Uchida *Journal of Economic Structures* (2026) 15:7
https://doi.org/10.1186/s40008-026-00373-w

Journal of Economic Structures

RESEARCH **Open Access**

Validation of the formulation of location quotient methods using regression analysis

Mizuki Uchida[1*]

*Correspondence:
Mizuki Uchida
mz-uchida@yachiyo-eng.co.jp
[1]Research Institute for Infrastructure Paradigm Shift, Yachiyo Engineering Co., Ltd., 5-20-8, Asakusabashi, Taito-Ku 111-8648, Tokyo, Japan

Abstract
The location quotient (LQ) method, a widely used nonsurvey method, has been improved through a series of empirical studies, resulting in variations that incorporate various factors. These variations include the simple LQ (SLQ) method, which considers the concentration of the supply sector; the cross-industry LQ (CILQ) method, which considers the concentration of both the supply and purchase sectors; and Flegg's LQ (FLQ) method, which modifies the CILQ method by accounting for regional size. However, LQ methods have long been criticized for their emphasis on mechanical convenience and lack of theoretical foundation, and previous studies have not adequately addressed these issues. In addition, the formulation of the conventional LQ method, which emphasizes simplicity, makes it difficult to statistically test the explanatory power of the factors introduced by each method. Therefore, we use regression models that generalize conventional LQ methods to elucidate the explanatory power of the included factors and test the validity of the conventional formulations. Our analysis reveals partial validation of the SLQ method, with the generalized SLQ method achieving an accuracy comparable to that of the conventional FLQ method. Furthermore, the variables introduced by the CILQ method, which represent the concentration in the purchase sector, have negligible explanatory power. Finally, the parameters of the FLQ method function primarily as constant terms and dummy variables rather than reflecting the effect of regional size.

Keywords Regional input–output tables, Nonsurvey methods, LQ methods, Regression analysis, SLQ, CILQ, FLQ

1 Introduction

In regional economic analysis, accurate assessment of the flow of goods and services is crucial. However, national input–output tables do not capture these flows between regions within the nation (Miller and Blair 2009). Thus, the regionalization of national input–output tables is a fundamental task in regional economic analysis. This process is typically accomplished using one of three primary methods—survey methods, nonsurvey methods, or semisurvey (hybrid or partial survey) methods.

Among the three methods, survey methods are generally considered to provide the most accurate results, as they directly capture comprehensive flows of goods and services. However, the considerable cost and time requirements associated with survey methods often make them impractical (Henry et al. 1980; Szabó 2015). Therefore, nonsurvey or semisurvey methods are usually preferred for constructing input–output tables in smaller regions.

To mitigate the practical limitations of survey methods, nonsurvey and semisurvey methods use estimates derived from existing statistical data to replace all or part of the survey data. The main difference between these approaches is their reliance on primary surveys. While nonsurvey methods are substantially less costly and time-consuming, they sacrifice accuracy compared with survey-based approaches. Thus, semisurvey methods that incorporate partial surveys are often preferred in contexts where both cost-effectiveness and accuracy are critical. However, because semisurvey methods are based on nonsurvey methods, it is critical to use the best nonsurvey method possible (Lahr 1993). Therefore, improving the accuracy of nonsurvey methods remains an important goal, especially for small regional analyses.

The location quotient (LQ) method is a widely used technique for regionalizing national input–output tables (Schaffer and Chu 1969; Szabó 2015). This method estimates regional input coefficients on the basis of the empirical observation that higher LQs, indicating greater industrial concentration, are correlated with lower-level import dependence. There are several variations in the calculation of LQs, and each LQ method is typically identified by the specific LQ used.

The simple LQ (SLQ) method is the basic LQ technique that has served as the basis for many subsequent developments. The primary advantage of this method is the mechanical convenience of SLQ calculations and censoring procedures. However, the SLQ method is known to have the problem of underestimating cross-hauling, i.e., the simultaneous import and export of the same type of product (Miller and Blair 2009).

Consequently, numerous alternative LQ methods have been developed to address the limitations of the SLQ method. These methods include the cross-industry LQ (CILQ) method (Schaffer and Chu 1969), which accounts for the concentration of both the supply and purchase sectors, and Flegg's LQ (FLQ) method (Flegg et al. 1995; Flegg and Webber 1997), which modifies the CILQ method to reflect regional size and is considered one of the more accurate LQ methods (Bonfiglio and Chelli 2008; Flegg and Tohmo 2013, 2016).

Conversely, LQ methods have been criticized for lacking theoretical support and for overemphasizing mechanical convenience (Round 1983). Hewings and Jensen (1987) state that implementing a regression-based direction such as that of Stevens et al. (1983) would be more productive than continuing to tinker with nonsurvey methods of dubious theoretical foundation. Although Flegg and Webber (1997) recognize the importance of such regression analysis in reexamination of the FLQ method, the majority of empirical studies have focused solely on comparing the predictive accuracy of different LQ methods. As a result, the fundamental question regarding the statistical validity of LQ method formulations remains largely unanswered.

To address this critical issue, we propose a generalized regression model framework for validating LQ methods, thereby reevaluating their formulations. Section 2 provides an overview of LQ methods and the regression-based approach. Section 3 presents representations of the SLQ, CILQ, and FLQ methods in the unified regression framework. Section 4 elucidates the explanatory power of the constituent variables and assesses the validity of the LQ method formulations by comparing the parameter estimation results and predictive accuracy between conventional and generalized approaches.

2 Literature review

2.1 Conventional LQ methods

The primary objective of LQ methods is to derive regional input coefficients from the corresponding national coefficients. In this context, we focus on the regionalization of Variant B tables, as described by Kronenberg (2012), since LQ methods are theoretically consistent only with these tables. We adopt the input and technical coefficient definitions delineated in Flegg et al. (1995). Let x_{ij}^{rs} denote the input from region *r* to region *s* in supply sector *i* and purchase sector *j*. Furthermore, let x_j^r represent the total input in region *r* and the purchase sector *j*. The regional input coefficient and the technical input coefficient for region *r* in nation *n* are defined, respectively, as follows:

$$a_{ij}^{rr} = \frac{x_{ij}^{rr}}{x_j^r} \tag{1}$$

$$a_{ij}^{r} = \frac{x_{ij}^{nr}}{x_j^r} \tag{2}$$

Specifically, the national technical coefficient, denoted by a_{ij}^n, is obtained by setting $r = n$ in Eq. (2), resulting in the following:

$$a_{ij}^{n} = \frac{x_{ij}^{nn}}{x_j^n} \tag{3}$$

The regional input coefficient a_{ij}^{rr} is derived by applying two adjustment factors to the national technical coefficient a_{ij}^n:

$$a_{ij}^{rr} = \gamma_{ij}^{rr}\gamma_{ij}^{r}a_{ij}^{n} \tag{4}$$

where γ_{ij}^r represents the difference in production technology between the region and the nation and γ_{ij}^{rr} indicates the share of inputs supplied by a region. Consequently, the primary objective of nonsurvey methods is to estimate the product $\gamma_{ij}^{rr}\gamma_{ij}^{r}$ (Miller and Blair 2009).

Given data limitations, it is standard practice to assume that regional production technology is identical to national production technology, i.e., $\gamma_{ij}^r = 1$ (Miller and Blair 2009). Under this assumption, conventional LQ methods typically adopt a censoring approach for the estimated adjustment factor $\hat{\gamma}_{ij}^{rr}$:

$$\begin{aligned} \hat{\gamma}_{ij}^{rr} &= \begin{cases} \mathrm{LQ}_{ij}^r & \text{if } \mathrm{LQ}_{ij}^r < 1 \\ 1 & \text{if } \mathrm{LQ}_{ij}^r \geq 1 \end{cases} \\ &= \min\{1, \mathrm{LQ}_{ij}^r\} \end{aligned} \tag{5}$$

The SLQ method, the most basic approach (Schaffer and Chu 1969), determines the LQ solely on the basis of the relative size of the supply sector *i*:

$$\mathrm{SLQ}_i^r = \frac{x_i^r / x^r}{x_i^n / x^n} \tag{6}$$

where x_i^r and x^r are the output of sector *i* and total output in region *r*, respectively, and x_i^n and x^n are the corresponding national values.

A theoretical limitation of the SLQ is that it underestimates cross-hauling. Specifically, if a region is specialized in sector *i* (i.e., $\mathrm{SLQ}_i^r > 1$), the method assumes that all local demand is met by local supply, thereby disregarding the potential for simultaneous imports arising from product differentiation. Stevens et al. (1988) note that cross-hauling is a common phenomenon and that the SLQ method generally produces biased results. Although the censoring procedure is a major contributor to this problem (Lahr et al. 2020), most scientific attention has been given to the LQ index calculation. As a result, numerous variations of the LQ index calculation have been proposed.

The CILQ method was an early attempt to improve the SLQ method (Schaffer and Chu 1969), where the CILQ for supply sector *i* and purchase sector *j* is defined as follows:

$$\mathrm{CILQ}_{ij}^r = \frac{\mathrm{SLQ}_i^r}{\mathrm{SLQ}_j^r} \tag{7}$$

However, since $\mathrm{CILQ}_{ii}^r = 1$, the CILQ cannot adjust the coefficients on the diagonal. Therefore, in the CILQ method, the following equation is used to adjust the diagonal coefficients using SLQ_i^r (Miller and Blair 2009):

$$\begin{aligned} \mathrm{LQ}_{ij}^r &= \begin{cases} \mathrm{CILQ}_{ij}^r & \text{if } i \neq j \\ \mathrm{SLQ}_i^r & \text{if } i = j \end{cases} \\ &= \frac{\mathrm{SLQ}_i^r}{(\mathrm{SLQ}_j^r)^{[i \neq j]}} \end{aligned} \tag{8}$$

where [*P*] denotes the Iverson bracket, which equals 1 if the proposition *P* is true and 0 otherwise. Thus, the exponent $[i \neq j]$ takes the value of 1 when $i \neq j$ and 0 when $i = j$.

Flegg and Webber (1997) noted that the SLQ method assumes that a regional supply sector is equally capable of meeting the demands of regional purchase sectors. In contrast, the CILQ method reflects the balance between the regional supply and purchase sectors. Consequently, the CILQ method can mitigate the underestimation of cross-hauling inherent in the SLQ method because it is not uniform across the supply sector, unlike the latter.

However, Flegg et al. (1995) reported that conventional LQ methods do not adequately resolve the underestimation of cross-hauling because they do not account for relative regional size. Consequently, Flegg et al. (1995) and Flegg and Webber (1997) proposed the FLQ method, which further incorporates the effect of regional size into the CILQ method. The FLQ method defines the LQ for supply sector *i* and purchase sector *j* as follows:

$$\begin{aligned} \mathrm{FLQ}_{ij}^r &= \begin{cases} \mathrm{CILQ}_{ij}^r \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]^\delta & \text{if } i \neq j \\ \mathrm{SLQ}_i^r \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]^\delta & \text{if } i = j \end{cases} \\ &= \frac{\mathrm{SLQ}_i^r}{(\mathrm{SLQ}_j^r)^{[i \neq j]}} \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]^\delta \end{aligned} \tag{9}$$

where

$$0 \leq \delta < 1$$

Here, δ is a parameter indicating that the degree of import dependence decreases with increasing regional size. This formulation is based on the idea that larger regions tend to have a more diverse range of industries. This increases the likelihood of finding suitable local suppliers and reduces reliance on imports (Flegg and Webber 2000). Several studies have shown that the FLQ method outperforms other LQ methods (Bonfiglio and Chelli 2008; Flegg and Tohmo 2013, 2016). Previous empirical studies, such as Flegg and Webber (1997), Flegg and Webber (2000) and Flegg and Tohmo (2013), suggest that a value of δ between 0.2 and 0.3 is optimal.

Given the high-level accuracy of the FLQ method, several extensions have been proposed. The simplest extension involves the refinement of the δ parameter. Kowalewski (2015) extended the original FLQ method and proposed the industry-specific LQ (SFLQ) method, where δ_j is defined for each purchase sector *j*.

A major challenge in applying the FLQ and SFLQ methods is estimating the parameter δ. Flegg et al. (2025) argued that it is difficult to select an appropriate value for δ in the FLQ method without surveys because the optimal value of δ varies by region and sector. They also criticized the lack of theoretical justification for adjusting the specialization coefficient according to regional size.

2.2 Alternative LQ formulations

In addition, other LQ methods have been proposed to improve the FLQ method. Hereafter, the estimated regional input coefficients are denoted by $\hat{a}_{ij}^{rr}$. Flegg and Webber (2000) proposed the AFLQ method, which estimates that the regional input coefficients are determined as follows:

$$\hat{a}_{ij}^{rr} = \begin{cases} \mathrm{FLQ}_{ij}^{r} \log_2\left(1 + \mathrm{SLQ}_j^r\right) a_{ij}^n & \text{if } \mathrm{SLQ}_j^r > 1 \\ \mathrm{FLQ}_{ij}^{r} a_{ij}^n & \text{if } \mathrm{SLQ}_j^r \leq 1 \end{cases} \tag{10}$$

While it is based on the FLQ method, the AFLQ method uses a different conditional branching mechanism than conventional LQ methods do, allowing for $a_{ij}^{rr} > a_{ij}^{n}$. However, Flegg and Tohmo (2013) and Lamonica and Chelli (2018) reported that compared with the FLQ method, the AFLQ method provides little or no improvement in accuracy.

Pereira-López et al. (2020) proposed the 2DLQ method. In this method, the regional input coefficients are determined as follows:

$$\hat{a}_{ij}^{rr} = \begin{cases} (\mathrm{SLQ}_i^r)^{\alpha} a_{ij}^n \left(\frac{x_j^r}{x_j^n}\right)^{\beta} & \text{if } \mathrm{SLQ}_i^r \leq 1 \\ \left[\frac{1}{2}\tanh(\mathrm{SLQ}_i^r - 1) + 1\right]^{\alpha} a_{ij}^n \left(\frac{x_j^r}{x_j^n}\right)^{\beta} & \text{if } \mathrm{SLQ}_i^r > 1 \end{cases} \tag{11}$$

By introducing parameters α and β, the 2DLQ method aims to capture the distinct characteristics of the supply and purchase sectors separately (Pereira-López et al. 2020). Like the AFLQ method, the 2DLQ method accommodates cases where $a_{ij}^{rr} > a_{ij}^{n}$ by using a different conditional branching mechanism than conventional LQ methods do. Pereira-López et al. (2020) show that the 2DLQ method is the generalized formulation of the FLQ method.

Both the AFLQ and 2DLQ methods are extensions of the FLQ method, using more complex conditional branching than does the FLQ method to accommodate cases where

$a_{ij}^{rr} > a_{ij}^{n}$. This increased complexity arises from their attempt to address two subproblems inherent in the regionalization of the input coefficients simultaneously.

Kwon and Choi (2024) proposed the KFLQ method to mitigate the overestimation of regional input coefficients inherent in the FLQ formula when both the supply and purchase sectors have low concentration. The KFLQ method defines the LQ for supply sector *i* and purchase sector *j* as follows:

$$\mathrm{KFLQ}_{ij}^{r} = \begin{cases} \mathrm{SLQ}_{i}^{r}\left[\log_2\left(1+\frac{x^r}{x^n}\right)\right]^{\delta} & \text{if } \left(i \neq j \wedge \mathrm{SLQ}_{i}^{r} < 1 \wedge \mathrm{SLQ}_{j}^{r} < 1\right) \vee i = j \\ \mathrm{CILQ}_{ij}^{r}\left[\log_2\left(1+\frac{x^r}{x^n}\right)\right]^{\delta} & \text{otherwise} \end{cases} \tag{12}$$

The KFLQ method incorporates the censoring procedure defined by Eq. (5). Empirical tests using the 2010 regional input–output tables for South Korea demonstrated that the KFLQ method yields more accurate estimates than traditional LQ methods do. Furthermore, Tohmo (2025b) reevaluated the KFLQ method using the 2015 South Korean tables and confirmed its consistent superiority to the standard FLQ method in reducing the mean absolute percentage error (MAPE) across most regions.

Furthermore, Flegg et al. (2025) proposed the hyperbolic tangent LQ (HTLQ) method as an alternative to the FLQ method. Unlike the complex formulations of the AFLQ and 2DLQ methods, the HTLQ method is designed for straightforward implementation while maintaining high accuracy. It accounts for the relative size of the purchase sector by transforming the CILQ using a hyperbolic tangent function. The formula is defined as follows:

$$\hat{a}_{ij}^{rr} = \begin{cases} 0 & \text{if } \mathrm{CILQ}_{ij}^{r} = 0 \\ \mu\left[\tanh\left(\mathrm{CILQ}_{ij}^{r} - 1\right) + 1\right] a_{ij}^{n} & \text{if } 0 < \mathrm{CILQ}_{ij}^{r} \leq 1 \\ a_{ij}^{n} & \text{if } \mathrm{CILQ}_{ij}^{r} > 1 \end{cases} \tag{13}$$

where μ $(0 < \mu < 1)$ is a parameter that acts as a proxy for regional self-sufficiency. In the absence of prior information, Flegg et al. (2025) recommended setting $\mu = 0.5$ because this simplifies the method and eliminates the need for complex parameter estimation.

A key advantage of the HTLQ method is its ability to allow for a nonzero intercept. This reflects a threshold effect, assuming a minimal regional supply structure once an industry is established. Based on empirical applications using South Korean tables for 2005 and 2015, the HTLQ method demonstrates superior accuracy over the FLQ method, outperforming it in the vast majority of regions on the basis of the standardized total percentage error (STPE). Crucially, fixing the parameter $\mu = 0.5$ yields consistently lower errors than the standard fixed FLQ does ($\delta = 0.3$) and often surpasses the FLQ even when its parameter is optimized, highlighting its practical utility when regional data are scarce.

Although the FLQ method improves accuracy by incorporating regional size, comparisons with alternative methods indicate that there is still room for refinement. For instance, while a correlation between the μ parameter of the HTLQ method and regional size has been confirmed—suggesting that μ partially accounts for the regional scale—the HTLQ method often outperforms the FLQ method even when μ is fixed. Consequently, the conventional view that improving FLQ accuracy relies primarily on considering the region's scale requires further verification.

2.3 Regression-based analysis of LQ methods

Among these variants of the LQ method, the SLQ method has at least a partial theoretical basis for the LQ index calculation. In particular, Stevens et al. (1983) show that the SLQ appears as an explanatory variable in the regression model for the regional purchasing coefficient (RPC), which is closely related to the regional input coefficient. Hewings and Jensen (1987) suggest that implementing a regression-based approach, as exemplified by Stevens et al. (1983), would be more productive than continuing to refine non-survey methods with dubious theoretical underpinnings.

Lahr et al. (2020) introduce a quasi-binomial regression model to estimate RPC and evaluate the predictive accuracy of models using SLQ and FLQ-equivalent methods. The quasi-binomial regression model that they used is suitable for constraining estimates of RPC within the interval of 0 to 1 and accommodates overdispersion or underdispersion through its flexible error structure. They report that a model with a constant and a censored SLQ as explanatory variables yields $R^2 = 0.142$. In addition, the FLQ-equivalent model, which adds the employment share as a variable, slightly improves R^2 to 0.195.

However, as noted by Flegg et al. (2021), Lahr et al. (2020) use row-only estimation, a departure from the cross-industry foundation of the original FLQ method. In addition, they use an additive model equation without logarithmic transformation of the SLQ. As a result, the appropriateness of the model proposed by Lahr et al. (2020) for testing the original SLQ and FLQ methods remains questionable.

Furthermore, Flegg et al. (2021) proposed a regression model to estimate δ given estimated regional input coefficients. Flegg et al. (2021) use the following relationship to estimate regional input coefficients:

$$\begin{aligned} \hat{a}_{ij}^{rr} &= \mathrm{FLQ}_{ij}^{r} a_{ij}^{n} \\ &= \frac{\mathrm{SLQ}_i^r}{(\mathrm{SLQ}_j^r)^{[i \neq j]}} \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]^{\delta^r} a_{ij}^{n} \end{aligned} \tag{14}$$

The following regression model is thus obtained:

$$\ln\left(\frac{\hat{a}_{ij}^{rr}}{\frac{\mathrm{SLQ}_i^r}{(\mathrm{SLQ}_j^r)^{[i \neq j]}} a_{ij}^{n}}\right) = \delta^r \ln\left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right] \tag{15}$$

Flegg et al. (2021) estimated the parameter δ^r via a regression-like approach. Specifically, this estimation was performed by dividing the mean of the regressand by the region-specific explanatory variable.

This model has three important limitations. First, the absence of an intercept deviates from standard regression practice and compromises the interpretation of δ^r as a true measure of regional size influence. Second, the inclusion of $\frac{\mathrm{SLQ}_i^r}{(\mathrm{SLQ}_j^r)^{[i \neq j]}}$ in the offset term assumes the validity of the CILQ method itself. Third, while we acknowledge that the impact of regional size may vary across regions because of other unobserved factors, attempting to capture this variation by regionally defining the parameter as δ^r creates an identification problem. The original FLQ method expresses the influence of regional size using a common parameter, δ, as shown in Eq. (9). However, by regionally defining this parameter as δ^r, it absorbs all region-specific heterogeneity, effectively neutralizing the explicit influence of the regional size variable. In other words, δ^r can be adjusted to fit the target coefficients regardless of the actual regional size. As a result, δ^r essentially functions as a regional dummy variable rather than a parameter representing the specific influence of regional size.

In contrast to the aforementioned methods, Flegg and Tohmo (2013) proposed a radically different approach, employing the following regression model to directly estimate δ:

$$\ln \hat{\delta} = -1.8379 + 0.33195 \ln R + 1.5834 \ln P - 2.8812 \ln I \tag{16}$$

where R represents regional size as a percentage of output, P represents a survey-based estimate of the average propensity to import for each region, and I represents a survey-based estimate of the average use of intermediate inputs for each region.

Furthermore, Kowalewski (2015) analyzed the factors that determine the parameter δ_j in the SFLQ method using survey-based data from Baden-Wurttemberg, Germany. The author estimated an OLS regression model in which the optimal δ_j was regressed against four explanatory variables: the coefficient of localization (CL_j), the simple LQ (SLQ_j^r), the share of imports (IM_j), and the share of value added (VA_j). The model is expressed as follows:

$$\delta_j = \alpha + \beta_1 \mathrm{CL}_j + \beta_2 \mathrm{SLQ}_j^r + \beta_3 \mathrm{IM}_j + \beta_4 \mathrm{VA}_j + e_j \tag{17}$$

The empirical results yield an R^2 value of 0.67, which indicates that CL_j is the only statistically significant predictor with a coefficient of $\beta_1 = 1.13$. Building on the work of Flegg and Tohmo (2019), Tohmo (2025a) applied this regression-based SFLQ method to the 2015 South Korean tables and demonstrated that the double-log specification of Eq. 17 outperforms the conventional linear model in terms of the MAPE.

These approaches focus solely on determining δ or δ_j and thus presuppose the validity of the CILQ method. Consequently, they are insufficient for validating the FLQ method itself.

On the basis of the previous analysis, the formulation of Lahr et al. (2020), despite using a more appropriate regression model than do other studies, remains inappropriate for validating the LQ methods owing to its focus on row-specific RPCs. Conversely, the regression models, like those proposed by Flegg et al. (2021), Flegg and Tohmo (2013) and Kowalewski (2015), which rely on a fixed formulation, are also inappropriate for validating the LQ methods.

Therefore, we aim to generalize conventional LQ methods into a regression model that allows us to statistically test these methods. By formulating conventional LQ methods as special cases within our regression model, we can statistically quantify the explanatory power of the variables used in the LQ methods and identify the divergence in the parameter estimates between conventional and generalized methods. In this study, we focus on the three most common LQ methods—the SLQ, CILQ, and FLQ methods. This focus is justified because many other LQ methods, such as the AFLQ and 2DLQ methods, are modified versions of these core methods and inherit their limitations.

3 Methods

3.1 Generalization of the LQ methods to regression models

To statistically validate the SLQ, CILQ, and FLQ methods, we extend the regression model proposed by Flegg et al. (2021). Assuming that $\mathrm{FLQ}_{ij}^{r} < 1$, we derive the following relationship from Eqs. 5 and 9:

$$\begin{aligned}\hat{\gamma}_{ij}^{rr} &= \mathrm{FLQ}_{ij}^{r} \\ &= \frac{\mathrm{SLQ}_{i}^{r}}{(\mathrm{SLQ}_{j}^{r})^{[i \neq j]}} \left[\log_2 \left(1 + \frac{x^r}{x^n}\right)\right]^{\delta_{ij}^{r}}\end{aligned} \tag{18}$$

where δ_{ij}^{r} is a parameter that extends the δ of the original FLQ method. Logarithmizing this equation yields the following equation:

$$\ln \hat{\gamma}_{ij}^{rr} = \ln \mathrm{SLQ}_{i}^{r} - [i \neq j] \ln \mathrm{SLQ}_{j}^{r} + \delta_{ij}^{r} \ln \left[\log_2 \left(1 + \frac{x^r}{x^n}\right)\right] \tag{19}$$

By generalizing this equation, we obtain the following regression model:

$$g(\hat{\gamma}_{ij}^{rr}) = \beta_0 + \beta_1 \ln \mathrm{SLQ}_{i}^{r} + \beta_2 [i \neq j] \ln \mathrm{SLQ}_{j}^{r} + \delta_{ij}^{r} \ln \left[\log_2 \left(1 + \frac{x^r}{x^n}\right)\right] \tag{20}$$

where $g(\cdot)$ is a link function and β_0, β_1, and β_2 are parameters.

The SLQ, CILQ, and FLQ methods are derived by setting the parameters in Eq. (20) with a $\ln(\cdot)$ link, as detailed in Table 1, where *N* is the number of industries and *S* is the number of regions. There are several approaches to defining δ_{ij}^{r} in the extended FLQ methods, and these variations are distinguished by subscripts. In particular, the FLQ (by column) method used in Kowalewski (2015) is referred to as the SFLQ method. To increase the comprehensiveness of our validation, we also include the FLQ (by row) method, despite its limited use in previous research.

We refer to models with fixed parameters, as specified in Table 1, as conventional methods and those with varying parameters as generalized methods. In general, the generalized FLQ methods have three additional parameters compared with the conventional methods because of the inclusion of β_0, β_1, and β_2, except for the FLQ (by region) method. In the FLQ (by region) method, however, the model specification differs fundamentally. Since the region-specific parameters effectively function as regional dummy variables, the inclusion of a global intercept is precluded. Therefore, we set the constant term to zero to ensure that all dummy variables are included in the estimation.

Both conventional and corresponding generalized methods are estimated using quasi-Poisson regression with an $\ln(\cdot)$ link, a type of generalized linear model (GLM).

Table 1 Parameter setting for LQ methods

	Parameter setting				Number of parameters		
Method	β_0	β_1	β_2	δ_{ij}^{r}	**Conventional**	**Generalized**	**Previous studies**
SLQ	0	1	–	–	0	2	Schaffer and Chu (1969)
CILQ	0	1	-1	–	0	3	Schaffer and Chu (1969)
FLQ (common)	0	1	-1	δ	1	4	Flegg and Webber (1997)
FLQ (by row)	0	1	-1	δ_i	*N*	$N+3$	–
FLQ (by column)	0	1	-1	δ_j	*N*	$N+3$	Kowalewski (2015)
FLQ (by region)	0	1	-1	δ^r	*S*	$S+2$	Flegg et al. (2021)

Quasi-Poisson regression is analogous to quasi-binomial regression in that both allow for flexible error structures. However, this method differs in that it models data greater than or equal to 0 with no upper bound, unlike quasi-binomial regression, which models proportions between 0 and 1. Quasi-Poisson regression assumes that the variance in the data is proportional to the expected value and typically uses the $\ln(\cdot)$ link. Importantly, this value can be estimated even when the target variable contains zeros, making it suitable for estimating gravity models (Silva and Tenreyro 2006).

These properties of quasi-Poisson regression provide two important advantages for estimating $\hat{\gamma}_{ij}^{rr}$ in Eq. (22). First, since the observed value of Eq. (22) often takes values of 0 or greater than 1, this method allows for the estimation of such data without exclusion or censoring. Second, the use of the $\ln(\cdot)$ link function fits well with the existing model formulation presented in Eq. (19).

From Eqs. (1) and (4), we derive the following relationship:

$$x_{ij}^{rr} = \gamma_{ij}^{rr}\gamma_{ij}^{r}a_{ij}^{n}x_{j}^{r} \tag{21}$$

Assuming that $\gamma_{ij}^{r} = 1$, we can obtain γ_{ij}^{rr} as follows:

$$\gamma_{ij}^{rr} = \frac{x_{ij}^{rr}}{a_{ij}^{n}x_{j}^{r}} \tag{22}$$

Therefore, treating x_{ij}^{rr} as the response variable, we use the natural logarithm of the denominator, $\ln\left(a_{ij}^{n}x_{j}^{r}\right)$, as an offset term to mitigate excessive variation associated with ratios when the denominator is small. Furthermore, in conventional methods, we include fixed parameters in the offset term to constrain the parameter estimation process.

Importantly, the $\hat{\gamma}_{ij}^{rr}$ estimates obtained from the quasi-Poisson regression differ from those of the traditional SLQ, CILQ, and FLQ methods in that they lack an upper bound of 1. To maintain consistency with traditional LQ methods and to clarify the validity of the censoring procedure, we compare the results of the generalized methods obtained with and without posterior censoring of $\hat{\gamma}_{ij}^{rr}$. In the case of posterior censoring, we impose an upper bound of 1, which is consistent with the standard method shown in Eq. (5).

3.2 Accuracy evaluation metrics for the LQ methods

We evaluate the accuracy of the LQ methods by comparing the actual and estimated regional input coefficients and Leontief inverse matrix coefficients as follows:

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1N} \\ \vdots & \ddots & \vdots \\ b_{N1} & \cdots & b_{NN} \end{bmatrix} = (I - A)^{-1} \tag{23}$$

where

$$A = \begin{bmatrix} a_{11}^{rr} & \cdots & a_{1N}^{rr} \\ \vdots & \ddots & \vdots \\ a_{N1}^{rr} & \cdots & a_{NN}^{rr} \end{bmatrix}$$

We denote the actual and estimated values of the regional input coefficient a_{ij}^{rr} and the Leontief inverse matrix coefficients b_{ij} as y_{ij} and $\hat{y}_{ij}$, respectively.

Lehtonen and Tykkyläinen (2014) proposed four metrics to evaluate the performance of LQ methods—the standardized total percentage error (STPE), mean weighted absolute error (MWAE), mean related bias of the inequality index (U^M), and variance in the inequality index (U^S), which are defined as follows:

$$\mathrm{STPE} = 100\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|\hat{y}_{ij} - y_{ij}|}{\sum_{i=1}^{N}\sum_{j=1}^{N} y_{ij}} \tag{24}$$

$$\mathrm{MWAE} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_j|\hat{y}_{ij} - y_{ij}| \tag{25}$$

$$U^M = \frac{[\mathrm{mean}(\hat{y}_{ij}) - \mathrm{mean}(y_{ij})]^2}{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{y}_{ij} - y_{ij})^2} \tag{26}$$

$$U^S = \frac{[\mathrm{sd}(\hat{y}_{ij}) - \mathrm{sd}(y_{ij})]^2}{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{y}_{ij} - y_{ij})^2} \tag{27}$$

where

$$\begin{aligned}\mathrm{mean}(\cdot) &: \text{mean function}\\ \mathrm{sd}(\cdot) &: \text{standard deviation function}\end{aligned}$$

Among these measures, the STPE and MWAE measure the absolute error between the actual and estimated values. However, the MWAE differs from the STPE in that it weights the importance of the industry on the basis of its share of employment. In addition, U^M and U^S quantify the differences between the expected value and the standard deviation of the actual and estimated values, respectively.

Flegg and Tohmo (2016) noted that the denominators of U^M and U^S are equivalent to the mean squared error (MSE), which means that increasing the MSE can decrease U^M and U^S; therefore, they proposed $\tilde{U}^M$ and $\tilde{U}^S$ as alternatives to U^M and U^S, respectively, which are defined as follows:

$$\tilde{U}^M = [\mathrm{mean}(\hat{y}_{ij}) - \mathrm{mean}(y_{ij})]^2 \tag{28}$$

$$\tilde{U}^S = [\mathrm{sd}(\hat{y}_{ij}) - \mathrm{sd}(y_{ij})]^2 \tag{29}$$

Furthermore, the above authors argued that the mean percentage error (MPE) and Theil's inequality index (U) should be added to the evaluation metrics, which are defined as follows:

$$\mathrm{MPE} = \frac{100}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\hat{y}_{ij} - y_{ij}}{y_{ij}} \tag{30}$$

$$U = 100\sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{y}_{ij} - y_{ij})^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} y_{ij}^2}} \tag{31}$$

Therefore, we evaluate the accuracy of the LQ methods using five metrics—STPE, MWAE, $\tilde{U}^M$, $\tilde{U}^S$, and U. We exclude the MPE because it diverges when the actual value is zero. For simplicity, we use the regional output value ratio w_j in the MWAE. Importantly, $\tilde{U}^M = 0$ and $\tilde{U}^S = 0$ do not necessarily mean that $\hat{y}_{ij} = y_{ij}$ for all i and j. Additionally, $\tilde{U}^S = 0$ does not necessarily mean that the expected values of $\hat{y}_{ij}$ and y_{ij} are the same. Conversely, $\mathrm{STPE} = 0$, $\mathrm{MWAE} = 0$ and $U = 0$ usually mean that $\hat{y}_{ij} = y_{ij}$ for all i and j.

3.3 Data sources

Since this study focuses on regionalizing national input–output tables, the regional input–output tables within a nation are preferred for validating LQ methods. Zhao and Choi (2015) and Flegg et al. (2021) employed the 2005 interregional input–output table for South Korea produced by the Bank of Korea (BOK), which covers 16 regions and 78 sectors, to examine the accuracy of various LQ-based nonsurvey methods. Zhao and Choi (2015) note that these data are considered a benchmark table because they are entirely survey-based. Subsequently, Tohmo (2025b) utilized the 2015 interregional input–output table for South Korea to evaluate the performance of the KFLQ method.

Therefore, we use the BOK's 2015 interregional input–output table for South Korea, which includes 17 regions and 33 industries ($S = 17, N = 33$), owing to its online availability. Importantly, the year and industry classification of our data differ from those used in Zhao and Choi (2015) and Flegg et al. (2021). In addition, the compilation methods and industry classifications of the BOK have been revised since 2010 (Kwon and Choi 2024); thus, caution is needed when this study is compared with previous research.

To estimate the regression models in this study, we need the national input coefficients from Eq. (22). Therefore, we aggregate and convert the data into a national input–output table and calculate these coefficients using Eq. (3). Similarly, the explanatory variables for the regression model in Eq. (20) are calculated from the interregional input–output table.

4 Results and discussion

4.1 Parameter estimation results for the LQ methods

Table 2 shows the parameter estimation results for the conventional and generalized methods. We omit the δ^r_{ij} estimation results for the FLQ (by row), FLQ (by column), and

Table 2 Estimation results of the regression models for the LQ methods

Type	Method	β_0	β_1	β_2	δ^r_{ij}
Conventional method	SLQ	–	–	–	–
	CILQ	–	–	–	–
	FLQ (common)	–	–	–	0.411***
	FLQ (by row)	–	–	–	dummy
	FLQ (by column)	–	–	–	dummy
	FLQ (by region)	–	–	–	dummy
Generalized method	SLQ	−0.768***	0.442***	–	–
	CILQ	−0.773***	0.441***	0.021**	–
	FLQ (common)	−0.471***	0.440***	0.025**	0.144***
	FLQ (by row)	−0.468***	0.502***	−0.022***	dummy
	FLQ (by column)	−0.488***	0.449***	0.038***	dummy
	FLQ (by region)	–	0.437***	0.008	dummy

The specific values for the dummy variables are presented in Fig. 1 due to their large number. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$

FLQ (by region) methods because of the large number of values represented by these dummy variables. The omitted δ_{ij}^r values are shown in a boxplot in Fig. 1.

In the following discussion, we evaluate the validity of the parameter fixing for β_0, β_1, and β_2 on the basis of the results presented in Table 2.

First, with respect to the coefficient on β_0, Table 2 shows that the intercept, which is fixed at zero in conventional methods, is significantly negative for all generalized LQ methods, except for the FLQ (by region) method, which has no intercept. Thus, these results do not support fixing β_0 to zero.

Second, the estimate for β_1, which corresponds to the coefficient for $\ln \mathrm{SLQ}_i^r$ derived from the SLQ method, ranges from approximately 0.4 to 0.5. Although these values deviate substantially from the unitary elasticity ($\beta_1 = 1$) assumed in the conventional method, they remain significantly positive. Notably, these results are close to the parameter estimates reported by Lahr et al. (2020), despite their use of RPC as the target variable and a censored SLQ_i^r as the explanatory variable. Thus, these results partially support the inclusion of this variable as a factor positively correlated with $\hat{\gamma}_{ij}^{rr}$, albeit with a smaller impact than traditionally assumed.

In contrast, the coefficient β_2, which corresponds to the term $[i \neq j] \ln \mathrm{SLQ}_j^r$ introduced by the CILQ method, substantially deviates from the value of -1 assumed by conventional methods. In fact, despite the conventional setting of -1, the estimated values are predominantly positive across most models, and in several cases, they are significantly positive. Therefore, there is little empirical support for the conventional assumption of fixing β_2 at -1. Moreover, the absolute values of β_2 are notably smaller than those of β_1, suggesting the weaker explanatory power of the former.

Next, we examine the estimation results of the region size parameters introduced by the FLQ method. Table 2 shows that for the FLQ (common) method, $\delta_{ij}^r \approx 0.411$ for

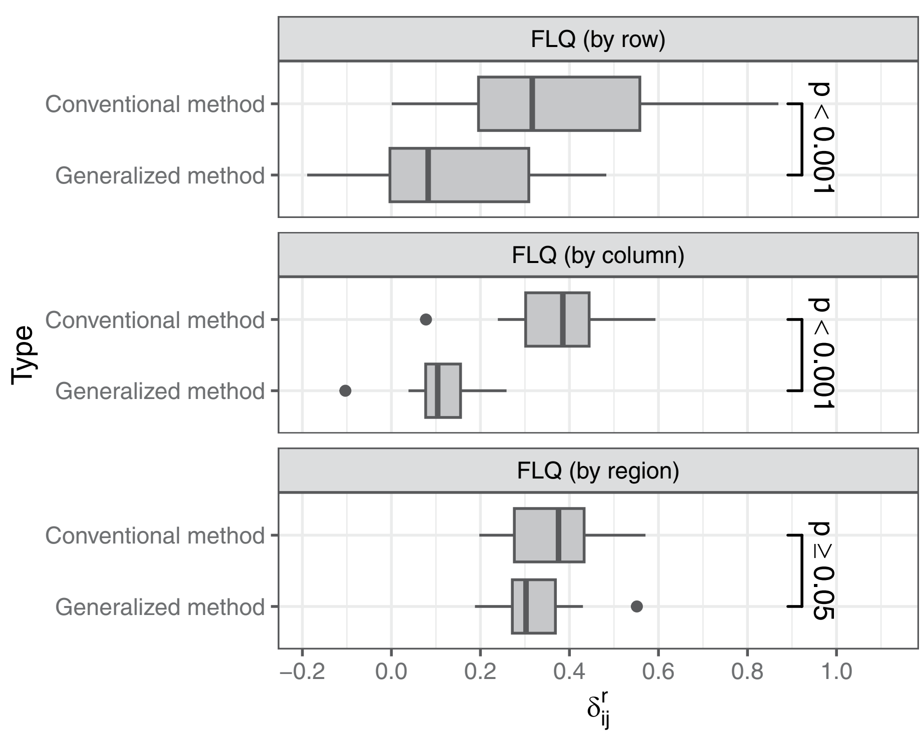

Fig. 1 Estimation results of the dummy variables δ_{ij}^r of the regression models for the LQ methods

Table 3 Comparison of data characteristics and estimation methods among studies

Study	Year	Regions (*S*)	Industries (*N*)	Estimation method
This study	2015	17	33	Quasi-Poisson regression model
Flegg et al. (2021)	2005	16	78	Regression-like model
Tohmo (2025b)	2015	17	33	Minimization of MAPE

Table 4 Comparison of estimated δ^r values in the FLQ (by region) method with those in prior studies

Region	This study	Flegg et al. (2021)	Tohmo (2025b)
Gangwon	0.228	0.24	0.2490
Gyeonggi	0.564	0.54	0.5140
South Gyeongsang	0.378	0.31	0.3190
North Gyeongsang	0.427	0.26	0.4010
Gwangju	0.265	0.32	0.2620
Daegu	0.247	0.26	0.2690
Daejeon	0.314	0.35	0.3790
Busan	0.292	0.30	0.3020
Seoul	0.559	0.37	0.3280
Sejong	0.296	–	0.4260
Ulsan	0.571	0.43	0.4800
Incheon	0.425	0.48	0.4140
South Jeolla	0.433	0.23	0.3800
North Jeolla	0.276	0.29	0.3010
Jeju	0.198	0.18	0.1770
South Chungcheong	0.435	0.39	0.4507
North Chungcheong	0.375	0.30	0.3770

the conventional method and $\delta_{ij}^r \approx 0.144$ for the generalized method, indicating a larger estimated value for the conventional method than for the generalized method. This trend is also observed for the FLQ (by row) and FLQ (by column) methods, as shown in Fig. 1, which shows a significant difference between the estimation results of the conventional and generalized approaches.

Furthermore, we compare the δ^r estimates obtained from the region-specific FLQ method in this study with those reported in prior studies (Flegg et al. 2021; Tohmo 2025b). Table 3 summarizes the data characteristics and estimation methods of the comparative studies. Table 4 presents the estimated values alongside those reported in previous studies. To further visualize the consistency of regional trends, Figure 2 displays a scatter plot matrix including the Pearson correlation coefficients (R) between these estimates.

Our estimates exhibit strong positive correlations with Flegg et al. (2021) and Tohmo (2025b) ($R = 0.71$ and 0.78, respectively), supporting the validity of our approach in identifying relative regional trends. However, the results are not identical. We attribute the observed differences in absolute values to variations in data characteristics—specifically, the target year and sectoral aggregation (N)—as well as the distinct estimation methods employed.

4.2 Accuracy evaluation results for the LQ methods

The accuracy evaluation results for the regional input coefficients and the Leontief inverse are presented in Figs. 3 and 4, respectively. In these plots, the bars represent the

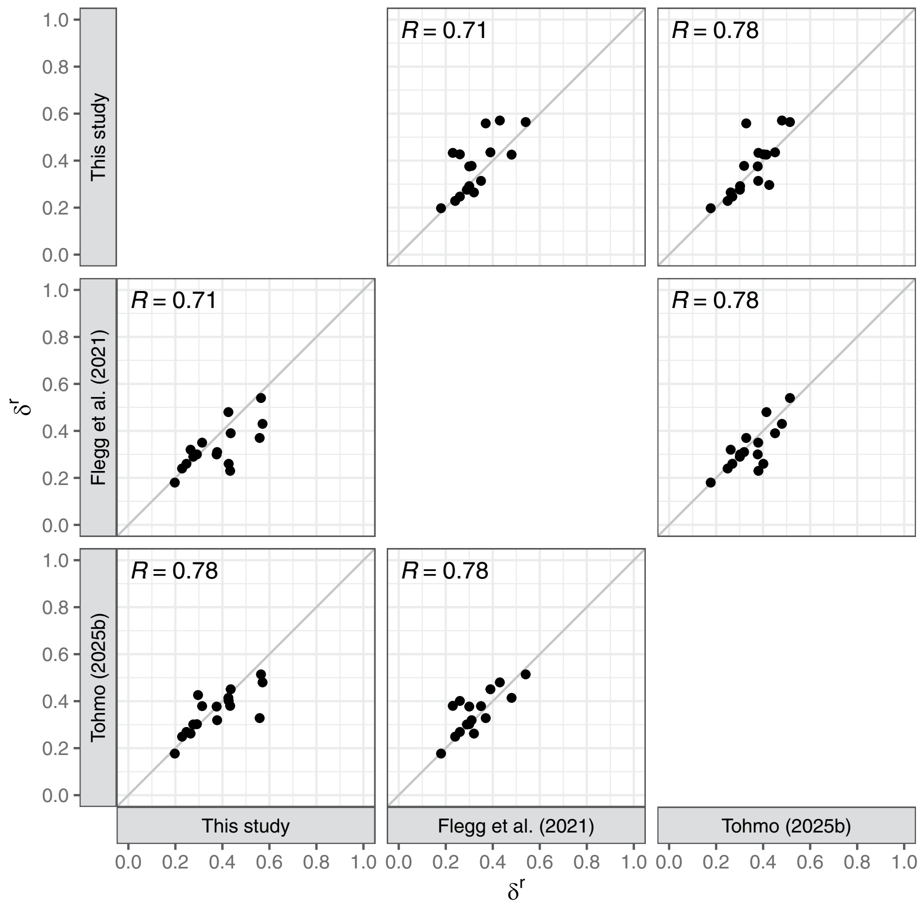

Fig. 2 Comparison of estimated δ^r values in the FLQ (by region) method with prior studies

mean errors by region, and the error bars represent the standard deviations of the errors within each region.

The generalized SLQ, CILQ, and FLQ methods showed minimal differences in accuracy with or without censoring (Figure 3). Therefore, it can be concluded that the censoring procedure provides limited benefit in this context.

In addition, the generalized methods generally show higher accuracy than their corresponding conventional methods do, except for the $\tilde{U}^S$ metric. According to Fig. 3, the generalized FLQ (by row) method achieves the greatest accuracy. However, importantly, this method has a larger number of variables, specifically $N + 3 = 36$. Furthermore, the observed increase in $\tilde{U}^S$ is thought to reflect the bias–variance tradeoff. Since these models utilize a larger number of parameters to effectively minimize systematic bias ($\tilde{U}^M$), they are likely to become more susceptible to noise in the data, which in turn increases the variance component ($\tilde{U}^S$).

Regardless of whether the method is generalized or conventional, the FLQ (by row) method consistently outperforms the FLQ (by column) method or the SFLQ method, despite having the same number of variables. This superior performance is likely due to the heterogeneity of the goods involved in cross-hauling, where the supply sector is a stronger determinant than the purchase sector is.

In addition, the accuracy of the generalized SLQ method is often comparable to that of the conventional FLQ method. While the generalized SLQ method has 2 parameters,

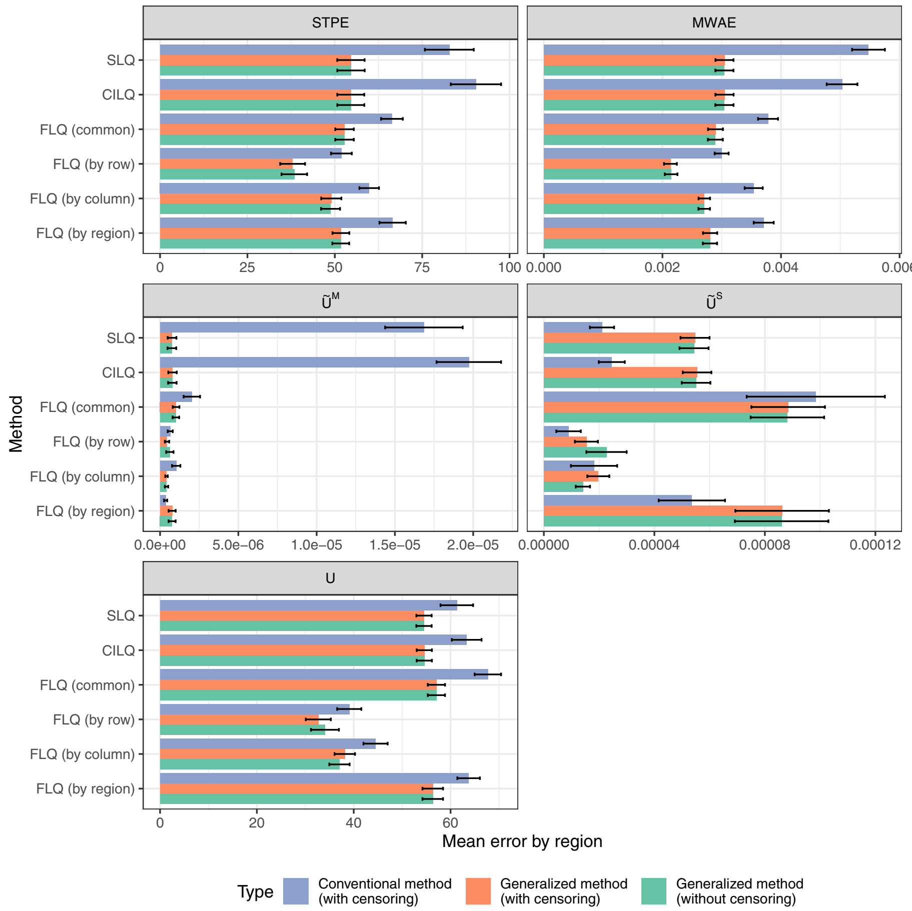

Fig. 3 Accuracy evaluation results for the regional input coefficients

which exceeds the single parameter of the FLQ (conventional) method, it uses fewer parameters than the FLQ (by row), FLQ (by column), and FLQ (by region) methods do, which have $N = 33$, $N = 33$, and $S = 17$ parameters, respectively. Thus, the generalized SLQ method achieves high accuracy with a relatively small number of parameters.

Given that the intercept in the generalized methods is significantly different from zero, this improvement in accuracy is consistent with the validation results of the HTLQ method by Flegg et al. (2025), which incorporates a coefficient. While direct comparisons require caution because of differences in functional forms, these findings suggest that explicitly including an intercept is effective for improving estimation accuracy.

Conversely, compared with the generalized SLQ method, the generalized CILQ method does not provide a substantial improvement in accuracy. This limited improvement is consistent with the observation that β_2, the coefficient introduced by the CILQ method, is smaller than β_1, the coefficient introduced by the SLQ method, indicating the lower effectiveness of the former. If the explanatory power of column-specific variables, such as $[i \neq j] \ln \mathrm{SLQ}_j^r$, appears to be limited, then a row-focused approach, such as the RPC regression used in Lahr et al. (2020), may be effective.

Figure 4 displays results broadly consistent with those in Fig. 3, with the notable exception that $\tilde{U}^S$ for the FLQ method is substantially smaller than that for the SLQ and CILQ methods. This outcome is likely attributable to the large systematic bias inherent in the conventional SLQ and CILQ methods. Specifically, the iterative calculations involved in

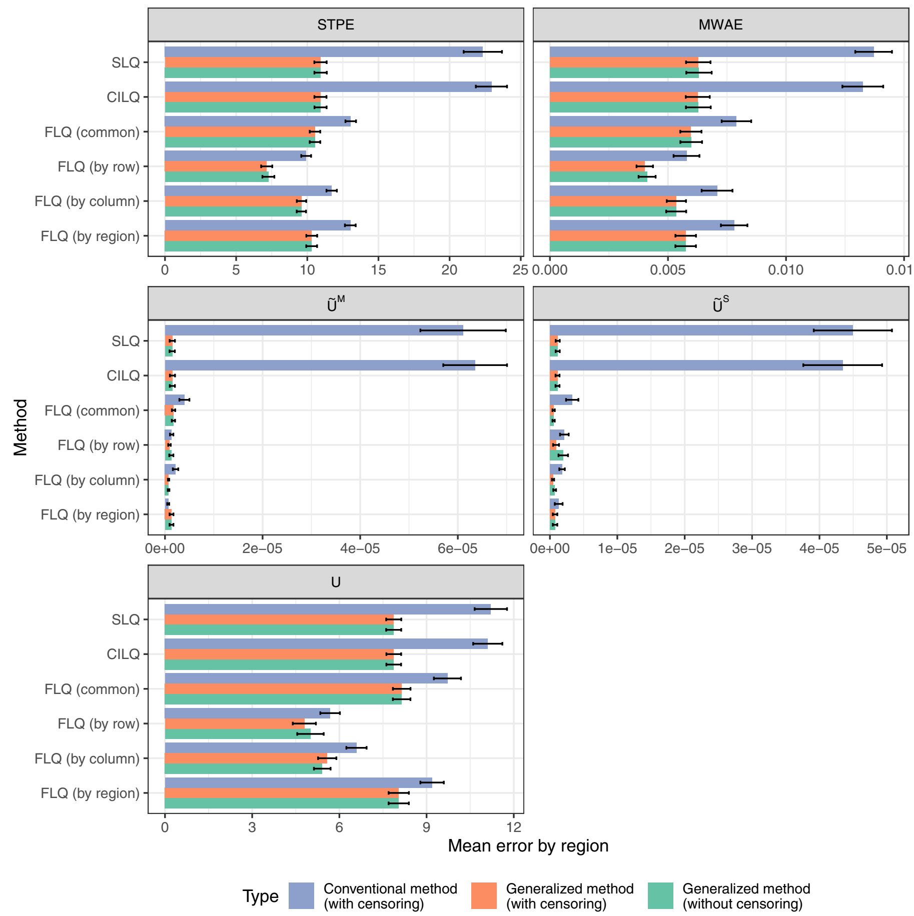

Fig. 4 Accuracy evaluation results for the Leontief inverse

matrix inversion can amplify this initial bias, causing the standard deviation of the estimated values to be significantly distorted. This distortion leads to inflated $\tilde{U}^S$ values for the conventional methods, whereas the FLQ method maintains greater stability.

4.3 Interpretation of the parameters of the FLQ method

On the basis of the previous results, the validity of $\ln \mathrm{SLQ}_i^r$ as introduced by the SLQ method is partially supported, in contrast to that of $[i \neq j] \ln \mathrm{SLQ}_j^r$ introduced by the CILQ method. What about the regional size variable, $\ln \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]$, introduced by the FLQ method?

Compared with the SLQ and CILQ methods, the FLQ (common) method is more accurate for both the regional input coefficients and the Leontief inverse. Does this indicate that the inclusion of regional size in the FLQ method is valid? To answer this question, we need to investigate why the estimated $\delta \approx 0.411$ in the conventional method is greater than the estimated $\delta \approx 0.144$ in the generalized method.

The distributions of the explanatory variables are shown in Fig. 5. As shown, $\ln \mathrm{SLQ}_i^r$ and $[i \neq j] \ln \mathrm{SLQ}_j^r$ are centered on zero, whereas $\ln \left[\log_2\left(1 + \frac{x^r}{x^n}\right)\right]$ consistently takes negative values because $x^r < x^n$. Consequently, a larger δ_{ij}^r value leads to a decrease in the expected value of the regression model.

We now return to the differences between the conventional and generalized FLQ methods. The conventional method fixes $\beta_0 = 0$, $\beta_1 = 1$, and $\beta_2 = -1$, but these fixed

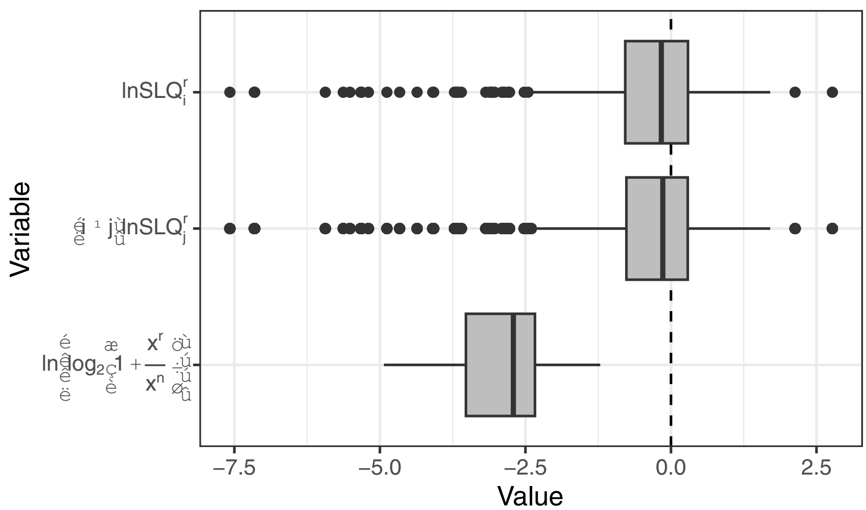

Fig. 5 Distribution of the explanatory variables of the LQ methods

values differ from the parameter estimates obtained via the generalized methods. Furthermore, Figure 5 shows that $\ln \mathrm{SLQ}_i^r$ and $[i \neq j] \ln \mathrm{SLQ}_j^r$ are distributed around zero.

Thus, the significantly negative estimated intercept β_0 in the generalized FLQ method implies that the expected value predicted by the conventional method is artificially inflated because of the constraint of setting $\beta_0 = 0$. Consequently, the larger estimated δ_{ij}^r in the conventional method likely serves to offset this upward bias in the expected value resulting from fixing the intercept at a zero value when it is actually negative.

These findings suggest that the parameter δ_{ij}^r in the conventional FLQ method may partially compensate for the constrained intercept in the regression model. To investigate these findings further, we perform additional comparisons between the conventional method and a modified method that assumes a uniform regional size. Consequently, we replace the regional size term in Eq. (19) with the following:

$$\beta_{ij}^r = \delta_{ij}^r \ln \left[\log_2 \left(1 + \frac{1}{S}\right)\right] \tag{32}$$

where S denotes the total number of regions. This approach results in the following modified regression equation:

$$\ln \hat{\gamma}_{ij}^{rr} = \ln \mathrm{SLQ}_i^r - [i \neq j] \ln \mathrm{SLQ}_j^r + \beta_{ij}^r \tag{33}$$

This formulation demonstrates that assuming a constant regional size is mathematically equivalent to including an intercept or region-specific dummy variables β_{ij}^r in the regression model.

Figures 6 and 7 show the accuracy results for the conventional CILQ, FLQ, and FLQ methods with constant regional size methods. Importantly, in the conventional FLQ (by region) method, δ^r already acts as a regional dummy variable; thus, the accuracy results for this method are indistinguishable from those for the constant regional size method.

Except for $\tilde{U}^S$, the difference in accuracy between the conventional FLQ method and the constant regional size method is less pronounced than the differences observed in Figs. 3 and 4. Furthermore, the accuracy of the conventional FLQ method is greater than

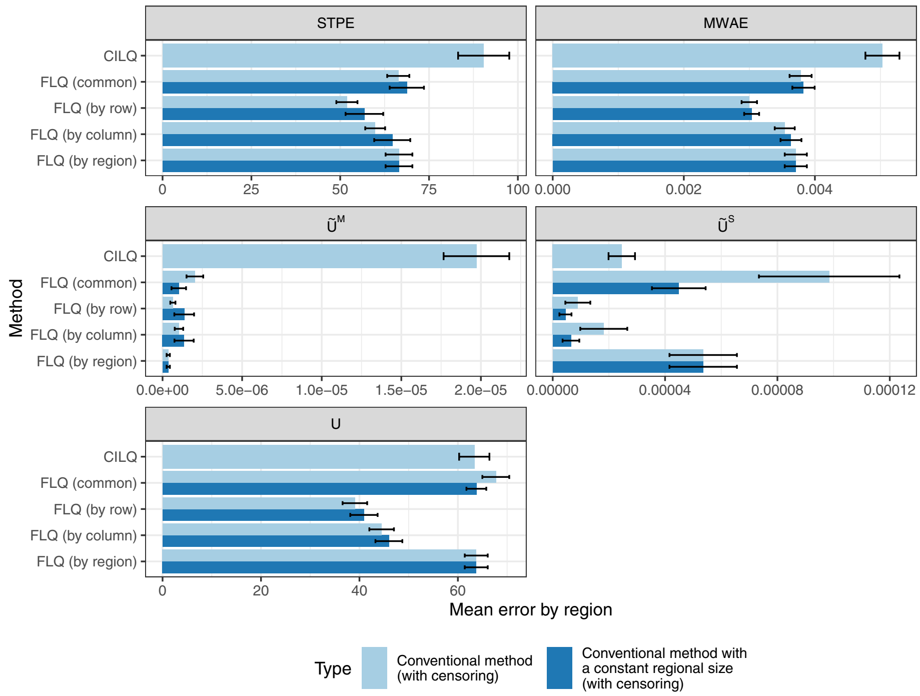

Fig. 6 Comparison of the accuracy of regional input coefficients when the conventional method and regional sizes are fixed

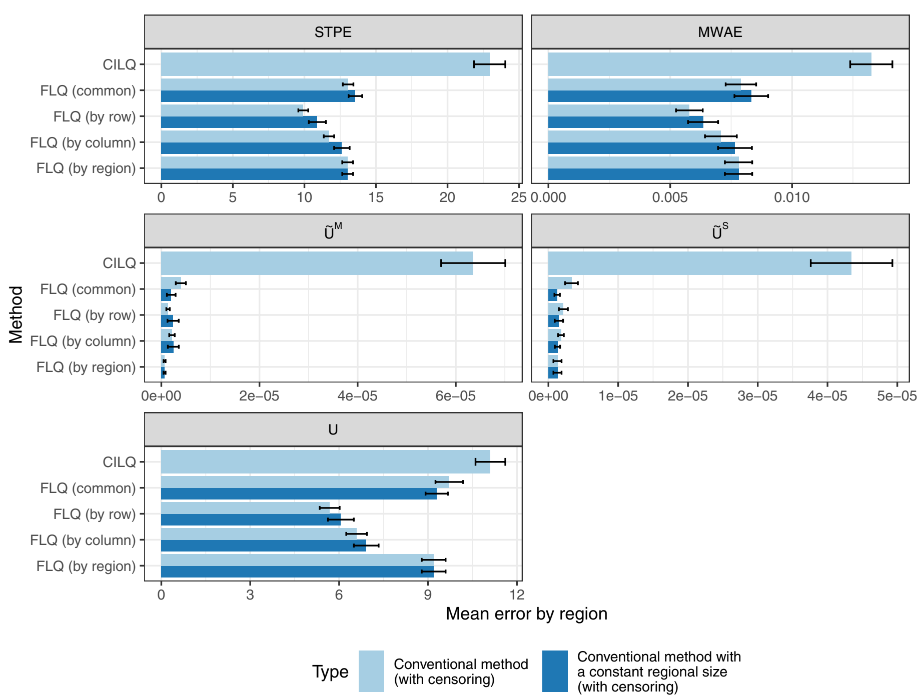

Fig. 7 Comparison of the accuracy of the Leontief inverse when the conventional method and regional sizes are fixed

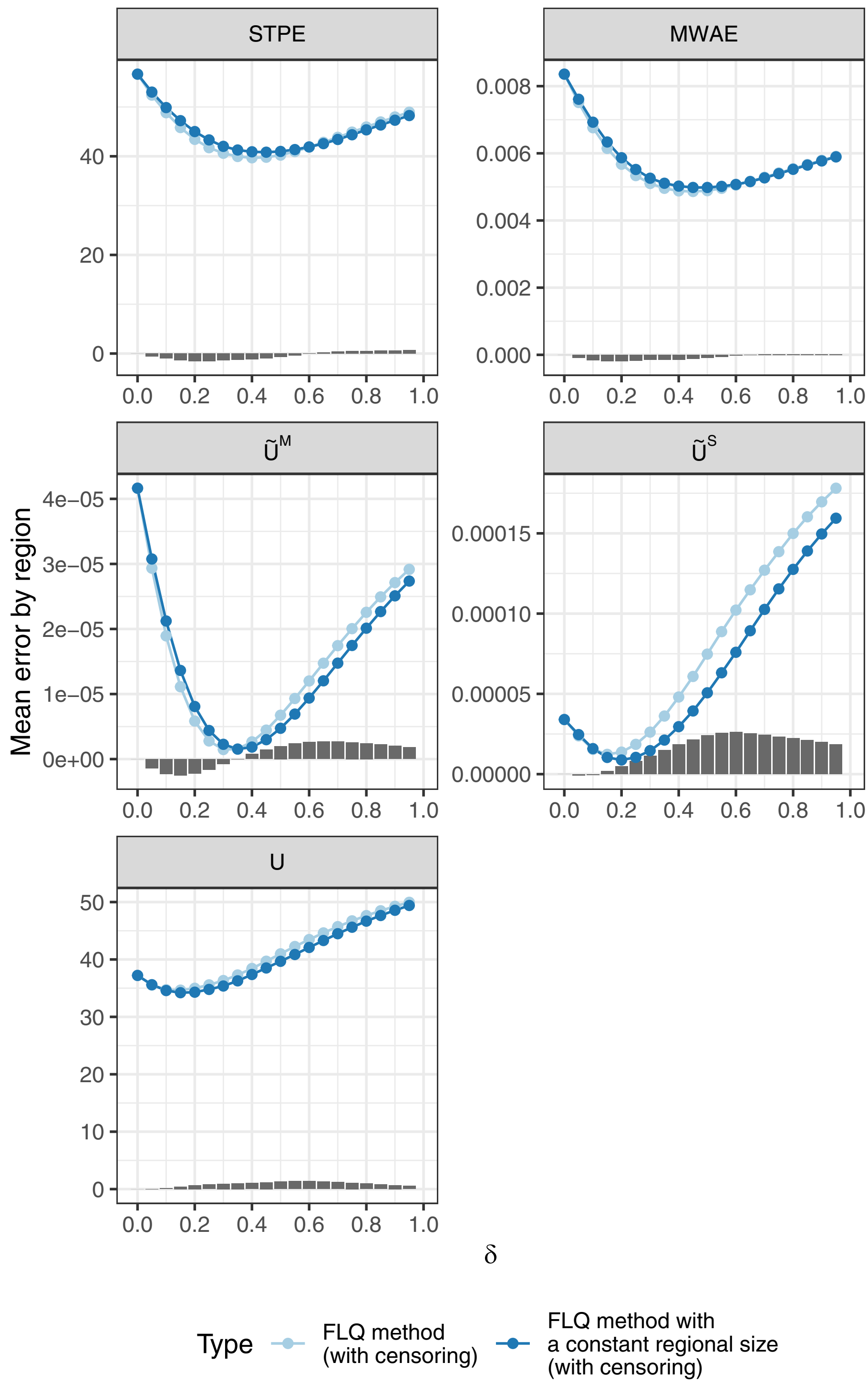

Fig. 8 Changes in accuracy due to δ_{ij}^{r} of the conventional FLQ method and FLQ method with a constant regional size

that of the CILQ method, a trend that is also evident for the FLQ method at constant regional size, and the accuracy gains for the CILQ method when these FLQ variations are used are comparable. Therefore, we cannot definitively conclude that the accuracy improvement in the FLQ methods is solely due to the consideration of regional sizes.

To further assess the impact of the regional size parameter, we examine the change in accuracy by varying δ via the FLQ (common) method. Specifically, we evaluate 20 cases with δ values ranging from 0–0.95 in increments of 0.05. The results are shown in Fig. 8, with the bar chart illustrating the difference in accuracy between the conventional method and the constant regional size method; negative values indicate greater accuracy for the conventional method. As observed in Fig. 8, the change in accuracy resulting from varying δ is similar for different methods, and there is no substantial difference in their accuracy when δ is the same.

Therefore, we conclude that the primary improvement in the accuracy of the FLQ method is not due to the explicit consideration of regional size. These findings suggest that δ_{ij}^{r} does not function primarily as a measure of regional size effects but rather functions more like a constant term or a dummy variable.

5 Conclusions

The LQ method, a nonsurvey-based approach, has gained widespread acceptance, primarily because of its ease of application. However, these methods have been subject to persistent criticism because of their lack of robust theoretical foundations. While extensive empirical research has been conducted to evaluate their performance, the formulation of these methods has historically lacked sufficient statistical verification, relying instead on heuristic assumptions regarding parameter specifications. Therefore, this study focuses on the main LQ methods, namely, the SLQ, CILQ, and FLQ methods, and reexamines their formulations using a generalized regression model.

First, conventional LQ methods fix the constant parameters in the generalized LQ model at zero. However, these parameters are significantly negative in most of the generalized methods, indicating that this fixation is not supported by the data. In addition, the coefficients of the variables introduced by the SLQ method are significantly greater than zero. While these findings contradict the parameter fixation of the conventional method, the observed positive correlation partially supports the SLQ method.

Next, it is often argued that the CILQ method has advantages over the SLQ method because it incorporates a balance between the regional supply and demand sectors. However, our analysis shows that this claim lacks statistical and factual support and that the additional variable introduced by the CILQ method has limited explanatory power.

Finally, the FLQ method, a parametric approach, is often considered one of the more accurate LQ methods. A key feature of the FLQ method is its incorporation of the effect of regional size on import dependence, a feature that is thought to contribute to its improved accuracy. However, our analysis suggests that most of the improvement in accuracy observed with the FLQ method is not due to the explicit consideration of regional size. Instead, the FLQ parameter appears to function more like a constant term or a dummy variable.

In general, our analysis shows that the regression model that generalizes the traditional LQ methods is more accurate than traditional methods and often achieves this with a smaller number of parameters. Notably, the finding that accuracy improves in generalized methods where the intercept is significantly different from zero aligns with the results of the HTLQ method by Flegg et al. (2025). This suggests that explicitly including a constant adjustment term is effective for estimation. Furthermore, the LQ by column has limited explanatory power, suggesting that a row-only approach, such as RPC, may provide a more productive way in which to regionalize than many traditional LQ methods do.

However, directly applying our proposed generalized regression model for parameter estimation contradicts the fundamental data constraints inherent in nonsurvey methods. That is, if complete regional input–output data were available, nonsurvey methods would be rendered redundant. On the other hand, this model is well suited for empirically validating the assumptions of LQ methods, which traditionally rely on heuristic parameter settings. To date, there has been a lack of sufficient statistical verification regarding why specific heuristic methods function effectively. Therefore, future research should aim to clarify the statistical justification for the functional forms employed, utilizing statistical frameworks such as the one proposed in this study.

Finally, this study does not include the effects of censoring, a common practice in LQ methods, in the estimation of the regression model; thus, accounting for these effects remains a challenge. Furthermore, while the regression equation demonstrates the potential for improving the accuracy of LQ methods, a comprehensive economic interpretation of the estimated relationships is not fully developed. Further research to address these limitations and provide a more robust economic basis for estimating regional input–output tables is needed.

Acknowledgements
The author is grateful to Professor Junji Nakazawa of Kochi University for his advice in the writing of this paper.

Author contributions
The author contributed to the design, methods, analysis, and writing of this study and read and approved the final version of the manuscript.

Funding
The Article Processing Charge was covered by the funds of PAPAIOS and JSPS (KAKENHI Grant No. JP21HP2002).

Data availability
The 2015 interregional input–output table for South Korea used in this study is publicly available on the website of the BOK—https://www.bok.or.kr/. The dataset used in this study will also be made available by the corresponding author upon reasonable request.

Declarations

Competing interests
Not applicable.

Received: 2 July 2025 / Revised: 26 January 2026 / Accepted: 3 March 2026
Published online: 26 March 2026

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(3):243–258
Flegg AT, Lamonica GR, Chelli FM, Recchioni MC, Tohmo T (2021) A new approach to modelling the input-output structure of regional economies using non-survey methods. J Econ Struct 10(1):1–31
Flegg AT, Pereira-López X, Sánchez-Chóez N, de la Torre Cuevas F, Tohmo T (2025) An examination of alternative LQ-based approaches to computing regional input–output coefficients. Comput Econ
Flegg AT, Tohmo T (2013) Regional input-output tables and the FLQ formula: a case study of Finland. Reg Stud 47(5):703–721
Flegg AT, Tohmo T (2016) Estimating regional input coefficients and multipliers: the use of FLQ is not a gamble. Reg Stud 50(2):310–325
Flegg AT, Tohmo T (2019) The regionalization of national input-output tables: a study of South Korean regions. Pap Reg Sci 98(2):601–620
Flegg AT, Webber CD (1997) On the appropriate use of location quotients in generating regional input-output tables: reply. Reg Stud 31(8):795–805
Flegg AT, Webber CD (2000) Regional size, regional specialization and the FLQ formula. Reg Stud 34(6):563–569
Flegg AT, Webber CD, Elliott MV (1995) On the appropriate use of location quotients in generating regional input-output tables. Reg Stud 29(6):547–561
Henry MS, Leholm A, Schaible G, Haskins J (1980) A semi-survey approach to building regional input-output models: an application to Western North Dakota. North Cent J Agric Econ 2(1):17

Hewings GJD, Jensen RC (1987) Chapter 8 regional, interregional and multiregional input-output analysis. Handbook of regional and urban economics (Vol. 1, pp. 295–355). Elsevier
Kowalewski J (2015) Regionalization of national Input-Output tables: empirical evidence on the use of the FLQ formula. Reg Stud 49(2):240–250
Kronenberg T (2012) Regional input-output models and the treatment of imports in the European System of Accounts (ESA). Jahrb Reg 32(2):175–191
Kwon H, Choi S-G (2024) An alternative approach to estimating regional input-output tables: the KFLQ method. Ann Reg Sci 72(2):561–578
Lahr ML (1993) A review of the literature supporting the hybrid approach to constructing regional input-output models. Econ Syst Res 5(3):277–293
Lahr ML, Ferreira JP, Többen JR (2020) Intraregional trade shares for goods-producing industries: RPC estimates using EU data. Pap. Reg, Sci
Lamonica GR, Chelli FM (2018) The performance of non-survey techniques for constructing sub-territorial input-output tables. Pap Reg Sci 97(4):1169–1202
Lehtonen O, Tykkyläinen M (2014) Estimating regional input coefficients and multipliers: is the choice of a non-survey technique a gamble? Reg Stud 48(2):382–399
Miller RE, Blair PD (2009) Input-output analysis: foundations and extensions. Cambridge University Press
Pereira-López X, Carrascal-Incera A, Fernández-Fernández M (2020) A bidimensional reformulation of location quotients for generating input-output tables. Spat Econ Anal 15(4):476–493
Round JI (1983) Nonsurvey techniques: a critical review of the theory and the evidence. Int Reg Sci Rev 8(3):189–212
Schaffer WA, Chu K (1969) Nonsurvey techniques for constructing regional interindustry models. Pap Reg Sci 23(1):83–101
Silva JMCS, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88(4):641–658
Stevens BH, Lahr ML, Treyz GI (1988) On the comparative accuracy of RPC estimating techniques, vol 131. Regional Science Research Institute, East Windsor, CT
Stevens BH, Treyz GI, Ehrlich DJ, Bower JR (1983) A new technique for the construction of non-survey regional input-output models. Int Reg Sci Rev 8(3):271–286
Szabó N (2015) Methods for regionalizing input-output tables. Reg Stat 01:44–65
Tohmo T (2025a) Estimating SFLQ-based regional input-output tables for South Korean regions. Natl Account Rev 7(1):125–142
Tohmo T (2025b) The KFLQ revisited: estimating regional input–output tables for regions in South Korea. Ann Reg Sci 74(1)
Zhao X, Choi S-G (2015) On the regionalization of input–output tables with an industry-specific location quotient. Ann Reg Sci 54(3):901–926

Publisher's Note