Regional Input−Output Tables and the FLQ Formula:
A Case Study of Finland

A.T. Flegg* & T. Tohmo**

* Department of Accounting, Economics and Finance, University of the West of England
tony.flegg@uwe.ac.uk

** School of Business and Economics, University of Jyväskylä, Finland
timo.j.tohmo@econ.jyu.fi

Abstract

This paper examines the use of location quotients (LQs) in constructing regional input−output tables. Its focus is on the modified FLQ formula proposed by Flegg and Webber (1997). Using data for 20 Finnish regions, ranging in size from very small to very large, we determine appropriate values for the unknown parameter δ in this formula. We also develop a regression model that can be used to help select an appropriate value for δ. We find that the FLQ yields results far superior to those from standard LQ-based formulae. Our findings should be very helpful to any regional analyst who is contemplating making use of the FLQ formula to generate an initial set of regional input−output coefficients. These coefficients could be used either as part of the RAS procedure or as the non-survey foundations of a hybrid model. We consider possible improvements to the FLQ formula but find that including a regional specialization term in this formula only marginally enhances its performance. On balance, we would recommend using the original FLQ formula.

Key words: Regional input−output tables Finland FLQ formula Location quotients Input coefficients Multipliers Hybrid models

JEL classifications: C67; O18; R15

Revised: 18th December 2010

INTRODUCTION

Regional economies differ from national economies in several respects, most notably in terms of trading relationships. For instance, intermediate inputs purchased from other regions within a given country represent a leakage from the regional economy but are classified as domestic production at the national level. For the regional input–output analyst, the estimation of interregional trade flows presents an awkward problem, which is compounded by the fact that a very limited amount of regional data is normally available.

In principle, the best way of obtaining the data required to construct a regional input–output table would be via a well-designed survey, yet that would be prohibitively expensive, as well as time consuming, in most cases. Consequently, analysts are forced to resort to indirect methods of estimation. A common approach is to attempt to 'regionalize' the national input–output table, so that it corresponds as far as possible to the industrial structure of the region under consideration. Of particular importance is the need to make an adequate allowance for interregional trade, as failure to do so is likely to lead to seriously overstated sectoral multipliers.

A straightforward and inexpensive way of regionalizing a national input–output table is to apply a set of employment-based *location quotients* (LQs). For instance, where *simple* LQs (SLQs) are used, the proportion of regional employment in each supplying sector is divided by the corresponding proportion of national employment in that sector. An SLQ < 1 indicates that a supplying sector is underrepresented in the regional economy and so is held to be unable to meet all of the needs of regional purchasing sectors for that input. In such cases, the national input coefficients for all purchasing sectors are scaled downwards by multiplying them by the SLQ. At the same time, a corresponding allowance for 'imports' from other regions is created. Conversely, where the SLQ ≥ 1, the supplying sector is judged to be able to fulfil all requirements of regional purchasing sectors, so that no adjustment to the national

input coefficients is needed. The estimated regional input coefficients derived via this process can subsequently be refined on the basis of any extra information available.

Unfortunately, the conventional LQs available − most notably, the SLQ and the *cross-industry* LQ (CILQ) − are known to yield greatly overstated regional sectoral multipliers. This occurs because these adjustment formulae tend to take insufficient account of interregional trade and hence are apt to understate regional propensities to import. In an effort to address this problem, Flegg *et al.* (1995) proposed a new employment-based location quotient, the FLQ formula, which took regional size explicitly into account. They posited an inverse relationship between regional size and the propensity to import from other regions. This FLQ formula was subsequently refined by Flegg and Webber (1997). A further refinement was proposed by Flegg and Webber (2000); this aimed to capture the effect of regional specialization on the magnitude of regional input coefficients.

It is worth noting that the potential uses of the FLQ formula go well beyond the mechanical production of a set of regional input coefficients. In particular, we believe that the FLQ is well suited for use as a key part of the hybrid modelling approach. Hybrid models were developed because of dissatisfaction with the inaccuracy of traditional LQ-based adjustments, along with the costs and delays associated with survey-based models.

According to Lahr (1993, p. 277), hybrid models 'combine non-survey techniques for estimating regional [input coefficients] with superior data, which are obtained from experts, surveys and other reliable sources (primary or secondary)'. Lahr goes on to emphasize the importance of using the best possible non-survey methods, so that 'the sectors and/or cells in the resulting [hybrid] model that do not receive superior data are as accurate as possible given the resources available' (1993, p. 278). Moreover, he remarks that 'the accuracy of the non-survey model is even more critical for many advanced hybrid techniques since researchers are likely to use information from the non-survey model to identify the superior data that

[need] to be obtained' (*ibid*.). In response to these points, we would argue that the FLQ offers a cost-effective way of building the non-survey foundations of a hybrid model.[1]

In addition, where the necessary data are available, FLQ-generated coefficients can be used as the initial values in the application of the RAS iterative procedure. This would be preferable, in our opinion, to using unadjusted national coefficients or coefficients generated by the SLQ or CILQ. Our reasoning here is that RAS employs a proportional scaling of the initial set of input coefficients and seeks to minimize differences between these initial coefficients and the final adjusted coefficients.[2] This argument suggests that enhanced results could be obtained by making use of a more realistic set of initial coefficients.

As discussed later, almost all of the evidence published so far has been strongly supportive of the FLQ formula. Even so, for this formula to be a useful addition to the regional analyst's toolbox, it is crucial that more guidance, based upon an examination of a wider range of regions, is made available with regard to the appropriate value(s) of an unknown parameter δ. This parameter and regional size jointly determine the size of the adjustment for interregional trade in the FLQ formula. The primary aim of our study is to offer some guidance on what value of δ to use. We also aim to shed some further light on the possible merits and demerits of the FLQ approach.

Our study makes use of the Finnish survey-based national and regional input−output tables for 1995, published by Statistics Finland (2000). These tables identify 37 separate sectors. We examine data for 20 regions of different size, in order to assess the relative performance of various LQ-based adjustment formulae and to determine appropriate value(s) for the parameter δ. These regions range in size from very small (0.5% of national output) to very large (29.7% of national output).

The rest of the paper is structured as follows. The first section is concerned with the role of LQs in a regional input−output model. The second section examines the properties of the

FLQ and how it differs from other LQs. This is followed by a review of empirical evidence on the performance of the FLQ. We then outline some key characteristics of Finnish regions. In the next three sections, we present our analysis of sectoral output multipliers and input coefficients for these Finnish regions. The fundamental assumption that regions use the same proportion of intermediate inputs as the nation is examined in the penultimate section. The final section contains our conclusions.

THE REGIONAL INPUT–OUTPUT MODEL

At the national level, we can define:

A to be an n × n matrix of interindustry technical coefficients,

f to be an n × 1 vector of final demands,

x to be an n × 1 vector of gross outputs,

I to be an n × n identity matrix,

where $\mathbf{A} = [a_{ij}]$. The simplest version of the input–output model is:

$$\mathbf{x} = \mathbf{Ax} + \mathbf{f} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{f} \tag{1}$$

where $(\mathbf{I} - \mathbf{A})^{-1} = [b_{ij}]$ is the Leontief inverse matrix.[3] The sum of each column of this matrix represents the type I output multiplier for that sector. The problem facing the regional analyst is how to transform the national coefficient matrix, $\mathbf{A} = [a_{ij}]$, into a suitable regional coefficient matrix, $\mathbf{R} = [r_{ij}]$. Herein lies the role of the LQs.

Now consider the formula:

$$r_{ij} = t_{ij} \times a_{ij} \tag{2}$$

where r_{ij} is the regional input coefficient, t_{ij} is the regional *trading coefficient* and a_{ij} is the national input coefficient.[4] r_{ij} measures the amount of regional input *i* needed to produce one unit of regional gross output *j*; it thus excludes any supplies of *i* 'imported' from other regions or obtained from abroad. t_{ij} measures the proportion of regional requirements of input *i* that can be satisfied by firms located within the region; hence, by definition, $0 \leq t_{ij} \leq 1$.

Using LQs, one can estimate the regional input coefficients via the formula:

$$\hat{r}_{ij} = LQ_{ij} \times a_{ij} \qquad (3)$$

where LQ_{ij} is the analyst's preferred location quotient. However, this adjustment is only made in cases where $LQ_{ij} < 1$.

CHOOSING AN LQ

The two most widely used LQs are the SLQ and the CILQ, defined as:

$$SLQ_i \equiv \frac{RE_i/TRE}{NE_i/TNE} \equiv \frac{RE_i}{NE_i} \times \frac{TNE}{TRE} \qquad (4)$$

$$CILQ_{ij} \equiv \frac{SLQ_i}{SLQ_j} \equiv \frac{RE_i/NE_i}{RE_j/NE_j} \qquad (5)$$

where RE_i denotes regional employment (or output) in supplying sector i and NE_i denotes the corresponding national figure. RE_j and NE_j are defined analogously for purchasing sector j. TRE and TNE are the respective regional and national totals. In addition, Round's semi-logarithmic LQ (Round, 1978) is sometimes used. This is defined as:

$$RLQ_{ij} \equiv SLQ_i/[\log_2(1+SLQ_j)] \qquad (6)$$

In evaluating these alternative formulae, it is helpful to refer to the criteria proposed by Round (1978). He suggested that any trading coefficient is likely to be a function of three variables in particular: (1) the relative size of the supplying sector i; (2) the relative size of the purchasing sector j; and (3) the relative size of the region. The first variable is captured here by RE_i/NE_i, the second by RE_j/NE_j and the third by TRE/TNE.

It is evident that the CILQ takes variables (1) and (2) explicitly into consideration, yet disregards (3), whereas the SLQ incorporates (1) and (3) but not (2). However, the SLQ takes account of regional size in a way that we would regard as counterintuitive: for a given RE_i/NE_i, the larger the region, the *larger* the allowance for imports from other regions.

6

Whilst the RLQ allows for all three variables, TRE/TNE enters into the formula in an implicit and seemingly rather strange way.[5] For instance, the effect of applying the logarithmic transformation to SLQ_j instead of SLQ_i is that a bigger allowance for regional imports would be made in a larger region than in a smaller one that was equivalent in all other respects.[6]

Flegg *et al.* (1995) attempted to overcome these problems in their FLQ formula. In its refined form (Flegg and Webber, 1997), the FLQ is defined as:

$$FLQ_{ij} \equiv CILQ_{ij} \times \lambda^* \text{ for } i \neq j \tag{7}$$

$$FLQ_{ij} \equiv SLQ_i \times \lambda^* \text{ for } i = j \tag{8}$$

where:

$$\lambda^* = [\log_2(1 + TRE/TNE)]^\delta \tag{9}$$

As with other LQ-based formulae, the FLQ is constrained to unity.[7]

Two aspects of the FLQ formula are worth emphasizing: its cross-industry foundations and the explicit role attributed to regional size. Thus, with the FLQ, the relative size of the regional purchasing and supplying sectors is taken into account when determining the adjustment for interregional trade, as is the relative size of the region.

The inclusion of the parameter δ in the FLQ formula makes it possible to refine the function $\log_2(1 + TRE/TNE)$ by altering its degree of convexity (see Flegg and Webber, 1997, Figure 2). $0 \leq \delta < 1$; as δ increases, so too does the allowance for interregional imports. $\delta = 0$ represents a special case where $FLQ_{ij} = CILQ_{ij}$.

Another facet of the FLQ formula is worth noting: the use of SLQ_i along the principal diagonal of the adjustment matrix rather than $CILQ_{ii} = 1$. This procedure, first suggested by Smith and Morrison (1974, p. 66), has also been adopted in our calculations of the CILQ. Its aim is to capture the size of industry *i*, along with the fact that much of the intrasectoral trade in a national input−output table becomes interregional trade in a regional table.

However, a possible shortcoming of the FLQ formula was highlighted by McCann and

Dewhurst (1998), who argued that regional specialization may cause a rise in the magnitude of regional input coefficients, possibly causing them to surpass the corresponding national coefficients. In response to this criticism, Flegg and Webber (2000) reformulating their formula by adding a specialization term, thereby giving rise to the following *augmented* FLQ:

$$AFLQ_{ij} \equiv CILQ_{ij} \times \lambda^* \times [\log_2(1 + SLQ_j)] \tag{10}$$

where the specialization term is applied only when $SLQ_j > 1$. The logic behind this refinement is that, other things being equal, increased sectoral specialization should raise the value of SLQ_j and hence raise the value of $AFLQ_{ij}$. This, in turn, would lower the allowance for imports from other regions. This refinement would make sense where the presence of a strong regional purchasing sector encouraged suppliers to locate close to the source of demand, resulting in greater intraregional sourcing of inputs.

EMPIRICAL EVIDENCE

There is abundant evidence illustrating the very poor performance of the SLQ and CILQ. For instance, in their classic study of data for the English city of Peterborough in 1968, Smith and Morrison (1974) used the SLQ and CILQ to estimate type I sectoral output multipliers. They found that the SLQ overstated these multipliers by 17.2% on average (p. 73). The CILQ generated a mean error of 24.9% but this figure was cut to 19.8% when the SLQ was placed along the principal diagonal of the CILQ (*ibid.*). Other relevant studies include Harrigan *et al.* (1980), Harris and Liu (1998), Sawyer and Miller (1983) and Stevens *et al.* (1989).

Flegg *et al.* (1995) carried out the first empirical test of the FLQ formula. Their re-examination of Smith and Morrison's data for Peterborough revealed that the weighted mean error in estimating the type I sectoral output multipliers could be reduced to about 0.3% by using $\delta \approx 0.3$.[8] Even so, one should be cautious in reading too much into this particular result for δ because of the high degree of aggregation used in Smith and Morrison's study (73

national sectors were aggregated into only 19 regional sectors). What is more, the sectors were aggregated prior to applying LQ-based adjustments for regional imports, which is likely to have biased the results (cf. Flegg *et al.*, 1995, p. 557).

Flegg and Webber (2000) used the survey-based input−output tables for the UK in 1990 and for Scotland in 1989 to construct consistent 104-sector coefficient matrices. They then derived alternative estimates of the Scottish input coefficients by using the FLQ, AFLQ, SLQ and CILQ to adjust the UK-wide data.

Flegg and Webber employed the following statistics, along with several others, to assess the relative performance of the alternative LQ-based formulae:

$$\mu_1 = \Sigma_j \, w_j \, \Sigma_i \, (\hat{r}_{ij} - r_{ij}) / n \tag{11}$$

$$\mu_2 = \Sigma_j \, w_j \, \Sigma_i \, |\hat{r}_{ij} - r_{ij}| / n \tag{12}$$

where $\hat{r}_{ij}$ is the LQ-based coefficient, r_{ij} is the survey-based coefficient, $n = 104$ is the number of sectors and w_j is the proportion of employment in sector *j*. μ_1 was clearly positive for the SLQ and CILQ, indicating a general overstatement of the Scottish input coefficients, whereas the FLQ with $\delta \approx 0.15$ yielded $\mu_1 \approx 0$. The FLQ also invariably outperformed the SLQ and CILQ in terms of μ_2, although a value of $\delta > 0.2$ was needed to minimize μ_2 (*ibid.*, Table 4).

An important additional finding to emerge from Flegg and Webber's study was that the AFLQ did not outperform the FLQ. This outcome is somewhat surprising since the AFLQ is the only one of the four alternative formulae to permit upward adjustments of input coefficients and, in this study, r_{ij} exceeded a_{ij} for 5,096 cases out of 10,816 (*ibid.*, p. 567). The high proportion of cases of $r_{ij} > a_{ij}$ may well be the reason why a relatively low optimal value of δ was found in this study for the FLQ.

Tohmo (2004) carried out another examination of the relative performance of the FLQ, SLQ and CILQ. He employed the survey-based input−output table for Finland in 1995 and a corresponding table for one of its regions, Keski-Pohjanmaa. These tables contained the

9

same 37 sectors. The mean error in estimating the type I sectoral output multipliers was 15.1% for the SLQ, 13.1% for the CILQ but only −0.3% for the FLQ (*ibid.*, Table 4).

A novel way of evaluating alternative LQ-based formulae was pursued by Bonfiglio and Chelli (2008). Using a Monte Carlo approach, they randomly generated 1,000 multiregional input−output tables for each of 20 'regions', with 20 sectors in each table. This process produced 400,000 sectoral output multipliers. By aggregating the regional tables, a 'national' table was produced. The various formulae were then applied to this national table in order to produce alternative estimates of multipliers. A big advantage of Bonfiglio and Chelli's approach is that it is capable of establishing results that should be valid in general, rather than being specific to a particular case study.

Bonfiglio and Chelli used the following key statistics in their evaluation:

$$mrd = (1/n) \, \Sigma_i \, \Sigma_j \, \Sigma_k \, (\hat{m}_{ijk} - m_{ijk}) / m_{ijk} \tag{13}$$

$$mrad = (1/n) \, \Sigma_i \, \Sigma_j \, \Sigma_k \, |\hat{m}_{ijk} - m_{ijk}| / m_{ijk} \tag{14}$$

$$\sigma = [(1/n) \, \Sigma_i \, \Sigma_j \, \Sigma_k \, \{(|\hat{m}_{ijk} - m_{ijk}| / m_{ijk}) - mrad\}^2]^{0.5} \tag{15}$$

where m_{ijk} and $\hat{m}_{ijk}$ are the respective true and estimated multipliers for sector *k* in region *j* relating to table *i*, *mrd* is the mean relative distance, *mrad* is the mean relative absolute distance, σ is the corresponding standard deviation and n = 400,000.

The simulations gave values for the *mrd* of 39.5%, 36.1% and 39.2% for the SLQ, CILQ and RLQ, respectively.[9] By contrast, for the FLQ with $\delta = 0.3$, the *mrd* was only 1.4%, which indicates minimal bias. However, this figure hides a substantial amount of offsetting of positive and negative errors. This point is substantiated by the fact that the *mrad* for the FLQ, again with $\delta = 0.3$, was 19.1%. The corresponding values for the SLQ, CILQ and RLQ were 40.3%, 38.0% and 40.2%, respectively. The similarity of the *mrd* and *mrad* statistics for the conventional LQs shows that they almost always overestimated the multipliers. The

10

estimates from the conventional LQs also exhibited much more dispersion: σ was 1.30 for the FLQ (with δ = 0.3) but 1.63 for the SLQ, 1.62 for the CILQ and 1.64 for the RLQ.[10] As regards the AFLQ, the simulations reveal that it performed slightly better than the FLQ.[11]

Bonfiglio and Chelli's findings are consistent with those of other studies in so far as they confirm the superior relative performance of the FLQ.[12] However, all adjustment formulae exhibit unusually high values of the *mrad*. A possible explanation of this phenomenon is that the regional input and import coefficients were randomly generated in the interval 0 to 1, yet the input coefficients in real input−output tables are typically fairly small, except for those on the principal diagonal. The wide range of possible values for the r_{ij} may well have introduced an artificial degree of variation into the randomly generated 'true' coefficients and hence rendered the simulations less accurate than they might otherwise have been.

The studies examined hitherto have all offered evidence supporting the use of the FLQ. A contrary position is taken by Riddington *et al.* (2006), who are highly critical of LQ-based approaches in general and of the FLQ in particular. The authors' brief was to measure the economic impact of water-based tourism at a 'local' level for seven defined areas covering the whole of Scotland, as well as for one specific area in the eastern Highlands (Moray, Badenoch and Strathspey, hereafter MBS) with 2.3% of total Scottish employment. To this end, Riddington *et al.* used gravity models to build input−output tables for each local area. In addition, for the MBS area, they built both survey-based and LQ-based models to measure the impact of angling expenditure.

A key part of the empirical evidence adduced by Riddington *et al.* concerns the values of just two multipliers: the type II output and expenditure multipliers.[13] The output multipliers are of particular interest. Here the survey yielded a multiplier of 1.700, whereas the gravity model, the SLQ and the CILQ gave estimates of 1.766, 1.549 and 1.858, respectively (Riddington *et al.*, 2006, Table 1).[14] The authors remark that they had expected the LQ-based

multipliers to be closer to the Scottish national multiplier of 2.013 (cf. *ibid*., p. 1077).

As regards the FLQ, Riddington *et al.* find that a near-zero value of δ is needed to reproduce the survey-based multipliers.[15] This suggests, of course, that regional size is an irrelevant variable, which seems unlikely in the light of the studies reviewed above. Moreover, there are several other reasons for querying this finding. The first is that it relates to only one sectoral multiplier and we would argue that is unrealistic to expect the FLQ or, indeed, any other technique to produce satisfactory results for every sector in every region. Secondly, the 'parent' table for Scotland that was used to produce the LQ-based tables contained 128 sectors, which were aggregated to only 14. Considerable aggregation bias can occur in such cases, particularly if the LQs are not applied prior to the aggregation (Flegg *et al.*, 1995; Sawyer and Miller, 1983). Riddington *et al.* do not say what procedure they used. Thirdly, as the authors themselves note, the MBS area has a significantly different 'shape' to that of Scotland as a whole; in particular, it has a very limited number of large employers, an industry (whisky) that exports 99% of its output, and very limited services (cf. *ibid*., p. 1076). Indeed, Riddington *et al.* remark that deriving a local table by modifying the Scottish national table makes less sense for the MBS area than for the other, much larger, regions they studied (*ibid*.). This may be one reason why the LQ-based multipliers were smaller than expected, so that a further FLQ-based adjustment for regional size was not required.[16] Finally, as the authors themselves note, their survey had its limitations, so that its use as an accurate 'benchmark' is questionable (cf. *ibid*., p. 1077).

We would argue, therefore, that it is by no means evident that a gravity modelling approach is superior to the FLQ, especially when its complexity, cost and extensive data requirements are borne in mind. Even so, it is worth reiterating that the FLQ only aims to provide a cost-effective way of generating an *initial* set of regional input coefficients. These should always be scrutinized to check for anomalies and, where appropriate, this analysis

should be backed up by surveys of key sectors. The MBS area is a good example of a region where survey-based data could offer a very useful check on the realism of the most important coefficients in the FLQ-based table.

FINNISH REGIONS

Before considering the relative performance of the different LQ-based formulae, it may be helpful to examine the characteristics of the 20 Finnish regions. The location of each region is identified in Figure 1.

Table 1 and Figure 1 near here

Table 1 reveals some marked differences in the characteristics of the regions, most notably in terms of their relative size. Regional size can be measured in several different ways and the first four columns of Table 1 illustrate some of possibilities. The measures are obviously closely related, and the close relationship between the share of output and the share of employees ($r = 0.997$) is reassuring because the regional modeller typically has to use employment data as a proxy for regional output data, which are not normally available.[17]

Uusimaa is by far the largest region. The central government is situated in this region and it is where firms maintain their headquarters. It has a high concentration of public sector jobs. Helsinki, the capital city of Finland, is located in Uusimaa. Electronics manufacturing is a major industry. Uusimaa is also an important node of foreign trade.

At the opposite extreme, Ahvenanmaa is clearly the smallest Finnish region. It is also unusual in being an insular region. It specializes in fishing and in services − especially transport − but it also has some manufacturing, the mainstay of which is the food industry.

The other 18 Finnish regions exhibit considerable diversity in terms of orientation. For instance, Satakunta, Pirkanmaa, Päijät-Häme, Kymenlaakso and Etelä-Karjala form a manufacturing belt with many manufacturing clusters. Also, in Varsinais-Suomi, Keski-

Suomi, Pohjanmaa and Pohjois-Pohjanmaa, the regional industrial structure is characterized by manufacturing, and the most specialized industries are wood, metals, petroleum, machinery, transport equipment, rubber, electronics and paper. By contrast, Itä-Uusimaa has only a few specialist manufacturing industries, most notably petroleum and chemicals. In the Kainuu region, agriculture, forestry and logging, and mining are more prominent than elsewhere. There is some manufacturing activity; this includes wood, along with medical and optical instruments. Kanta-Häme has many manufacturing industries with above-average concentration; these include food, metals, textiles and furniture.

Extraction characterizes Etelä-Savo, Pohjois-Savo, Pohjois-Karjala, Etelä-Pohjanmaa, Keski-Pohjanmaa and Lappi, and manufacturing's share of employment is below the average for Finland. The specialist manufacturing industries in these regions are food, wood, furniture, textiles and leather. Some large-scale industry – largely paper, metals, chemicals, and rubber and plastic products – is also located in these regions. Keski-Pohjanmaa and Etelä-Pohjanmaa also have many small businesses. Lappi is an atypical region insofar as it is sparsely populated and shares borders with three other countries (Norway, Sweden and Russia). Only its southern border is with another Finnish region.

Table 1 also displays some information on the degree of specialization in each region. When measured in terms of Herfindahl's index, *H*, for all industries, it is evident that Ahvenanmaa is the most specialized region in Finland. Using the same criterion, Uusimaa is the next most specialized region. However, there is not a great deal of variation in the value of *H* for the remaining regions. The table also reveals that, for most regions, manufacturing is more highly concentrated than are industries in general.

Another way of attempting to capture the extent of sectoral specialization is by counting the number of sectors that are overrepresented in a regional economy, i.e. those with an SLQ > 1. Six regions stand out as being highly specialized inasmuch as they have 18 or more

sectors (out of a possible 37) with an SLQ > 1. Four of these regions (Etelä-Savo, Pohjois-Savo, Pohjois-Karjala and Etelä-Pohjanmaa) are heavily involved in extraction. By contrast, Kanta-Häme has many manufacturing sectors that show above-average concentration, whereas the focus in the Kainuu region is on agriculture, forestry and logging, and mining.

At the other extreme, there are two regions where only seven sectors have an SLQ > 1 and one region with only four SLQs > 1. Here it is worth noting that Etelä-Karjala and Kymenlaakso form part of the manufacturing belt mentioned above, whereas Itä-Uusimaa has only a few specialist manufacturing industries. Indeed, Table 1 shows that Itä-Uusimaa has a noticeably lower value of *H* in terms of manufacturing than the other two regions.

Of course, merely counting the number of sectors that have an SLQ > 1 does not take any account of the extent to which such sectors are overrepresented in the regional economy, so this approach could be misleading. For instance, both Kainuu and Pohjois-Savo have 20 sectors with an SLQ > 1, yet the largest SLQ in Kainuu is 5.61, well above the maximum value of 2.84 in Pohjois-Savo.

The last column of Table 1 shows the number of cases (out of a maximum of $37^2 = 1369$) where $r_{ij} > a_{ij}$. Such instances are allowed for via the specialization term, $\log_2(1 + SLQ_j)$, in the augmented FLQ formula (10), which is applied only when $SLQ_j > 1$. It is a little surprising that there is not a more obvious positive association between the last two columns. For instance, there are three regions in which 20 sectors have an SLQ > 1, yet these regions yield very different numbers of sectors with $r_{ij} > a_{ij}$. This number ranges from 149 (= 10.9%) for Etelä-Pohjanmaa to 312 (= 22.8%) for Uusimaa, with a mean of 197 (= 14.4%).[18]

REGIONAL OUTPUT MULTIPLIERS

As a first step in our evaluation of alternative LQ-based adjustment formulae, we examine their relative success in estimating output multipliers. In doing so, we use each formula to

regionalize the survey-based national input–output table for 1995 (Statistics Finland, 2000). We then estimate the type I output multiplier for each sector in each region. Finally, we compare these LQ-based estimates with the survey-based estimates for each region (*ibid.*).

Our focus on output multipliers is motivated by their importance in regional analysis, along with the fact that many earlier studies have attempted to derive satisfactory estimates of such multipliers. This multiplier analysis will be complemented later in the paper by a detailed consideration of regional input coefficients.

We now need to consider possible ways of assessing the accuracy of the estimated sectoral multipliers in a given region. The following measures will be examined here:

$$\mu_1 = (100/n)\,\Sigma_j\,(\hat{m}_j - m_j)/m_j \qquad (16)$$

$$\mu_2 = (100/n)\,\Sigma_j\,(\hat{m}_j - m_j)/(m_j - 1) \qquad (17)$$

$$\mu_2^* = 100\,(\overline{\hat{m}} - \overline{m})/(\overline{m} - 1) \qquad (18)$$

$$\mu_3 = 100\,\Sigma_j\,q_j\,(\hat{m}_j - m_j)/m_j \qquad (19)$$

$$\mu_4 = 100\,\sqrt{\frac{\Sigma_j(\hat{m}_j - m_j)^2}{\Sigma_j m_j^{\,2}}} \qquad (20)$$

$$\mu_5 = (1/n)\,\Sigma_j\,|\hat{m}_j - m_j|/m_j \qquad (21)$$

$$sd = [(1/n)\,\Sigma_j\,\{(|\hat{m}_j - m_j|/m_j) - \mu_5\}^2]^{0.5} \qquad (22)$$

where $\hat{m}_j$ is the estimated type I output multiplier for sector *j* (column sum of the LQ-based Leontief inverse matrix) in a given region, m_j is the corresponding survey-based multiplier, q_j is the proportion of regional output produced in sector *j* and n = 37 is the number of sectors.

μ_1 is a statistic that has been used in many earlier studies, thereby facilitating comparisons. Another useful characteristic of μ_1 is its ability to measure bias. This is the main reason why μ_1 is our preferred measure. Even so, it might be argued that the inclusion of the unitary components (direct effects) in the formula tends to exaggerate the apparent

16

precision of the estimated multipliers, as only the indirect effects need to be estimated. Hence some authors (e.g. Lahr, 2001) have opted to use a formula such as μ_2. On the other hand, suppose that the true multiplier is very small, say 1.02, while the estimated multiplier is 1.04. μ_1 and μ_2 would record errors of 2% and 100%, respectively. However, from a practical point of view, we would argue that what is important is that the true multiplier is very small, so that 1.04 should be judged to be a good estimate, as it would be by μ_1.

Unfortunately, it was impossible to compute μ_2 because, in five regions, $m_j = 1$ for the Hunting and fishing sector. It was also very close to unity in four other regions. μ_2^* was formulated in an effort to circumvent this problem. A very different refinement of μ_1, to take account of the relative importance of sectors, has been incorporated in μ_3.

A problem with the measures discussed so far is that an apparently good result could come about as a consequence of some very large positive and negative errors averaging out to zero. Our next two formulae, μ_4 and μ_5, represent alternative ways of overcoming this problem. μ_4 is Theil's index of inequality (Theil *et al.*, 1966), whereas μ_5 is the well-known mean absolute proportional deviation. Finally, sd (standard deviation) has been included to capture dispersion in the absolute proportional errors.

Table 2 near here

Table 2 illustrates the impact on the accuracy of the simulations for individual regions of altering the value of δ in steps of 0.05, using μ_1 as the criterion. What is most striking is that the optimal values of δ are clustered in the interval 0.25 ± 0.05. Indeed, the modal value of $\delta = 0.25$ produces good estimates of sectoral multipliers for seven regions, with an error well below 1%. The mean error of 0.4% (representing a modest overstatement) when $\delta = 0.25$ is also very satisfactory, especially when compared with the outcomes for other values of δ. On the other hand, a δ of 0.25 is clearly too high for the three smallest regions, yet too low for several of the larger regions. Here it is worth noting that, when $\delta = 0.4$, the errors are

negative for all regions except Itä-Uusimaa, which suggests that it would not normally be right to specify such a high value. Conversely, apart from Ahvenanmaa, the errors are invariably positive − many strikingly so − when $\delta = 0.15$, which indicates that a value this low would not be a sensible choice in most circumstances.

Although 0.25 is demonstrably the best single value of δ, it does not generate acceptable results for many regions, so some further analysis is needed to establish the reasons why. One reason has already been alluded to, namely the tendency for the required value of δ to rise with regional size. This tendency is evinced by the fact that the mean error rises from 0.4% to 1.6% once the size of regions is taken into account. Another reason why a δ above or below 0.25 might be required is that certain regions exhibit unusually high or low propensities to import from other regions.

Figure 2 near here

Figure 2 displays alternative LQ-based estimates of each region's propensity to import products produced in other Finnish regions, along with survey-based estimates for comparison.[19] In this diagram, the regions are arrayed from smallest to largest in terms of their share of national output. Whilst the FLQ with $\delta = 0.25$ yields good estimates of these propensities for eight regions and adequate estimates for four others, it is striking how much the propensity to import is overstated in Ahvenanmaa, Kainuu, Pohjois-Karjala and Lappi.[20] This means that a smaller δ is needed to provide satisfactory estimates. On the other hand, the propensity is noticeably understated in Itä-Uusimaa and Varsinais-Suomi, which implies that a $\delta > 0.25$ is required. A final reason why a $\delta \neq 0.25$ might be required is to capture any divergence from the national pattern in a region's mix of intermediate and primary inputs.

In line with the above discussion, we used data for the 20 regions to estimate the following regression equation:

$$\ln \delta = -1.8379 + 0.33195 \ln R + 1.5834 \ln P - 2.8812 \ln I + e \qquad (23)$$

18

where R is regional size measured in terms of output and expressed as a percentage; P is a survey-based estimate of each region's propensity to import from other regions, divided by the mean value of this propensity for all regions; I is a survey-based estimate of each region's average use of intermediate inputs (including inputs imported from other regions), divided by the corresponding national proportion of intermediate inputs; e is a residual.[21]

With $R^2 = 0.915$, the expected signs for all estimated coefficients, and t ratios of 11.66, 6.25 and -3.33 for the three regressors, the regression appears satisfactory. This assessment is bolstered by the fact that it comfortably passes the χ^2 diagnostic tests for heteroscedasticity ($p = 0.591$), functional form ($p = 0.447$) and normality ($p = 0.559$).

To illustrate the potential use of this regression, let us assume that an analyst is examining a region such as Lappi, for which $R = 3.7$, $P = 0.854$ and $I = 0.981$. When compared with other regions of roughly the same size, Lappi stands out in Table 2 as requiring a lower δ. Equation (23) works well in this instance, as it yields a δ of 0.202, which is a little below the optimal value of $\delta \approx 0.209$.[22] It also gives accurate estimates for several other regions, including Ahvenanmaa ($\delta = 0.129$) and Uusimaa ($\delta = 0.375$). Indeed, the absolute error is below 0.01 in eight regions, between 0.01 and 0.02 in four more, and between 0.02 and 0.03 in another six. Unfortunately, in the remaining two regions, Itä-Uusimaa and Etelä-Pohjanmaa, the errors are fairly large (-0.043 and 0.042, respectively).

Nevertheless, would an analyst have the necessary data to make use of our regression equation? Calculating the value of R using either output or employment data would be straightforward. However, getting a reasonable estimate of P would be more challenging, yet a knowledgeable analyst would surely know whether the region under examination had an unusually high propensity to import from other regions (e.g. Itä-Uusimaa, $P = 1.262$) or an unusually low propensity (e.g. Lappi, $P = 0.854$). A well-informed guess would be required. Likewise, the analyst would need to assess whether the region had an atypical mix of

intermediate and primary inputs relative to the national pattern (e.g. Etelä-Savo, $I = 0.931$, or Pohjanmaa, $I = 1.030$). Equation (23) could then be used to derive a figure for δ, which should be better than merely assuming that $\delta = 0.25$. Any other relevant information that was available could be employed to fine tune this estimate. As a default, the analyst could assume that $P = I = 1$. This would entail estimating δ via the equation $\ln \delta = -1.8379 + 0.33195 \ln R$. In the case of Lappi, for example, this would mean using a δ of 0.246 rather than 0.202.

<div align="center">**Table 3 near here**</div>

The discussion so far has focused on our preferred measure μ_1, so we now need to see whether the other criteria listed earlier generate comparable results. At the outset, we should note from Table 3 that μ_2^* also identifies $\delta = 0.25$ as optimal. Moreover, in the case of μ_4 and μ_5, the loss in accuracy from using $\delta = 0.25$ rather than the optimal value is negligible. The fact that $\delta = 0.25$ yields the minimum standard deviation is also an important finding, as it suggests that the FLQ can simultaneously minimize bias and dispersion.[23] Even so, the sectoral weighting underlying μ_3 indicates that the larger sectors typically need a $\delta < 0.25$.

Table 3 reveals that the FLQ − regardless of which measure is used − yields much more accurate results than the SLQ and CILQ. The most obvious explanation of this outcome is that the SLQ and CILQ do not make sufficient downward adjustments to the national input coefficients − to allow for interregional trade − and hence greatly overstate regional propensities to import. The strong upward bias in input coefficients and hence multipliers is also demonstrated in Table 3 by the fact that the mean values of μ_1 and μ_5 are very similar for the SLQ and likewise for the CILQ. This bias is also clearly evident in Figure 2, although it is true that the SLQ does perform well in Ahvenanmaa.

A key reason why the conventional LQs tend to understate imports is that they disregard the problem of *cross-hauling*, which occurs when a sector simultaneously imports and exports the same commodity. This is a chronic problem in small regions that do not represent

<div align="center">20</div>

a functional economic area (Robison and Miller, 1988) but it is also problematic in larger regions (Kronenberg, 2009). The SLQ rules out the possibility of cross-hauling *a priori*. It presupposes that a region will import from other regions, yet not export to them, if $SLQ_i < 1$ but do the opposite if $SLQ_i \geq 1$. The CILQ does not preclude cross-hauling, as some cells in a given row of the adjustment matrix can have $CILQ_{ij} < 1$, while others can have $CILQ_{ij} \geq 1$. Hence imports and exports of commodity *i* can occur simultaneously. The problem here is that the CILQ does not make adequate allowance for cross-hauling, whereas the FLQ attempts to do so by taking regional size into account.[24]

Although it is obviously desirable to have μ_1 as close to zero as possible, so as to eliminate any systematic tendency towards overstatement or understatement, $\mu_1 = 0$ could still arise if some very large positive and negative errors happened to average out to zero. Thus, when assessing accuracy, we do also need to look at measures such as μ_4 and μ_5. When $\delta = 0.25$, the FLQ exhibits an average error of 11.9% for μ_4 and 8.2% for μ_5. The higher figure for μ_4 is due to the fact that, with this measure, simulation errors are squared, which puts greater emphasis on larger errors. For instance, μ_5 would treat an error of ± 0.1 as 10 times larger than an error of ± 0.01, whereas μ_4 would treat the former as 100 times larger than the latter. It is by no means clear that this extreme emphasis on avoiding larger errors is warranted.

However, is an error of 8.2% unacceptably large? This is ultimately up to the analyst to decide, by weighing up the costs and benefits of increasing the degree of accuracy, e.g. the costs of surveying key sectors. It should also be borne in mind that the 8.2% average error arose as a result of using a single figure for δ; this error could be reduced in many cases by using a value tailored to the specific characteristics of the region under consideration.

Many authors have suggested that measures using absolute values are the most satisfactory way of assessing the accuracy of estimated multipliers and coefficients (see, for

example, Bonfiglio and Chelli, 2008; Sawyer and Miller, 1983). What is more, the negligible difference of 0.053 in the mean values of μ_5 for $\delta = 0.25$ and $\delta = 0.3$ may mask some useful information. Hence a set of disaggregated results is presented in Table 4.

Table 4 near here

Table 4 strongly confirms the earlier finding that the optimal δ tends to rise with regional size. Indeed, five of the largest regions require $\delta = 0.35$, whereas two of the smallest require $\delta = 0.15$. Furthermore, compared with Table 3, there has been a shift towards the right in the distribution of results, with 14 of the 20 regions now clustered in the range 0.3 ± 0.05. However, only three regions are located in the centre of this range.

One might ask whether this rightward shift is explicable in terms of the characteristics of the FLQ. To explore this issue, we examined the sectoral distribution of errors for each region when $\mu_1 \approx 0$. We then increased δ by 0.025 to create a new distribution of errors.

A fundamental property of the FLQ is that a rise in the value of δ decreases the extent to which any multipliers are overestimated, while increasing the amount of any underestimation. When μ_1 is used as the criterion, overestimates can be offset by underestimates but this is not possible with μ_5, as all errors of a given size are treated equally. We identified nine regions where the rise in the value of δ decreased the degree of overestimation by considerably more than it increased the degree of underestimation. For these regions, μ_5 typically reached a minimum at a somewhat higher δ than it did for μ_1. On the other hand, there were only two regions where there was a noticeable movement in the opposite direction. The overall shift rightwards in the distribution of results is, therefore, unsurprising.

In the light of the above results, should we accept 0.3 as the best single value of δ? We would suggest not. Our justification for this view is that raising δ from 0.25 to 0.3 would introduce bias that was not previously there, yet yield only a minimal gain in terms of accuracy. We would argue that μ_1 is the best criterion to use in selecting a value of δ,

whereas μ_5 is the best statistic for measuring the resulting degree of accuracy. On this basis, we would recommend using $\delta = 0.25$ as the best single value.

THE FLQ VERSUS THE AFLQ

At this stage, it seems worthwhile to see whether more accurate estimates of multipliers can be obtained by using the augmented FLQ (AFLQ), which is defined in equation (10) and includes a measure of regional specialization. However, when Table 3 was recalculated using the AFLQ, the results were found to be very similar indeed to those for the FLQ, in terms of both accuracy and the pattern of errors. Even so, as Table 5 shows for $\delta = 0.3$, the AFLQ does yield slightly more accurate results on average than the FLQ.

Table 5 near here

How can one explain the minimal rise in accuracy brought about by using the AFLQ? One possible explanation is that, on average across the 20 regions, only 14.4% of sectors have $r_{ij} > a_{ij}$ (see Table 1). Thus a new formula designed to address the problem of $r_{ij} > a_{ij}$ is unlikely to yield dramatically improved results relative to one that does not admit of this possibility. Another possible explanation is that the specialization term $\log_2(1 + SLQ_j)$ in equation (10) is mis-specified in terms of its focus on the purchasing sector *j* rather than on the supplying sector *i*. This argument suggests that we should use $\log_2(1 + SLQ_i)$ instead.

Furthermore, there is a potential problem with using $\log_2(1 + SLQ_j)$ to capture the effects of greater specialization: a rise in SLQ_j will raise the denominator of the CILQ (recall that $CILQ_{ij} \equiv SLQ_i/SLQ_j$), which will tend to dampen the effects of the change in SLQ_j.[25] However, contrary to expectations, Table 5 shows that using SLQ_i rather than SLQ_j does not produce better results.

Our findings for the AFLQ are in line with those of Bonfiglio and Chelli (2008, Table 1), whose Monte Carlo study revealed the AFLQ to be only slightly more accurate than the FLQ.

It is worth noting too the Scottish findings of Flegg and Webber (2000); they found that the AFLQ was no better than the FLQ.

REGIONAL INPUT COEFFICIENTS

The estimation of input coefficients has received much less attention in the literature than the estimation of multipliers. This focus on multipliers can probably be explained by their importance in impact analyses, along with the belief that errors in individual coefficients tend to have little impact on the multipliers – unless these coefficients happen to be large. For instance, Jensen and West (1980) show that more than fifty per cent of the smaller coefficients in an input–output table can be set equal to zero before a ten per cent error appears in the sectoral multipliers. Even so, an examination of coefficients is still worthwhile, as it can highlight any problematic sectors and identify problems such as bias.

The following criteria will be used to assess the accuracy of the estimated coefficients:

$$\gamma_1 = \Sigma_j \Sigma_i (\hat{r}_{ij} - r_{ij}) / (n^2 - z) \tag{24}$$

$$\text{mse} = \Sigma_j \Sigma_i (\hat{r}_{ij} - r_{ij})^2 / (n^2 - z) \tag{25}$$

$$\gamma_2 = \Sigma_j \Sigma_i |\hat{r}_{ij} - r_{ij}| / (n^2 - z) \tag{26}$$

$$\gamma_3 = \Sigma_j \Sigma_i r_{ij} |\hat{r}_{ij} - r_{ij}| / (n \Sigma_i r_{ij}) \tag{27}$$

$$\gamma_4 = 100 \Sigma_j \Sigma_i |\hat{r}_{ij} - r_{ij}| / \Sigma_j \Sigma_i r_{ij} \tag{28}$$

$$\gamma_5 = 100 \sqrt{\frac{\Sigma_j \Sigma_i (\hat{r}_{ij} - r_{ij})^2}{\Sigma_j \Sigma_i r_{ij}^2}} \tag{29}$$

where r_{ij} is the survey-based coefficient in a given region, $\hat{r}_{ij}$ is the corresponding LQ-based coefficient, $n = 37$ is the number of sectors and z is the number of cells for which $r_{ij} = 0$.[26]

γ_1 is a measure of the extent to which a particular LQ-based method tends to overestimate or underestimate the input coefficients. Minimal bias is clearly a key desirable

property but it is also important to consider the dispersion in the values of $(\hat{r}_{ij} - r_{ij})$. Our second formula, the mean squared error (mse), is designed to capture both bias and dispersion. It does so by summing the squared bias and variance of $(\hat{r}_{ij} - r_{ij})$.[27]

The third formula, γ_2, complements γ_1 by providing a check on the possibility that positive and negative errors could offset each other, giving rise to a spuriously accurate set of results. γ_3 is a more radical refinement of γ_1. It takes into account the argument, noted above, that errors in the largest coefficients tend to have the most impact on the estimated multipliers. γ_3 captures the relative size of a particular coefficient via the weight $r_{ij} / \Sigma_i\, r_{ij}$. A similar formula is proposed by Lahr (2001, p. 238), although he focuses on the coefficient matrix as a whole rather than on particular columns within it. For this reason, our version should be more informative as to the source of simulation errors.

γ_4 expresses the mean absolute deviation as a percentage of the mean value of r_{ij} (Sawyer and Miller, 1983). This relative measure has the merit that comparisons can be made with the findings from other studies. This is not true for the measures considered hitherto.

Our final measure is Theil's index of inequality (Theil *et al.*, 1966). A very useful feature of γ_5 is that the mse component of the formula can be decomposed into proportions due to bias, variance and covariance (Stevens *et al.*, 1989).[28] Nonetheless, a demerit of γ_5 should be noted: the use of squared simulation errors means that this statistic can be distorted by extreme values.

Whilst all of the statistics discussed above possess some desirable properties, each measure does have one or more shortcomings. It seems unwise, therefore, to rely on a single statistic as a measure of performance. Instead, by looking at a range of statistics with different properties, one might hope to encompass important characteristics such as bias, variance and the relative size of coefficients. In selecting a subset of possible statistics to discuss, we avoided those that incorporated expressions such as $(\hat{r}_{ij} - r_{ij}) / r_{ij}, |\hat{r}_{ij} - r_{ij}| / r_{ij}$ or

$(\hat{r}_{ij} - r_{ij})^2 / r_{ij}$, which can be inflated by near-zero values of r_{ij}. The χ^2 statistic is a case in point.[29] Indeed, some further shortcomings of the χ^2 statistic are identified by Knudsen and Fotheringham (1986), who also recommend avoiding the use of r^2.[30]

Table 6 near here

If we focus solely on γ_1, then Table 6 shows that $\delta = 0.25$ is incontestably the best single value for estimating coefficients. This outcome is consistent with the result for μ_1 in the case of the multipliers (see Table 3). By contrast, the mse and γ_5 suggest an optimal δ of 0.2. This difference occurs because these statistics place more emphasis than γ_1 upon avoiding large errors, a feature that necessitates a somewhat smaller δ in this instance. The results for γ_2 and γ_4 are also at variance with that for γ_1. Here a higher value, $\delta = 0.3$, is identified as best. This finding can be explained by the use of absolute values. A similar phenomenon was observed with respect to the multipliers; it is also evident in the Scottish findings of Flegg and Webber (2000, Table 4). However, the decreases in the mean values of γ_2 and γ_4 when δ is raised from 0.25 to 0.3 are very small indeed.

γ_3 is out of line with the other statistics in suggesting the use of $\delta = 0.1$, which yields a minimum of 2.747. Evidently, the largest coefficients in a given column of the regional matrix **R** tend to be understated when a higher value of δ is used. In many cases, it is the intrasectoral coefficients that dominate the columns. Although we would not recommend using $\delta = 0.1$ in general, this finding does suggest that it would be worthwhile to consider treating the largest coefficients differently from the remaining coefficients.

The results for γ_4 and γ_5 reveal a high degree of error in estimating the r_{ij}, especially when compared with the multipliers. For instance, when $\delta = 0.25$, the average absolute proportionate error is 8.2% for multipliers (see Table 3) but 69.9% for coefficients. These errors are, however, well below those for the SLQ and CILQ.

That the errors are greater for coefficients than for multipliers is unsurprising: the

elements of the difference matrix, $\mathbf{D} = [\hat{r}_{ij} - r_{ij}]$, are bound to exhibit much more dispersion than the differences in the column sums of the Leontief inverse matrices, $\mathbf{d'} = [\hat{m}_j - m_j]$.[31] A great deal of offsetting of errors occurs when computing multipliers. Therefore, if one's objective is to obtain good estimates of multipliers, this may still be possible even if the coefficients are subject to considerable error.

Table 7 near here

Table 7 displays some helpful information on the source of simulation errors. The covariance component is, in all cases, the predominant source of error, whereas bias is the least important. The covariance component captures the lack of a perfect correlation between the $\hat{r}_{ij}$ and r_{ij} distributions, whereas the variance component arises when these distributions have different standard deviations. The results reveal that the superior performance of the FLQ relative to the conventional LQs is due to its ability to reduce all three sources of error. However, the biggest reduction is in the covariance component of the mse, whereas the smallest is in the bias component. This is rather surprising because the former is thought to be harder to reduce than the latter (cf. Stevens *et al.*, 1989, p. 248). A final point worth noting is that, while the variance component rises along with the value of δ, the covariance component does the opposite.

INTERMEDIATE INPUTS

All LQ-based methods assume identical regional and national technology, i.e. that national and regional firms use the same proportions of different inputs to produce a given commodity. Unfortunately, this assumption cannot be tested directly with Finnish data because each sector's imports from other regions are not disaggregated by type of input. Instead, we shall test the assumption that each region uses the same mix of intermediate and primary inputs as the nation. Primary inputs include 'compensation of employees', 'other

value added' and imports from abroad.[32]

Table 8 displays the results of running the following regression:

$$I_{jr} = \alpha + \beta I_{jn} + \varepsilon_{jr} \qquad (30)$$

where I_{jr} is a survey-based estimate of the proportion of intermediate inputs (including inputs imported from other regions) used by sector j in region r, I_{jn} is the corresponding national proportion of intermediate inputs, and ε_{jr} is a random error term. *Note:* n = 36 for nine regions owing to the exclusion of the Hunting and fishing sector, for which intermediate inputs were either non-existent or extremely close to zero.

While Table 8 shows that the null hypotheses $\alpha = 0$ and $\beta = 1$ cannot be rejected at the 5% level for twelve regions, it also shows that there are statistically significant differences between the regional and national use of intermediate inputs for the remaining eight regions. These differences are most pronounced for Itä-Uusimaa and Pohjois-Karjala.

However, an inspection of the plots of actual versus fitted values revealed that several regressions had obviously been affected by extreme observations. This visual examination was backed up by an appraisal of the standardized residuals. Most notably, in Pohjois-Karjala, the presence of the atypical sectors 6 (Meat and fish) and 7 (Fruit and vegetables; animal and vegetable oils) had the effect of sharply increasing the slope of the regression line.[33] For this region, $I_{6r} = 1.2277$ and $I_{7r} = 1.1742$, whereas $I_{6n} = 0.8365$ and $I_{7n} = 0.8376$. The fact that both I_{6r} and I_{7r} exceed unity can be explained by the existence of subsidies.[34] By contrast, $I_{6r} = 0.5553$ in Itä-Uusimaa, resulting in a relatively flat regression line.[35] The data for sectors 6 and 7 also had a marked impact on the results for several other regions.

Another problematic sector was Hunting and fishing. Although this sector was excluded from the analysis in nine cases, its presence elsewhere did cause some problems in terms of unduly influencing the position of the regression lines in several regions and worsening the

goodness of fit.[36] Likewise, the rather low R^2 for Ahvenanmaa can be attributed to the outlying observations for three sectors: sector 7, as defined above; Electricity, gas and heat supply; and Construction and maintenance of railways, highways and roads.[37]

Nevertheless, apart from a few atypical observations, the plots for most regions exhibited a fairly close relationship approximating a 45° line, which suggests that it is reasonable to assume the same mix of intermediate and primary inputs across regions.

CONCLUSION

Regional analysts rarely have the necessary regional data to build input–output models directly and so are forced to resort to indirect methods of estimation. A straightforward and inexpensive approach is to use regional and national sectoral employment (or output) figures to compute a set of location quotients (LQs). Our focus has been on the FLQ adjustment formula proposed by Flegg and Webber (1997), which takes the relative size of a region explicitly into account. In this way, the FLQ seeks to minimize the strong upward bias in the estimated sectoral multipliers that characterizes conventional LQs such as the simple LQ (SLQ) and cross-industry LQ (CILQ).

A difficulty in applying the FLQ is the need to specify the value of an unknown parameter, δ. Some survey-based evidence on its possible value is presented for Scotland by Flegg and Webber (2000) and for the Italian Marche region by Bonfiglio (2009). However, the generality of results obtained from a single region is always open to question, so our primary aim has been to provide more guidance, drawn from a detailed examination of a wide range of regions of different size, on the appropriate value(s) of δ.

In our case study, we examined data for all 20 Finnish regions, which range in size from 0.5% to 29.7% of national output. We used the Finnish survey-based national and regional input–output tables for 1995, which identify 37 separate sectors, as a benchmark to evaluate

the relative performance of the FLQ and other LQ-based adjustment formulae in estimating input coefficients and type I sectoral output multipliers. We employed a wide range of statistical criteria to assess the results. As expected, the FLQ outperformed the conventional LQs by a wide margin, regardless of which criterion was used. Moreover, an analysis using Theil's index of inequality revealed that the FLQ produced a much closer match between the estimated and survey-based input coefficients. This was true irrespective of whether one measured this match in terms of bias, variance or covariance.

Using the mean proportional difference, μ_1, as the criterion, we found that $\delta = 0.25$ was the best single value for estimating multipliers. This criterion has the desirable property that it minimizes bias. When averaged across regions, the unweighted mean value of μ_1 was 0.4%, compared with -1.9% for $\delta = 0.3$. By contrast, when the results were weighted by size of region, $\delta = 0.3$ gave the best outcome. There was a discernible tendency for the optimal value of δ to rise with regional size. It is also worth noting that the use of absolute values suggested that 0.3 rather than 0.25 might be the best single value of δ. Nevertheless, when averaged across regions, the outcomes for these two values of δ were very similar indeed.

It is interesting that our findings are broadly consistent with those of Bonfiglio and Chelli (2008), who used a Monte Carlo approach to generate 400,000 sectoral output multipliers. Their study indicated an optimal $\delta \approx 0.3$. On the other hand, using Scottish data, Flegg and Webber (2000) found that the FLQ with $\delta \approx 0.15$ yielded $\mu_1 \approx 0$. This relatively low value of δ can probably be explained by the fact that almost half of the Scottish survey-based coefficients were larger than those in the parent table that was being used in the simulation.

Although the FLQ was demonstrably better than the conventional LQs, its performance was not as good as we would have wished. For instance, the unweighted mean absolute proportional error, with $\delta = 0.3$, was 8.1% for multipliers. The weighted mean was 7.6%. As expected, the errors in estimating coefficients were much larger than those for multipliers.

This is exemplified by the fact that the unweighted mean absolute proportional error, again with $\delta = 0.3$, was 69.7% for coefficients. Whether these FLQ-based estimates are sufficiently accurate to be useful in a practical modelling context is a moot point, something that is ultimately up to the potential user to decide. However, most analysts do seem to be far more interested in good estimates of multipliers than they are in good estimates of coefficients.

The above-mentioned results were derived using a single value of δ, yet the evidence shows that the optimal value of δ is unlikely to be the same in different regions. Therefore, using a value of δ that is tailored to the individual characteristics of regions should enhance the performance of the FLQ formula. Here it is worth noting the finding by Bonfiglio (2009), based upon a Monte Carlo analysis, that only one third of regions can be expected to have an optimal δ in the interval 0.3 ± 0.05.[38] Our results indicated a more limited set of possibilities, with most regions clustered in the interval 0.25 ± 0.05. However, when we applied a set of weights to the simulations to capture the relative size of the coefficients in each column of the regional coefficient matrix, we found that the largest coefficients might need a δ of 0.15 or less. It was evident that these coefficients required special scrutiny.

To take account of the tendency for the optimal value of δ to rise with regional size, we developed a regression model to assist in choosing a value of δ. Along with regional size, this model seeks to capture any marked differences in regional propensities to import from other regions or in the mix of intermediate and primary inputs, so it should help to reduce the simulation errors that are bound to occur when applying the FLQ formula.[39] We also re-estimated the multipliers using a modified version of the augmented FLQ formula (AFLQ) proposed by Flegg and Webber (2000). However, despite the fact that the AFLQ takes an additional factor − regional specialization − into account, it gave only slightly better results.

Another issue we explored was how far the mix of intermediate and primary inputs in each region differed from that in Finland as a whole. For most regions, there was a fairly

close relationship, approximating a 45° line, between the regional and national proportions of intermediate inputs. Even so, we did find statistically significant differences between the regional and national use of intermediate inputs in several regions, although these differences were accentuated in many cases by the presence of atypical sectors. These sectors included hunting and fishing; electricity, gas and heat supply; fruit and vegetables; and meat and fish.

Taken as a whole, however, the results indicated that our assumption of a common ratio of intermediate to primary inputs across regions was reasonable. This is important because it suggests that the values of δ obtained here are, to a large extent, indicative of the required adjustments for interregional trade *per se*, with differences between regional and national ratios of intermediate to primary inputs being of much less concern.[40]

Whilst we believe that the results reported in this paper are supportive of the use of the FLQ, it must be emphasized that this formula can only be expected to generate a useful *initial* set of regional input coefficients. These initial coefficients should always be evaluated by the analyst on the basis of informed judgement, surveys of selected industries, etc. Here it would be wise to focus on the larger coefficients, since it is errors in these that have the greatest impact on the multipliers. In addition, Lahr (1993, p. 287) stresses the importance of obtaining superior data for resource-based and 'miscellaneous' sectors. Where aggregation of national sectors is necessary, it is essential that the FLQ be applied prior to aggregation. It is crucial too that any regional peculiarities be taken into account, although the accuracy of the FLQ-based simulations was not affected in any obvious way in our study by a region's location or by its orientation towards manufacturing or extraction.

It is worth emphasizing, finally, that the potential uses of the FLQ formula go well beyond the mechanical production of a set of regional input coefficients. In particular, we would argue that the FLQ offers a cost-effective way of building the non-survey foundations of a hybrid model. Also, where the necessary data are available, FLQ-generated coefficients

can be used as the initial values in the application of the RAS iterative procedure. This should yield more accurate results than could be obtained by using unadjusted national coefficients or coefficients generated by the SLQ or CILQ.[41]

NOTES

1. See Harris and Liu (1998), Lahr (1993, 2001) and West (1990) for more discussion of hybrid models.

2. Cf. Bonfiglio and Chelli (2008, p. 244). In a regional context, the RAS procedure would involve minimizing an expression of the form $D = \Sigma_i \, \Sigma_j \, \tilde{r}_{ij}(\ln\tilde{r}_{ij} - \ln\hat{r}_{ij})$, where $\tilde{r}_{ij}$ is the RAS-adjusted estimate of the regional input coefficient and $\hat{r}_{ij}$ is the LQ-based initial estimate, subject to the constraints of known values of the vectors of sectoral intermediate sales, intermediate purchases and gross output. For a very helpful exposition of the RAS procedure, see Miller and Blair (2009, pp. 313−336). Also see Dietzenbacher and Miller (2009).

3. See Miller and Blair (2009, chapter 2).

4. This type of analysis can also be applied to inter-country trade. See, for example, Oosterhaven and Hoen (1998, pp. 507−509).

5. In a personal communication, Jeffery Round explained that his motivation in developing this formula was to devise a simple expression that allowed for all three factors, yet avoided the need to introduce an additional parameter. In addition, he wished to mediate

between the SLQ and CILQ outcomes, in such a way that the SLQ, CILQ and RLQ all equalled unity when $SLQ_i = SLQ_j = 1$.

6. See Flegg *et al.* (1995, pp. 550−551) and Flegg and Webber (1997, p. 796) for more discussion of Round's formula.

7. The logarithmic transformation in (9) ensures that $\lambda^* \to 1$ as TRE $\to$ TNE.

8. See Flegg *et al.* (1995, Table 9). In its original form, the FLQ incorporated an unknown parameter β, which can be converted into a δ via the formula: $\delta = \beta \times (x/y)$, where R is regional size, $x = \log_{10}[0.30103R / \log_{10}(1 + R)]$ and $y = \log_{10}[3.32193 \log_{10}(1 + R)]$. Given R = 0.0015 for Peterborough, $\beta = 5$ is equivalent to $\delta \approx 0.3$.

9. It should be noted that the SLQ was not used along the principal diagonal of the CILQ.

10. In fact, for the FLQ, σ declined from 1.30 for $\delta = 0.3$ to 0.88 for $\delta = 0.9$. However, we would not recommend using a $\delta > 0.3$ because of the strong negative bias that this would introduce. See Bonfiglio and Chelli (2008, Table 1).

11. The *mrad* for the AFLQ is 18.8% for $\delta = 0.3$ and 18.3% for $\delta = 0.4$. The AFLQ requires a somewhat higher value of δ than the FLQ to achieve a minimum *mrad*. This is in line with expectations, as will be explained later in the paper.

12. Miller and Blair (2009, pp. 361−363) offer an illustrative example using a highly aggregated survey-based Chinese data set with three regions and three sectors. The output multipliers for the FLQ and AFLQ exhibit a marked negative bias. This occurs because the authors used $\delta = 0.3$, which is evidently too high for this data set. Even so, the FLQ and AFLQ still perform better than the SLQ and CILQ. The most accurate results are obtained from the RAS and RPC (regional purchase coefficient) methods, although it should be borne in mind that these methods also have the most exacting data requirements. The FLQ and AFLQ yield very similar results.

13. Type II multipliers take the induced spending by households into account.

14. It is not stated whether the SLQ was used along the principal diagonal of the CILQ.

15. Cf. Riddington *et al.* (2006, p. 1078). For the MBS region, $\beta = 0.25$ entails $\delta = 0.026$.

16. The CILQ-based multiplier is noticeably skewed by the data for the food processing sector. This is an atypical sector in the MBS area since nearly all of its output is exported, while most processed foods are imported from plants throughout Scotland (cf. Riddington *et al.*, 2006, p. 1077). When this dominant sector is excluded, the multipliers become 1.666 (survey), 1.750 (gravity model), 1.499 (SLQ) and 1.659 (CILQ).

17. Output has a correlation of 0.998 with value added and 0.988 with population.

18. For a more detailed discussion of regional specialization and industrial concentration in Finland, see Tohmo (2007, chapters 2−5). Also see Tohmo *et al.* (2006).

19. The domestic import propensity, κ, for a given region was calculated using the formula:

$$\kappa = \Sigma_j \Sigma_i (a_{ij} - \hat{r}_{ij}) / n$$

κ represents the average proportion of gross output that is imported from other Finnish regions.

20. We deem an estimate to be 'good' if it has an absolute error below 0.015 and 'adequate' if this error is in the range 0.015 to 0.025.

21. A log-linear model has some attractive theoretical properties, such as a lognormal error term and the fact that $\delta = 0$ when $R = 0$. It also gave more accurate results than a linear formulation. We attempted − albeit unsuccessfully − to refine our regression by adding a measure of industrial concentration or specialization, $\ln H$, where H is Herfindahl's index (all industries). The lack of statistical significance of $\ln H$ ($t = -0.15$) can probably be ascribed to the limited amount of variation in H across regions (see Table 1). Only Ahvenanmaa and Uusimaa stand out as having noticeably different values of H.

22. An approximate optimal δ was derived for each region by redoing the calculations with smaller steps of 0.025 and then applying linear interpolation.

23. Bonfiglio and Chelli (2008, Table 1) obtained a contrary result. See note 10.

24. Given the reciprocal nature of the CILQ, a maximum of only $\frac{1}{2}(n^2 - n)$ of the national input coefficients will be adjusted downwards. A novel solution to the cross-hauling problem is proposed by Kronenberg (2009). Gerking *et al.* (2001, p. 396) stress the importance of applying LQs at the most disaggregated level possible, in order to minimize bias due to cross-hauling.

25. This is the reason why the AFLQ requires a higher δ than the FLQ.

26. z averaged 150.6 (or 11%) across the 20 regions.

27. $\text{mse} \equiv \{1/(n^2 - z)\}\sum_j\sum_i (\hat{r}_{ij} - r_{ij})^2 \equiv \gamma_1^2 + \{1/(n^2 - z)\}\sum_j\sum_i\{(\hat{r}_{ij} - r_{ij}) - \gamma_1\}^2$

28. $\{1/(n^2 - z)\}\sum_j\sum_i (\hat{r}_{ij} - r_{ij})^2 \equiv \{m(\hat{r}_{ij}) - m(r_{ij})\}^2 + \{sd(\hat{r}_{ij}) - sd(r_{ij})\}^2 + 2(1 - r)\times sd(\hat{r}_{ij})\times sd(r_{ij})$

    where m( ) and sd( ) denote the mean and standard deviation, respectively, and r is the correlation coefficient between $\hat{r}_{ij}$ and r_{ij}. When divided by the mse, the terms on the right-hand side of this identity can be interpreted as the proportions of the mse due to bias, to differences in standard deviations and to the lack of a perfect correlation ($r = 1$) between $\hat{r}_{ij}$ and r_{ij}. Cf. Theil *et al.* (1966, pp. 29–30).

29. Flegg and Webber (2000) attempted to circumvent this problem by excluding cases where $r_{ij} < 0.001$.

30. In fact, Knudsen and Fotheringham (1986) argue that the following statistic is the best way of comparing the performance of either (i) two or more models in replicating the same data set or (ii) a single model in different systems:

$$\gamma_6 = 100\frac{\sqrt{(1/n^2)\sum_j\sum_i (\hat{r}_{ij} - r_{ij})^2}}{(1/n^2)\sum_j\sum_i r_{ij}}$$

This statistic is similar to γ_5 in its use of squared simulation errors, yet there is an important difference between the two statistics: γ_6 expresses the root mean squared error

as a percentage of the mean value of r_{ij}, whereas γ_5 uses r_{ij}^2 as the basis for comparison. This distinction means that:

$$\frac{\gamma_6}{\gamma_5} = n \frac{\sqrt{\sum_j \sum_i r_{ij}^2}}{\sum_j \sum_i r_{ij}}$$

The fact that this ratio depends on n seems arbitrary. Moreover, in our view, it is more logical to relate the squared simulation errors to r_{ij}^2, as in Theil's formula, than to r_{ij}.

31. See Miller and Blair (2009, pp. 324–327) for a numerical example. The detailed results of Sawyer and Miller (1983) provide a very clear illustration of the point that errors in coefficients are likely to be far greater than those in multipliers.

32. 'Other value added' is essentially a measure of profit or surplus. It equals 'value added at basic prices' *minus* 'compensation of employees' *plus* 'subsidies on production' *minus* 'other taxes on production'. For example, for the agricultural sector in Keski-Pohjanmaa, $0.7566 = 0.5341 − 0.0789 + 0.3014 − 0.0000$. *Source*: Statistics Finland (2000), Regional accounts (data for 1995).

33. Omitting these two sectors changed the results to $\hat{\alpha} = −0.064$, $\hat{\beta} = 1.068$ and $R^2 = 0.806$. The null hypotheses $\alpha = 0$ and $\beta = 1$ could not be rejected at the 5% level ($t = −1.74$ and 0.73, respectively). The further exclusion of the outlying sector, Electricity, gas and heat supply, produced $\hat{\alpha} = −0.059$, $\hat{\beta} = 1.036$ and $R^2 = 0.857$ ($t = −1.95$ and 0.47).

34. Product subsidies for sectors 6 and 7 were $−0.4398$ and $−0.2836$, respectively, for Pohjois-Karjala in 1995, compared with $−0.1250$ and $−0.1343$ for Finland.

35. When sector 6 was excluded, $\hat{\alpha} = 0.038$, $\hat{\beta} = 0.862$ and $R^2 = 0.830$ for Itä-Uusimaa.

36. For instance, the exclusion of the Hunting and fishing sector in Pohjanmaa altered the outcome to $\hat{\alpha} = −0.011$, $\hat{\beta} = 1.021$ and $R^2 = 0.874$.

37. Omitting these three atypical sectors changed the outcome for Ahvenanmaa to $\hat{\alpha} = 0.024$,

$\hat{\beta} = 0.900$ and $R^2 = 0.748$.

38. Using survey-based data for the Italian Marche region in 1974, Bonfiglio (2009) computed a δ of 0.66. This region is, however, somewhat unusual, in view of its below-average proportion of intermediate inputs and above-average propensity to import from other regions (*ibid.*, Table 5). Indeed, the author's results show a probability of only 0.008 of getting $0.6 \leq \delta \leq 0.7$ (*ibid.*, Table 1).

39. Bonfiglio (2009) presents an alternative regression model for determining the value of δ. He includes variables to capture regional size (*RSRP*) and the propensity to import from other regions (*PROP*), as we do, but does not include a measure of the divergence between regional and national proportions of intermediate inputs. Unlike us, he finds an inverse relationship between the optimal δ and regional size. He also defines the import propensity differently.

40. Another way of adjusting for differences between regional and national proportions of intermediate inputs would be to apply Round's 'fabrication' adjustment, whereby each column of the national coefficient matrix is scaled prior to applying LQs (Miller and Blair, 2009, pp. 356−357, 361−362; Sawyer and Miller, 1983). However, this adjustment presupposes that $\alpha = 0$ in equation (30). Our preferred approach, which we believe to be both simpler and more comprehensive, is to use the regression model (23) to help determine an appropriate value of δ.

41. This statement is based on the *a priori* arguments presented in the Introduction. We intend to carry out an empirical analysis to substantiate it and to measure the extent of any improvements in accuracy.

REFERENCES

Bonfiglio A. (2009) On the parameterization of techniques for representing regional economic structures, *Economic Systems Research* **21,** 115−127.

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

Dietzenbacher E. and Miller R.E. (2009) RAS-ing the transactions or the coefficients: it makes no difference, *Journal of Regional Science*, **49**, 555−566.

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.

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.

Flegg A. T. and Webber C. D. (2000) Regional size, regional specialization and the FLQ formula, *Regional Studies* **34**, 563−569.

Gerking S., Isserman A., Hamilton W., Pickton T., Smirnov O. and Sorenson D. (2001) Anti-suppressants and the creation and use of non-survey regional input−output models, in Lahr M.L. and Miller R.E. (Eds) *Regional Science Perspectives in Economic Analysis.* North-Holland, Amsterdam.

Harrigan F.J., McGilvray J.W. and McNicoll I.H. (1980) Simulating the structure of a regional economy, *Environment and Planning A* **12**, 927−936.

Harris R. I. D. and Liu A. (1998) Input−output modelling of the urban and regional economy: the importance of external trade, *Regional Studies* **32**, 851−862.

Jensen R. C. and West G. R. (1980) The effect of relative coefficient size on input−output multipliers, *Environment and Planning A* **12**, 659−667.

Kronenberg T. (2009) Construction of regional input−output tables using nonsurvey methods: the role of cross-hauling, *International Regional Science Review* **32**, 40−64.

Knudsen D.C. and Fotheringham A.S. (1986) Matrix comparison, goodness-of-fit, and spatial interaction modeling, *International Regional Science Review* **10,** 127−147.

Lahr M.L. (1993) A review of the literature supporting the hybrid approach to constructing regional input−output models, *Economic Systems Research* **5,** 277−293.

Lahr M.L. (2001) A strategy for producing hybrid regional input−output tables, in Lahr M.L. and Dietzenbacher E. (Eds) *Input−Output Analysis: Frontiers and Extensions*, 211−242. Palgrave, Basingstoke.

McCann P. and Dewhurst J. H. L. (1998) Regional size, industrial location and input−output expenditure coefficients, *Regional Studies* **32**, 435−444.

Miller R. E. and Blair P. D. (2009) *Input−Output Analysis: Foundations and Extensions*, 2nd edition. Cambridge University Press, Cambridge.

Oosterhaven J. and Hoen A.R. (1998) Preferences, technology, trade and real income changes in the European Union: an intercountry decomposition analysis for 1975−1985, *Annals of Regional Science* **32,** 505−524.

Riddington G., Gibson H. and Anderson J. (2006) Comparison of gravity model, survey and location quotient-based local area tables and multipliers, *Regional Studies* **40**, 1069−1081.

Robison M.H. and Miller J.R. (1988) Cross-hauling and nonsurvey input−output models: some lessons from small-area timber economies, *Environment and Planning A* **20**, 1523−1530.

Round J. I. (1978) An interregional input−output approach to the evaluation of nonsurvey methods, *Journal of Regional Science* **18**, 179−194.

Sawyer C.H. and Miller R.E. (1983) Experiments in regionalization of a national input−output table, *Environment and Planning A* **15**, 1501−1520.

Smith P. and Morrison W. I. (1974) *Simulating the Urban Economy: Experiments with Input−Output Techniques*. Pion, London.

Statistics Finland (2000) *Regional Input−Output 1995: tables and compilation methods*. Official Statistics of Finland, Helsinki.

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

Theil H., Beerens G.A.C., DeLeeuw C.G. and Tilanus C.B. (1966), *Applied Economic Forecasting*, 15−43. Elsevier, New York.

Tohmo T. (2004) New developments in the use of location quotients to estimate regional input−output coefficients and multipliers, *Regional Studies* **38**, 43−54.

Tohmo T., Littunen H. and Tanninen H. (2006) Backward and forward linkages, specialization and concentration in Finnish manufacturing in the period 1995−1999, *European Journal of Spatial Development*, April, 1−27, http://www.nordregio.se/EJSD/index.html.

Tohmo T. (2007) *Regional Economic Structures in Finland: Analyses of Location and Regional Economic Impact*, Jyväskylä Studies in Business and Economics 57. University of Jyväskylä, Jyväskylä, Finland.

West G.R. (1990) Regional trade estimation: a hybrid approach, *International Regional Science Review* **13**, 103−118.

Table 1. *Characteristics of Finnish regions in 1995*

Region	Value added (%)	Output (%)	Population (%)	Employees (%)	Herfindahl's index (1995)		SLQ > 1 (number of sectors)	$r_{ij} > a_{ij}$ (number of sectors)
					Manufacturing	All industries		
Ahvenanmaa	0.6	0.5	0.5	0.7	0.189	0.276	14	207
Keski-Pohjanmaa	1.1	1.2	1.4	1.3	0.157	0.088	15	208
Kainuu	1.5	1.3	1.9	1.6	0.162	0.080	20	231
Etelä-Savo	2.5	2.3	3.4	2.9	0.141	0.080	19	216
Itä-Uusimaa	1.7	2.5	1.7	1.6	0.110	0.067	4	155
Pohjois-Karjala	2.6	2.5	3.5	3.0	0.115	0.077	18	210
Etelä-Pohjanmaa	2.8	2.9	3.9	3.5	0.127	0.082	20	149
Kanta-Häme	2.8	3.0	3.2	3.1	0.119	0.072	18	220
Etelä-Karjala	2.9	3.2	2.7	2.5	0.207	0.091	7	154
Päijät-Häme	3.4	3.2	3.9	3.7	0.122	0.075	13	203
Pohjanmaa	3.4	3.5	3.4	3.4	0.114	0.071	12	156
Lappi	3.7	3.7	4.0	3.4	0.173	0.085	15	181
Pohjois-Savo	4.3	3.9	5.1	4.5	0.126	0.085	20	196
Kymenlaakso	3.9	4.4	3.8	3.7	0.230	0.096	7	150
Keski-Suomi	4.6	4.5	5.1	4.7	0.161	0.079	12	208
Satakunta	4.2	5.2	4.8	4.6	0.117	0.069	12	172
Pohjois-Pohjanmaa	6.0	6.0	7.0	6.1	0.168	0.083	13	249
Pirkanmaa	8.1	7.7	8.5	8.2	0.112	0.071	14	167
Varsinais-Suomi	8.4	8.9	8.5	8.9	0.122	0.075	11	204
Uusimaa	31.6	29.7	23.8	28.6	0.134	0.118	15	312
Mean					0.145	0.091	14	197

Source: Statistics Finland (2000), Regional accounts

Table 2. *Mean percentage differences from survey for the FLQ: sectoral output multipliers for 20 Finnish regions in 1995 (measure μ_1)*

	Value of δ					
	0.15	0.2	0.25	0.3	0.35	0.4
Ahvenanmaa	**−1.50**	−4.29	−6.71	−8.72	−10.40	−11.80
Keski-Pohjanmaa	2.74	**−0.11**	−2.80	−5.18	−7.29	−9.24
Kainuu	2.16	**−0.86**	−3.40	−5.60	−7.45	−9.01
Etelä-Savo	4.79	2.01	**−0.34**	−2.42	−4.20	−5.76
Itä-Uusimaa	12.78	9.81	6.97	4.56	2.38	**0.44**
Pohjois-Karjala	5.94	2.51	**−0.47**	−3.02	−5.23	−7.15
Etelä-Pohjanmaa	1.52	**−0.98**	−3.13	−4.96	−6.59	−8.12
Kanta-Häme	6.80	3.99	1.49	**−0.76**	−3.18	−5.32
Etelä-Karjala	5.72	2.95	**0.43**	−1.79	−3.69	−5.31
Päijät-Häme	4.93	2.06	**−0.45**	−2.75	−4.81	−6.64
Pohjanmaa	6.62	2.85	**−0.56**	−3.52	−6.08	−8.23
Lappi	2.59	**0.38**	−1.59	−3.64	−5.61	−7.39
Pohjois-Savo	6.11	2.79	**−0.45**	−3.38	−5.87	−8.05
Kymenlaakso	7.62	4.98	2.47	**0.24**	−1.74	−3.53
Keski-Suomi	6.17	3.42	**1.02**	−1.29	−3.57	−5.60
Satakunta	10.30	7.09	4.27	1.75	**−0.56**	−2.67
Pohjois-Pohjanmaa	5.71	3.02	**0.61**	−1.60	−3.65	−5.60
Pirkanmaa	12.19	3.17	5.82	3.13	**0.72**	−1.44
Varsinais-Suomi	7.53	4.87	2.46	**0.30**	−1.72	−3.59
Uusimaa	3.60	2.84	2.07	1.28	0.50	**−0.26**
Unweighted mean	5.72	2.62	**0.39**	−1.87	−3.90	−5.71
Weighted mean	5.95	3.23	1.59	**−0.30**	−2.05	−3.63

Note: In this and in subsequent tables, minima are shown in bold type.

Table 3. *Assessment of accuracy using different criteria: sectoral output multipliers for 20 Finnish regions in 1995 (unweighted)*

Method	Criterion					
	μ_1	μ_2*	μ_3	μ_4	$\mu_5 \times 100$	sd
SLQ	14.7	59.8	14.2	20.4	15.7	0.1167
CILQ	15.0	63.3	12.3	19.9	16.4	0.1061
FLQ ($\delta = 0.15$)	5.7	26.4	3.4	13.1	9.9	0.0763
FLQ ($\delta = 0.2$)	2.6	10.6	**0.5**	**11.9**	8.5	0.0682
FLQ ($\delta = 0.25$)	**0.4**	**−0.7**	−1.7	11.9	8.2	**0.0673**
FLQ ($\delta = 0.3$)	−1.9	−12.2	−3.7	12.3	**8.1**	0.0680

Table 4. *Mean absolute percentage differences from survey for the FLQ: sectoral output multipliers for 20 Finnish regions in 1995 (measure $\mu_5 \times 100$)*

	Value of δ					
	0.15	0.2	0.25	0.3	0.35	0.4
Ahvenanmaa	**8.28**	8.78	9.71	11.03	12.12	13.18
Keski-Pohjanmaa	9.40	9.18	**9.04**	9.19	10.35	11.54
Kainuu	**8.96**	9.52	10.25	11.02	11.74	12.46
Etelä-Savo	8.87	7.99	**7.99**	8.13	8.95	9.53
Itä-Uusimaa	13.47	11.19	9.10	7.66	6.99	**6.79**
Pohjois-Karjala	10.41	9.27	**8.90**	9.05	9.33	9.74
Etelä-Pohjanmaa	7.98	**7.95**	8.56	9.09	9.81	10.62
Kanta-Häme	12.27	10.84	9.84	9.38	**9.14**	9.59
Etelä-Karjala	8.09	**6.95**	7.09	7.47	7.95	8.88
Päijät-Häme	9.57	8.64	8.18	**7.99**	8.37	9.23
Pohjanmaa	10.95	8.50	7.18	**6.56**	7.28	9.07
Lappi	9.20	**8.72**	8.88	9.32	9.88	10.52
Pohjois-Savo	8.89	7.66	**7.46**	7.90	9.14	10.35
Kymenlaakso	10.29	8.97	8.06	7.81	**7.61**	7.90
Keski-Suomi	9.31	8.08	7.34	**7.12**	7.46	8.55
Satakunta	10.85	8.24	6.34	5.27	**5.10**	5.89
Pohjois-Pohjanmaa	9.09	7.74	**7.08**	7.53	8.23	8.98
Pirkanmaa	12.87	9.94	7.55	6.49	**6.07**	6.17
Varsinais-Suomi	10.90	9.09	7.68	7.18	**7.09**	7.64
Uusimaa	7.87	7.77	7.71	7.66	**7.64**	7.67
Unweighted mean	9.88	8.54	8.20	**8.14**	8.51	9.22
Weighted mean	9.58	8.15	7.81	**7.63**	7.80	8.29

Table 5. *Mean absolute percentage differences from survey for the FLQ and AFLQ: sectoral output multipliers for 20 Finnish regions in 1995 (measure $\mu_5 \times 100$, unweighted)*

Method	Value of δ				
	0.15	0.2	0.25	0.3	0.35
FLQ	9.9	8.5	8.2	8.1	8.5
AFLQ (column-based)	10.7	9.2	8.3	7.9	8.0
AFLQ (row-based)	10.3	9.0	8.3	8.0	8.0

Table 6. *Assessment of accuracy using different criteria: input coefficients for 20 Finnish regions in 1995 (unweighted)*

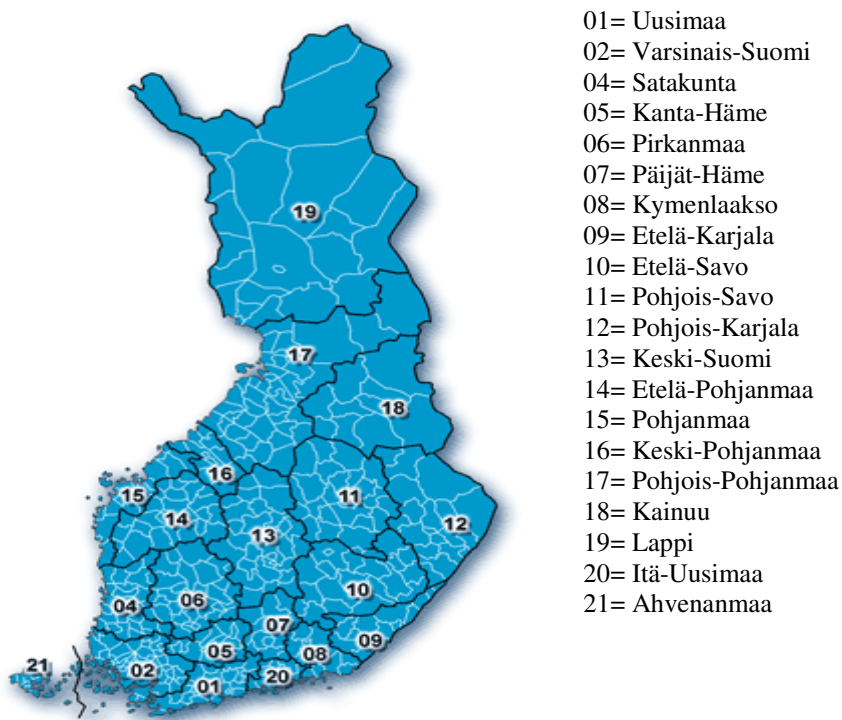| Method	Criterion					
	$\gamma_1 \times 10^3$	mse $\times 10^3$	$\gamma_2 \times 100$	$\gamma_3 \times 100$	γ_4	γ_5
SLQ	2.905	0.331	0.474	2.986	78.9	75.3
CILQ	3.119	0.318	0.510	2.886	85.2	75.2
FLQ ($\delta = 0.15$)	1.293	0.265	0.441	2.770	73.3	67.6
FLQ ($\delta = 0.2$)	0.665	**0.261**	0.428	2.846	71.0	**66.7**
FLQ ($\delta = 0.25$)	**0.062**	0.267	0.422	2.996	69.9	66.9
FLQ ($\delta = 0.3$)	−0.514	0.283	**0.421**	3.184	**69.7**	68.1
FLQ ($\delta = 0.35$)	−1.057	0.302	0.424	3.386	69.9	69.8

Table 7. *Decomposition of mean squared error (mse): input coefficients for 20 Finnish regions in 1995 (unweighted)*

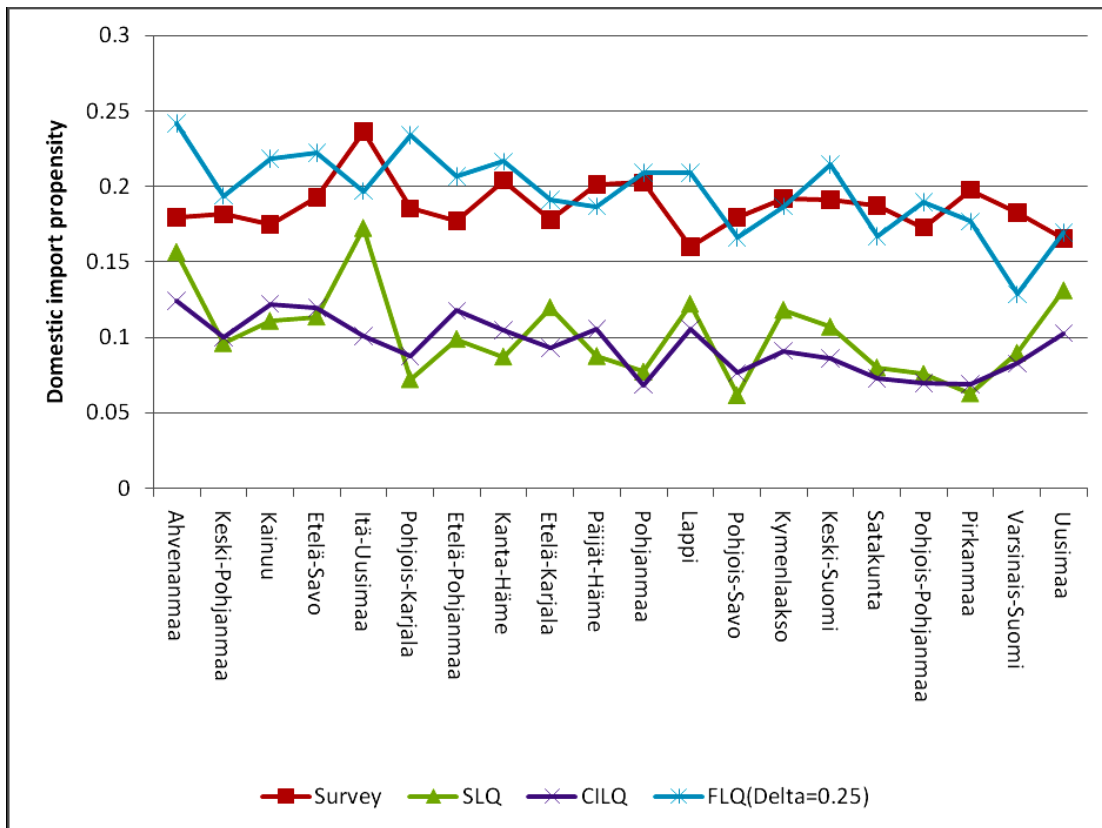| Method	mse $\times 10^3$	Source of error		
		Bias	Variance	Covariance
SLQ	0.3310	0.0100	0.0702	0.2508
CILQ	0.3179	0.0109	0.0651	0.2419
FLQ ($\delta = 0.15$)	0.2652	0.0026	**0.0300**	0.2327
FLQ ($\delta = 0.2$)	**0.2615**	0.0012	0.0346	0.2257
FLQ ($\delta = 0.25$)	0.2674	**0.0008**	0.0489	0.2178
FLQ ($\delta = 0.3$)	0.2827	0.0011	0.0728	0.2088
FLQ ($\delta = 0.35$)	0.3023	0.0021	0.1026	**0.1976**

Table 8. *Regressions of regional on national sums of intermediate input coefficients for 20 Finnish regions in 1995*

	Intercept (|*t*|)	Slope (|*t*|)	n	R^2
Ahvenanmaa	0.055 (1.16)	0.840 (1.45)	37	0.621
Keski-Pohjanmaa	−0.059 (1.67)	1.124 (1.50)	37	0.842
Kainuu	−0.016 (0.40)	0.985 (0.16)	36	0.775
Etelä-Savo	−0.010 (0.26)	0.954 (0.54)	36	0.788
Itä-Uusimaa	0.070 (2.44)	0.769 (3.49)	37	0.793
Pohjois-Karjala	−0.187 (4.44)	1.422 (4.36)	36	0.864
Etelä-Pohjanmaa	−0.066 (2.28)	1.139 (2.10)	36	0.898
Kanta-Häme	−0.049 (1.35)	1.085 (1.01)	37	0.827
Etelä-Karjala	−0.071 (2.60)	1.133 (2.13)	36	0.906
Päijät-Häme	−0.005 (0.14)	1.008 (0.11)	36	0.825
Pohjanmaa	0.052 (1.12)	0.899 (0.93)	37	0.663
Lappi	−0.095 (2.79)	1.220 (2.77)	37	0.871
Pohjois-Savo	−0.074 (2.46)	1.190 (2.73)	36	0.896
Kymenlaakso	0.003 (0.07)	0.955 (0.46)	37	0.732
Keski-Suomi	−0.061 (1.60)	1.140 (1.61)	36	0.835
Satakunta	0.018 (0.61)	0.910 (1.35)	37	0.842
Pohjois-Pohjanmaa	−0.036 (0.98)	1.110 (1.29)	37	0.828
Pirkanmaa	−0.011 (0.46)	1.014 (0.24)	36	0.901
Varsinais-Suomi	0.052 (1.81)	0.849 (2.25)	37	0.821
Uusimaa	0.073 (2.78)	0.843 (2.57)	37	0.846
Unweighted mean	−0.021	1.029		
Weighted mean	0.007	0.975		

Notes: H_0 is $\alpha = 0$ for the intercept and $\beta = 1$ for the slope. The critical value of *t* at the 5% level (two-tailed test) is approximately 2.03.

01= Uusimaa
02= Varsinais-Suomi
04= Satakunta
05= Kanta-Häme
06= Pirkanmaa
07= Päijät-Häme
08= Kymenlaakso
09= Etelä-Karjala
10= Etelä-Savo
11= Pohjois-Savo
12= Pohjois-Karjala
13= Keski-Suomi
14= Etelä-Pohjanmaa
15= Pohjanmaa
16= Keski-Pohjanmaa
17= Pohjois-Pohjanmaa
18= Kainuu
19= Lappi
20= Itä-Uusimaa
21= Ahvenanmaa

Figure 1. Finnish regions. Source: Statistics Finland

Figure 2. Estimates of domestic import propensities produced by the survey and by LQ-based methods