Spatial Economic Analysis, 15:476–493. https://doi.org/10.1080/17421772.2020.1729996

Pereira-López, X., Sánchez-Chóez, N. G., and Fernández-Fernández, M. (2021). Performance of bidimensional location quotients for constructing input–output tables. Journal of Economic Structures, 10:7. https://doi.org/10.1186/s40008-021-00237-5

Round, J. I. (1978). An interregional input-output approach to the evaluation of nonsurvery methods. Journal of Regional Science, 18:179–194. https://doi.org/10.1111/j.1467-9787.1978.tb00540.x

Sánchez-Chóez, N. G., Pereira-López, X., and Fernández-Fernández, M. (2022). A modification of the cross-industry location quotient for projecting sub-territorial input-output tables. Revista de Economía Mundial, 2022:25–50. https://doi.org/10.33776/rem.v0i62.5130

Stevens, B. H., Treyz, G. I., and Lahr, M. L. (1989). On the comparative accuracy of RPC estimating techniques. In R. E. Miller, K. R. Polenske, and A. Z. Rose (Eds.), Frontiers of Input-Output Analysis (pp. 245–257). Oxford University Press, Oxford.

Stöllinger, R. (2021). Testing the smile curve: functional specialisation and value creation in gvcs. Structural Change and Economic Dynamics, 56:93–116. https://doi.org/10.1016/j.strueco.2020.10.002

Thissen, M., Lankhuizen, M., van Oort, F., Los, B., and Diodato, D. (2018). EUREGIO: The construction of a global IO DATABASE with regional detail for Europe for 2000-2010 (No. 2018-084/VI; Tinbergen Institute Discussion Paper). http://www.tinbergen.nl

Többen, J., and Kronenberg, T. H. (2015). Construction of multi-regional input–output tables using the charm method. Economic Systems Research, 27:487–507. https://doi.org/10.1080/09535314.2015.1091765

Zhao, X., and Choi, S.-G. (2015). On the regionalization of input–output tables with an industry-specific location quotient. The Annals of Regional Science, 54:901–926. https://doi.org/10.1007/s00168-015-0693-x

APPENDIX

Table 4. Ranking of alternative parametric regionalization techniques: year 2005. Source: table 1.

	Region	Regional size (%)	CILQ	FLQ	AFLQ	2D-LQ	RFLQ
1	Gyeonggi	20.14	5	3	4	**1**	2
2	Seoul	18.20	3	2	4	**1**	5
3	North Gyeongsang	8.42	5	3	4	**1**	2
4	South Gyeongsang	7.35	5	3	4	2	**1**
5	Ulsan	7.11	5	3	4	2	**1**
6	South Jeolla	6.46	5	4	3	**1**	2
7	South Chungcheong	6.33	5	4	3	**1**	2
8	Incheon	5.49	5	3	4	2	**1**
9	Busan	5.06	5	3	2	**1**	4
10	North Chungcheong	2.92	5	3	4	**1**	2
11	Daegu	2.88	5	3	4	**1**	2
12	North Jeolla	2.74	5	3	4	**1**	2
13	Gangwon	2.16	5	2	3	**1**	4
14	Gwangju	2.15	5	4	3	**1**	2
15	Daejeon	1.93	5	4	3	**1**	2
16	Jeju	0.67	5	3	4	**1**	2

Table 5. Ranking of alternative parametric regionalization techniques: year 2015. Source: table 2.

	Region	**Regional size (%)**	**CILQ**	**FLQ**	**AFLQ**	**2D-LQ**	**RFLQ**
1	Gyeonggi	22.85	5	3	4	2	**1**
2	Seoul	18.97	3	2	4	**1**	5
3	North Gyeongsang	7.00	5	3	4	**1**	2
4	South Gyeongsang	6.93	5	3	4	2	**1**
5	Ulsan	6.32	5	3	4	2	**1**
6	South Jeolla	4.89	5	3	4	**1**	2
7	South Chungcheong	6.96	5	3	4	2	**1**
8	Incheon	4.96	5	3	4	**1**	2
9	Busan	4.73	5	4	3	**1**	2
10	North Chungcheong	3.47	5	3	4	**1**	2
11	Daegu	2.82	5	4	3	**1**	2
12	North Jeolla	2.82	5	3	4	**1**	2
13	Gangwon	1.97	5	4	2	**1**	3
14	Gwangju	2.07	5	3	4	**1**	2
15	Daejeon	1.92	5	4	3	**1**	2
16	Jeju	0.81	5	3	4	**1**	2
17	Sejong	0.50	5	3	4	**1**	2

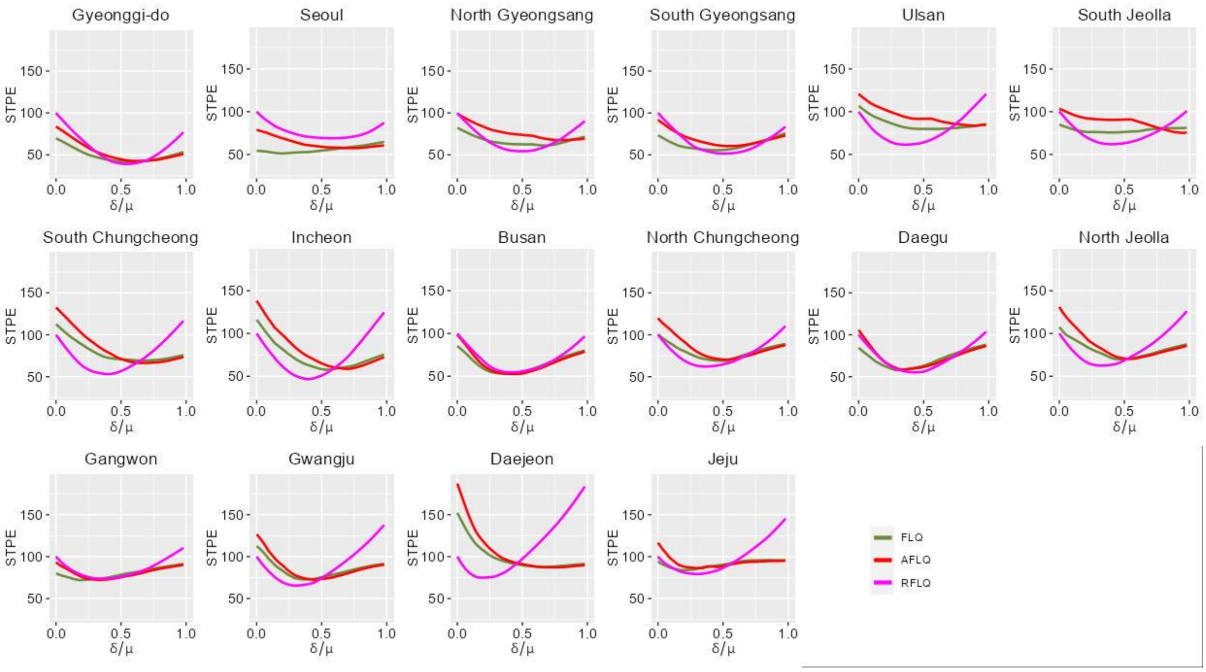

Figure 3. STPE sensitivity to values of parameters for Korean regions: year 2005
Source: own elaboration

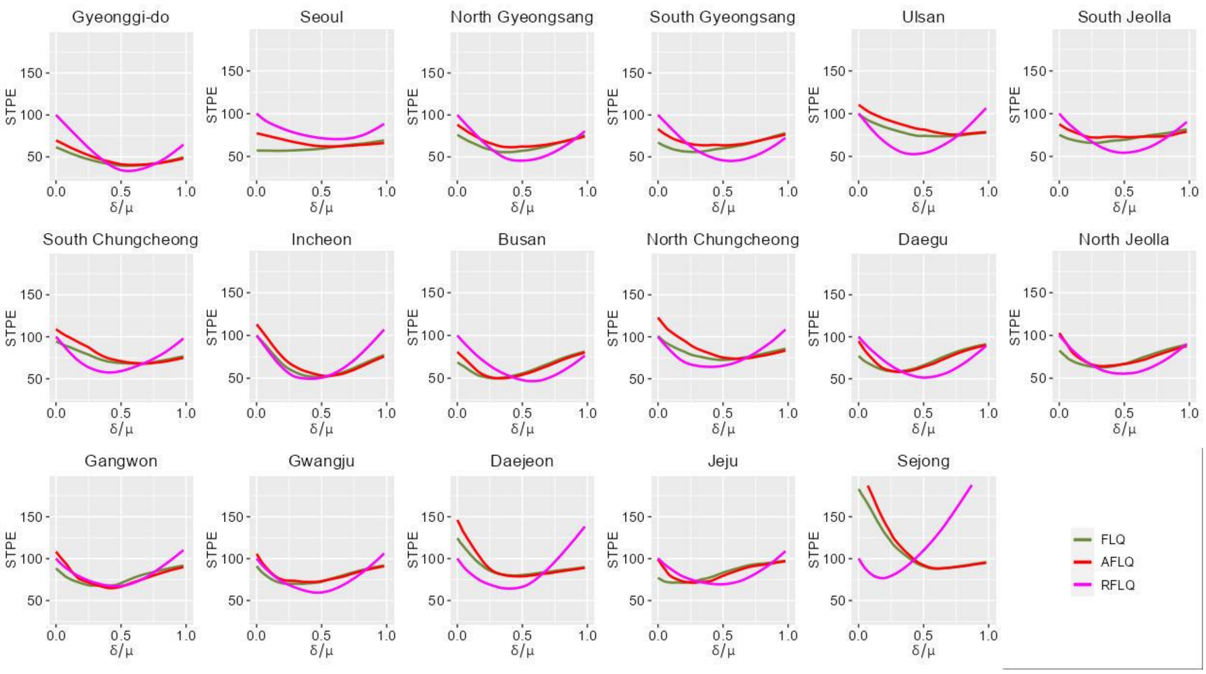

Figure 4. STPE sensitivity to values of parameters for Korean regions: year 2015
Source: own elaboration

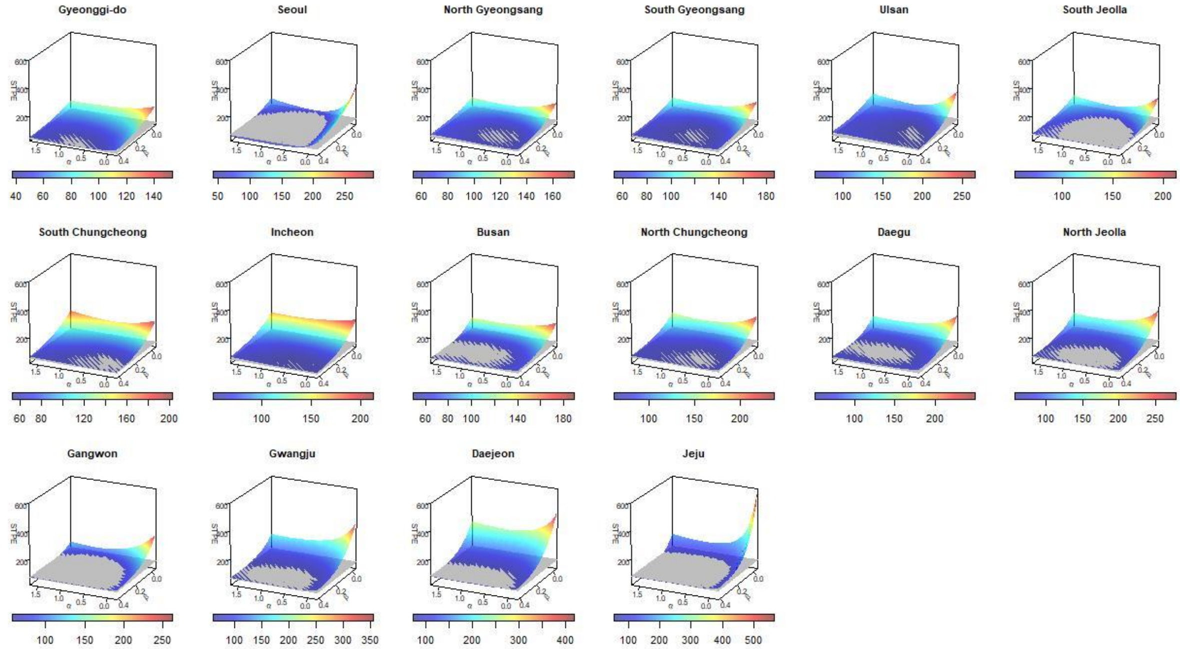

Figure 5. Comparison between STPE values for NP1 (grey) and 2D-LQ: year 2005
Source: own elaboration

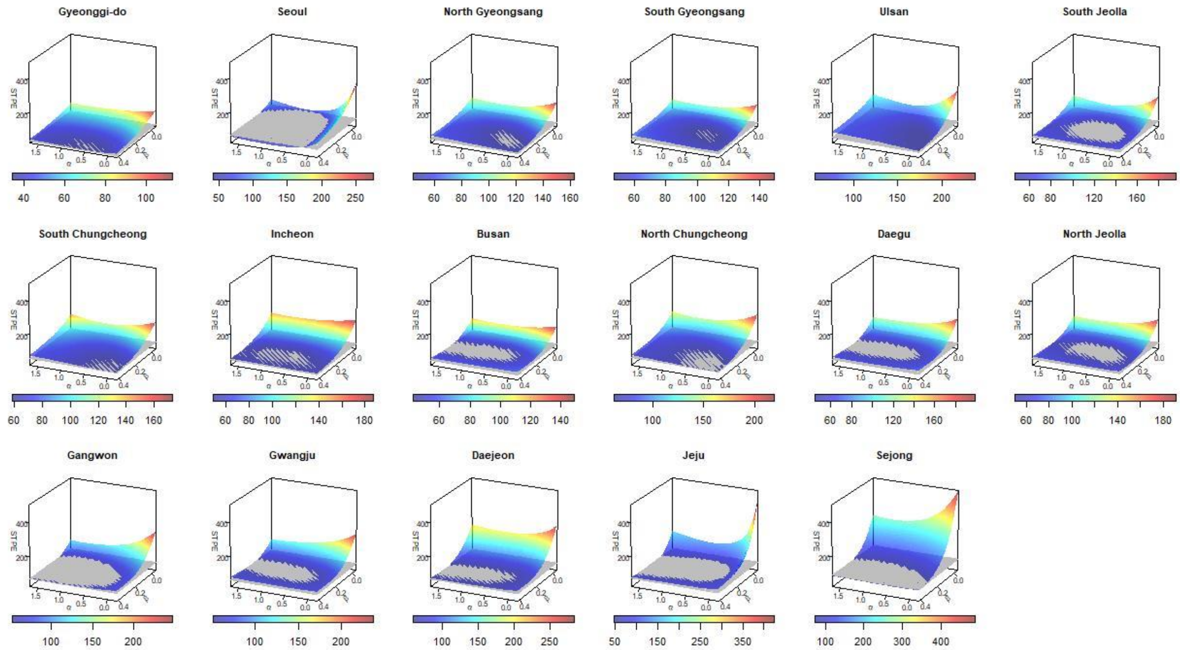

Figure 6. Comparison between STPE values for NP1 (grey) and 2D-LQ: year 2015
Source: own elaboration