A NOTE ON METHODS OF ESTIMATING REGIONAL INPUT-OUTPUT TABLES: CAN THE FLQ IMPROVE THE RAS ALGORITHM?

Steven Brand

July 2012

Author contact details:

Plymouth University

Drake's Circus

Plymouth

PL4 8AA

sbrand@plymouth.ac.uk

+441752585674

Approx word count excluding references: **3,400**

Abstract

This paper challenges the assertion made by FLEGG and TOHMO, 2011 that the FLQ non-survey method of estimating regional trade will improve the accuracy of STONE's, 1969 RAS algorithm when applied within the context of constructing regional input-output tables. The paper firstly raises a number of conceptual concerns with the assertion before presenting empirical evidence regarding the efficacy of the claim using Finland regional input-output data for 1995. The paper finds evidence to reject the view that the FLQ and other location quotients can provide an improvement to the RAS algorithm.

Keywords : Regional input-output; RAS algorithm; location quotients; the FLQ formula

JEL: C67; O18; R15

Introduction

Improving the methods of estimating intra-national trade for use in regional input-output tables continues to be a topic of debate and innovation amongst regional analysts (see for example KRONENBERG, 2009; FLEGG and TOHMO, 2012). Recently in this journal, FLEGG and TOHMO, 2011 present evidence for 20 regions of Finland to suggest that, amongst the range of conventional methods of modelling regional input-output coefficients, the FLQ approach generates estimates that are closest to empirically estimated coefficients. An intriguing conclusion from their analysis is that the modelled FLQ coefficients can provide a superior basis for implementing the partially survey-based RAS algorithm:

'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.'

The argument is based upon the assumption that estimated coefficients which are closer to their 'observed' values will yield improvements in the efficacy of Stone's bi-proportionate matrix scaling algorithm, known as RAS (STONE, 1969). Despite the strength of the assertion, the authors do not provide empirical evidence to support their claim. The following therefore provides a brief conceptual and empirical analysis of the hypothesis.

The paper is organised as follows. Firstly a brief review of non-survey methods of estimating regional input-output tables is given, followed by a consideration of STONE's, 1969 RAS algorithm within the context of regional input-output trade estimation. The paper then considers whether, conceptually, the FLQ could be of additional use to the RAS algorithm. An empirical analysis of the FLQ's ability to improve the RAS algorithm is then presented and conclusions follow.

Location Quotient Methods of Estimating Regional Input-Output Tables : A Brief Review

Regional trade estimators can be applied to the intermediate matrix of an industry-symmetric input-output table as follows

$$r_{ij} = c_{ij}a_{ij} = \left(1 - m_{ij}\right)a_{ij} \qquad [1]$$

where i and j represent selling and purchasing industries respectively;

r_{ij} are regional input-output coefficients, representing the fraction of *j's* total input that is bought from regional suppliers of i;

a_{ij} is the fraction of j's total input that is bought from national suppliers of i;

c_{ij} , $0 \leq c_{ij} \leq 1$, is the propensity of j to purchase from regional suppliers of i rather than national suppliers (hereby known as *regional consumption propensities*)

m_{ij} , $0 \leq m_{ij} \leq 1$, is the propensity of j to import from national suppliers of i rather than regional suppliers

Non-survey methods of estimating r_{ij} typically make the assumption that the a_{ij} can be taken from the national input-output table. This assumption implies that there is no difference between regional and national industries in their input structure, or in their pattern of international import demand and means that the only task in specifying the regional intermediate matrix is the estimation of regional consumption propensities.

The most commonly applied non-survey method of estimating regional consumption propensities is to use location quotients, of which there are a several varieties. The Simple Location Quotient (SLQ) can be expressed as

$$\hat{c}_{ij} = q_i = \min \quad (s_i/\bar{s}\,, 1) \qquad [2]$$

where q_i is the SLQ of the supplying industry i

s_i is the region's share of national output[1] in industry i

$\bar{s}$ is the region's share of total national output (*i.e.* the weighted average of all s_i)

The Cross-Industry Location Quotient (CILQ) estimates the regional consumption propensity as the ratio of the SLQs between i and j

$$\hat{c}_{ij} = q_{ij} = \min\ (q_i/q_j\,, 1) \qquad [3]$$

A third and, seemingly popular, variant on these specifications is Flegg's Location Quotient, the FLQ (FLEGG *et al.*, 1995). The FLQ is motivated by empirical findings which suggest the application of either [2] or [3] yield estimates of regional consumption propensities that are too high (see for example LAHR, 1993 for a review of studies). The main innovation of the FLQ is the application of a scale parameter λ to standard location quotients, the preferred choice being:

$$\hat{c}_{ij} = \min\ (\lambda q_i, 1) \text{ when } i = j, \qquad [4a]$$

$$\hat{c}_{ij} = \min\ (\lambda q_{ij}, 1) \text{ otherwise} \qquad [4b]$$

Where the scale parameter, λ tends to unity with increasing regional size as follows:

$$\lambda = \log_2(1 + \bar{s})^{\delta} \qquad [4c]$$

The term δ, $0 \leq \delta < 1$is a parameter which determines the rate of increase of λ with respect to regional share of national output, with a smaller value implying a greater allowance for regional import propensity.

RAS as an Estimator of Regional Trade

Whilst location quotient methods are commonly implemented without use of empirically derived information on regional transactions, hybrid methods use an incomplete set of empirical estimates to form the regional input-output table (WEST, 1990). One such method is an application of the bi-proportional matrix scaling algorithm known as RAS and proposed by

STONE, 1969. Its application as a regional trade estimation methodology can be outlined as follows.

Assume identical national and regional production functions[2] so that the national matrix of technical coefficients **A** can be formed into a matrix **T** showing regional industry purchases from the national market by

$$\mathbf{T} = \mathbf{A}\hat{\mathbf{x}} \qquad [5]$$

Where **x** is a vector of regional industry output and ^ denotes a diagonalised vector

Matrix **T** has elements t_{ij}, an *i* dimensional intermediate column vector with elements $t_i = \sum_j t_{ij}$ and a *j* dimensional intermediate row vector with elements, $t_j = \sum_i t_{ij}$. Define the total value of regional transactions purchased from the national market as *t* given, for example by $\sum_i t_i$.

The regional application of the RAS algorithm seeks to estimate **R**, the unknown transactions between regional industries, elements r_{ij}, from **T** knowing only the observed intermediate row and column vectors of **R**, elements r_i and r_j, which are formed in the same way as t_i and t_j in **T**. The total value of intra-regional industry transactions, *r* is therefore known, e.g. $r = \sum_i r_i$

In the first step of the algorithm, a column vector **z** is formed with elements r_i/t_i, *i.e.* the propensities to purchase each industry's output from the regional rather than national market, and this is applied to **T** as a multiplicative scalar across its rows:

$$\mathbf{R}_0^* = \hat{\mathbf{z}}\mathbf{T} \qquad [6]$$

i.e. so that the intermediate column vector of $\mathbf{R}^*_0$ now has elements $r_{0i}^* = r_i$.

In the second step, a row vector **s** is formed with elements r_j/r_{0j}^* where the r_{0j}^* are formed from the intermediate row vector of $\mathbf{R}^*_0$; the columns of $\mathbf{R}^*_0$ are then scaled to correspond to the r_j as follows:

$$\mathbf{R}_1^* = \mathbf{R}_0^*\hat{\mathbf{s}} \qquad [7]$$

The first and second steps of the algorithm (*i.e.* equations [6] and [7]) are then repeated, **z** being formed with elements r_i/r^*_{1i} and applied to $\mathbf{R}^*_1$ forming $\mathbf{R}^*_2$; **s** is then correspondingly formed from r_j and r^*_{2j} etc. This process is iterated until both **z** and **s** approach unit vectors, the resulting matrix $\mathbf{R}^*_n$ being the final estimate of **R**[3].

The FLQ and RAS

One of the central conclusions of FLEGG and TOHMO's, 2011 analysis is the idea that since the FLQ yields estimates of regional input output coefficients (or transactions) that are closer to empirically based estimates relative to unadjusted national coefficients (or transactions), the FLQ is likely to improve the performance of the RAS algorithm. This hypothesis is now examined conceptually.

Essentially, the innovation that the FLQ introduces to conventional location quotient-based regional trade estimation is a general scale parameter λ, $0<\lambda<1$ which is reflective of regional size. In non-survey applications of the FLQ, the parameter λ is unknown and has to be estimated. However, any analyst with the regional RAS dataset clearly already has an FLQ-like scale parameter prior to the application of **z** and **s**: it is given by the region's average propensity to purchase regional output from the national market, *i.e.*

$$\lambda = r/t \qquad [8]$$

Hence the RAS analyst would have little call for the non-survey specification of λ by equation [4c]. Indeed, with **z** and **s** providing a total of i^2 *industry-specific* scalars, λ should be redundant. For example, applying λ to **T**, λ**T**, the elements of **z** become

$$z_i = r_i/t_i\lambda \qquad [9]$$

and λ immediately cancels out of the algorithm with the application of equation [6].

Suppose the RAS analyst had applied a column vector of simple location quotients, **q** [$0<q_i\leq 1$], to **T** as a first estimation step.

$$\mathbf{T}^* = \hat{\mathbf{q}}\mathbf{T} \quad [10]$$

and subsequently believed that an FLQ-type general scalar was desirable, again they would surely want to scale $\mathbf{T}^*$ to the known value of intra-regional purchases r, *i.e.* they would estimate

$$\lambda = r/t^* \quad [11]$$

rather than use the estimator of equation [4c]. Either way however, the analyst would quickly find that the application of **q** and the matrix scalar λ was redundant in RAS since the elements of the vector **z** become

$$z_i = r_i/t_i q_i \lambda \quad [12]$$

and $q_i\lambda$ immediately cancels out of the scaled matrix $\lambda\mathbf{T}^*$ in the first step of the algorithm, equation [6].

One could prevent the cancellation of q_i (but not λ) by starting the RAS algorithm from the row vector **s**, but that very much highlights the dubious logic of combining a scaled simple location quotient with a RAS dataset: the scalar λ is obsolete and since the location quotients **q** are a proxy for the empirical regional consumption propensities, **z**, why bother using **q**, the 'wrong' column scaling vector, when the optimal vector **z** is known and available?

In the case of the cross-industry location quotient or its variant, with q_i on the principal diagonal, whilst the FLQ-like scalar λ (= r/t^*) cancels out of the matrix immediately, the quotients do not. However, the empirical evidence regarding the performance of the cross-industry specification (*e.g.* SMITH and MORRISON, 1974) suggests it offers no particular improvement in performance as a trade estimator in relation to other location quotients and hence its value to the analyst in possession of a regional RAS dataset would have to be questionable.

The analysis so far has centred upon a 'FLQ-like' scale parameter applied as a scalar to a matrix that has been pre-adjusted by location quotients, *i.e.* for a location quotient, q, as $\min(q,1)\lambda$.

However the conventional FLQ applies λ as $\min(q\lambda,1)$, *i.e.* as a scalar to the location quotient, which may subsequently be constrained to unity[4]. In this case, the scalar λ does not cancel out of a row where $q_{ij}\lambda>1$ for any j for example. Whilst it seems doubtful that this relatively minor technical detail could do much to improve the application of RAS, a brief empirical test of the main possibilities discussed in this section is presented.

An Empirical Test of the Efficacy of the FLQ and Location Quotients in Relation to the Performance of RAS

Data and Method

The data for the empirical test are the publically available Finland regional input-output data for 1995 as described by FLEGG and TOHMO (2011). Specifically the data set contains a 37-industry-symmetric table of intra-regional transactions and primary inputs, including an aggregate estimate of regional imports, for each of 20 regions. The corresponding national table is also available.

The empirical test aims to consider whether the RAS procedure is improved by the following cases:

1. Applying the FLQ prior to RAS using the value of λ as would conventionally be estimated by non-survey methods, *i.e.* using FLEGG and TOHMO's (2011) suggested general value of δ= 0.25, applied $\min(q_{ij}\lambda,1)$ for $i\neq j$ and $\min(q_i\lambda,1)$ otherwise, such that $q_{ij}\lambda$ does not cancel out of the process.
2. Applying the FLQ prior to RAS using the value of λ available from the regional RAS data, $\lambda=r/t$, using the conventional FLQ constraint $\min(q_{ij}\lambda,1)$ for $i\neq j$ and $\min(q_i\lambda,1)$ otherwise, *i.e.* such that $q_{ij}\lambda$ does not cancel out of the process[5].
3. Applying the cross-industry specification prior to RAS with the principal diagonal replaced with the simple location quotient. This is essentially the case where λ is applied

min(q,1)λ and λ cancels out of the process but the location quotient adjustments do not cancel out. This is subsequently referred to as the 'modified CILQ'.

Each is benchmarked against the application of the conventional RAS procedure, *i.e.* from the values of the national matrix with no pre-use of location quotients, referred to subsequently as case 4.

The FLEGG and TOHOMO, 2011 analysis largely concentrates on the accuracy of Type I multipliers and presents a considerable range of distance measures in order to form a view regarding relative performance of methods. Type I multipliers are overwhelmingly determined by the intermediate row sum vector (BURFORD and KATZ, 1981) which is known in each case 1-4 and therefore will not generate substantive differences between estimation methods. Hence, this analysis is presented in terms of the accuracy of estimated transactions and coefficients using a simple measure of mean absolute difference between estimated and empirical values as an illustrative guide to relative estimation accuracy.

It should be noted that the test cannot be restricted to the accuracy of regional trade estimation alone because the regional technical coefficient matrix is not provided within the dataset[6]. Since this assumption is common to all techniques this is not considered too distortive to the analysis. The national technical matrix, transformed into regional transactions as per equation [5], is therefore used to generate cases 1-3 and the benchmark case.

Results

Firstly it is worth giving some consideration as to whether the FLQ's scale parameter λ, calibrated using δ= 0.25 relates to the empirically estimated average regional consumption propensity. Given the construction of equations [4a,b], one would expect λ to at least proxy average regional consumption propensity, and hence a positive correlation should be anticipated.

The structure of the Finland regional economy is slightly unusual in that it has one very large region, representing nearly 30% of national output. The remaining 19 regions fall within the range 0.5%-9%, arguably covering a range of regional size more typically encountered by analysts (this range broadly covers NUTS1-NUTS3 regions in the UK, for example).

With δ= 0.25 the FLQ seems to pick up the broad difference between the very large region: average regional consumption propensity = 70%, FLQ λ= 78% and the smaller set of regions: unweighted mean of average regional consumption propensities = 52%, FLQ unweighted average λ = 46%. However, if one excludes the large outlying region and instead focus upon the association between the FLQ and the average regional consumption propensity over the range of regional sizes normally encountered by practitioners, the correlation coefficient is slightly negative but not significantly different from zero, ρ=-0.21 (p=0.37). This result suggests that the FLQ scalar is not a particularly useful general predictor of average regional consumption propensity.

Figure 1 illustrates the mean absolute difference between RAS-estimated and empirically observed transactions, expressed as a percentage of the observed average value of regional intermediation (*i.e.* r for each region) for each of cases 1-3 and the benchmark. The figure shows quite unambiguously that the application of any location quotient prior to application of the RAS algorithm results in a poorer set of estimated regional transactions compared to the use of unadjusted national values. This is the case in 19 out of 20 regions, with 1 region showing a slightly improved average mean absolute difference over the benchmark for the conventional FLQ and the modified CILQ. Of the location quotients, the modified CILQ performs 'least worst', yielding a lower estimation error than the conventional FLQ in 16 out of 20 regions. The modified CILQ average mean absolute difference was on average 3.3% higher than the 26.7% mean absolute difference achieved by the conventional RAS. The conventional FLQ and the survey-based FLQ performed similarly: the conventional FLQ was 4.2% higher than the RAS benchmark and the FLQ using the survey-based estimate of λ 4.6% higher than the benchmark[7].

FIGURE 1 AROUND HERE

Figure 2 shows the result for input-output coefficients. In each case the mean absolute differences are slightly higher than for transactions, for the benchmark case, a 29% mean absolute difference versus 26.7%; however the relative performance of estimators remains unchanged on the previous analysis.

FIGURE 2 AROUND HERE

Conclusion

In conclusion, this paper has argued that FLEGG and TOHMO's, 2011 assertion that the FLQ can improve the estimation performance of the RAS procedure is questionable conceptually. Moreover, it has demonstrated empirically that, for the 20 regions of Finland's economy, the use of an unadjusted national transactions matrix as the starting point for RAS yields generally more accurate estimates of regional transactions than matrices pre-adjusted by location-quotients, including the FLQ. The FLEGG and TOHMO hypothesis regarding the efficacy of the FLQ to RAS is therefore rejected.

However, this is not to say that the conventional RAS procedure cannot be improved upon. Indeed, it has long been established that the more is known about the contents of the target matrix, in this case **R**, such as known values of individual regional transactions, or subsets within **R** *e.g.* regional purchases from a broad group of industries such as manufacture *etc.* the better the performance of RAS should be (see for example ISRAILEVICH, 1986; LAHR and DE MESNARD, 2004). However, it is difficult to see how a location quotient calculated from industry employment or gross output could deliver that kind of information – information that is somehow richer than that embodied within the intermediate row and column sums of regional transactions.

References

Burford, R.L. and Katz, J.L. (1981) A method for estimation of input-output-type output multipliers when no I-O table exists, *Journal of Regional Science*, **21**, 151-161

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., Tohmo, T (2012) A Comment on Tobias Kronenberg's ''Construction of Regional Input-Output Tables Using Nonsurvey Methods: The Role of Cross-Hauling'', International Regional Science Review, 0160017612446371

Israilevich, P. R. (1986) *Biproportional Forecasting of Input–Output Tables*, Unpublished PhD dissertation, Philadelphia, University of Pennsylvania.

Kronenberg, T. (2009). 'Construction of Regional Input-Output Tables Using Nonsurvey Methods: The Role of Cross-Hauling', International Regional Science Review, **32**(1), 40-64

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. and de Mesnard, L. (2004), Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis, *Economic Systems Research*, **16** (2), 115-134

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

West, G (1990) Regional Trade Estimation: A Hybrid Approach, *International Regional Science Review*, **13**, 103-118

Figure 1: Mean Absolute Difference between RAS-Estimated and Empirically Estimated Transactions for Regions of Finland, 1995

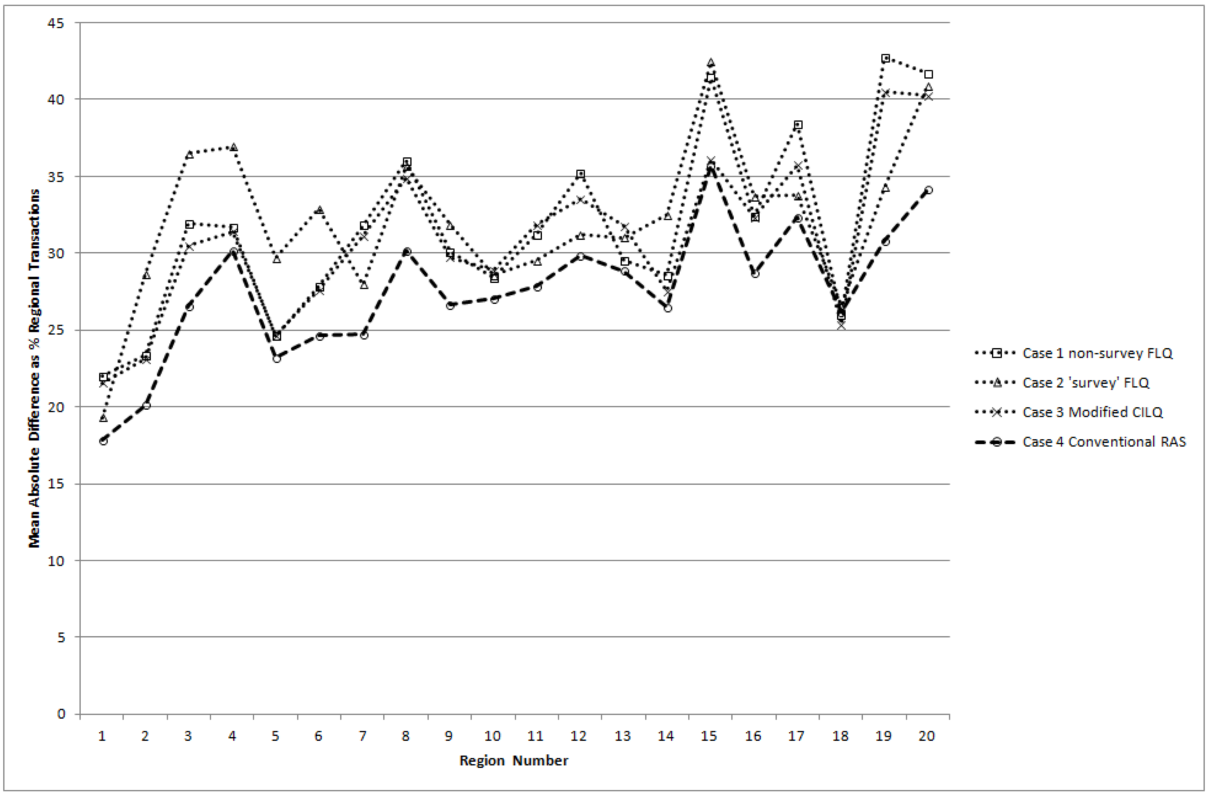
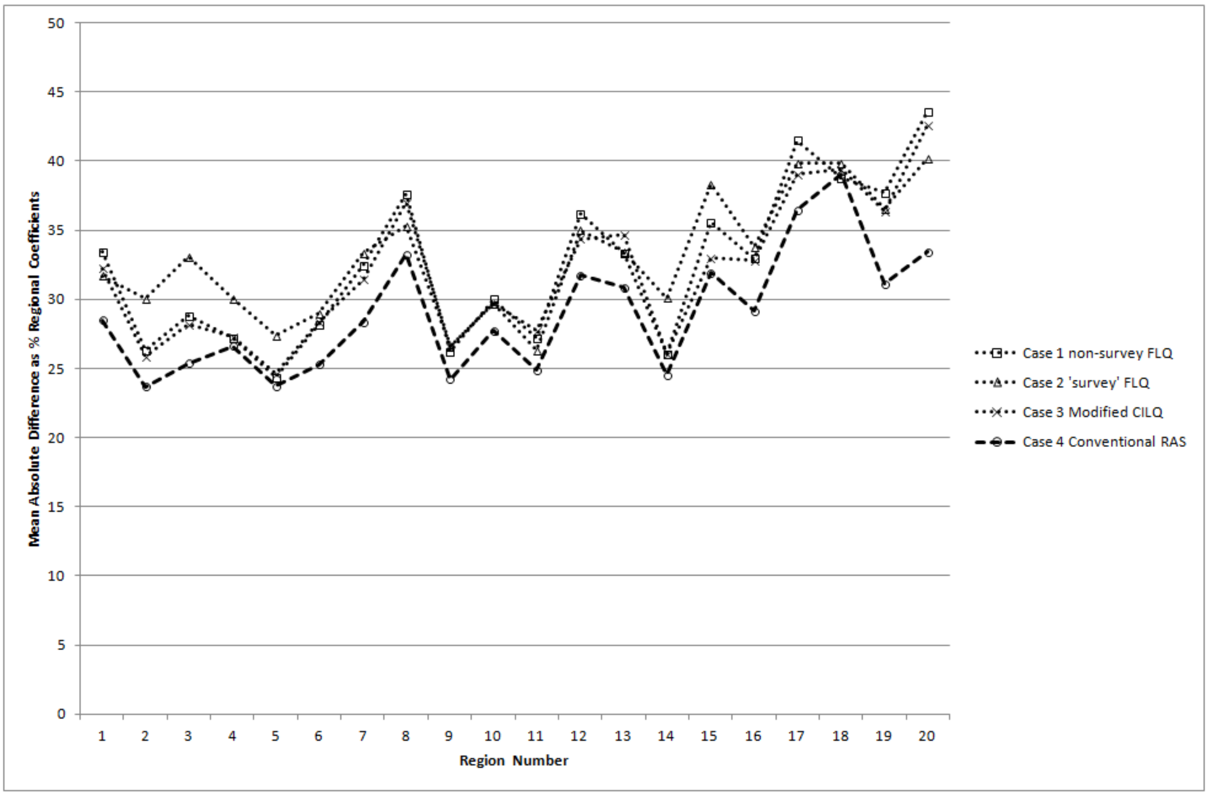


Figure 2: Mean Absolute Difference between RAS-Estimated and Empirically Estimated Coefficients for Regions of Finland, 1995



[1] In the absence of output data, employment may be used to calculate regional shares. It is assumed throughout the paper that output data are available.

[2] This assumption is not essential to the RAS procedure but however simplifies the problem and its exposition to one of estimating regional trade, which is the subject under consideration, rather than a 'mixed' problem of estimating differences in industry production functions and regional trade.

[3] It should be noted that, when used as an estimator of regional trade, the constraint $r_{ij} \leq t_{ij}$ for all i , j should be applied throughout the algorithm.

[4] The application of the non-conventional FLQ formula min(q,1)λ may actually offer more appeal to the non-survey analyst than min(qλ,1) for a number of reasons. Firstly it breaks the adjustment process into two distinct stages: one where the location quotient adjustment is made; and a second where the analyst can think more clearly about how the matrix should be subsequently scaled *e.g.* enabling the analyst to scale the table to satisfy a preconceived idea of average regional import propensity. An analyst with a preconceived, perhaps survey-based, estimate of average regional import propensity would find it more difficult to realise that idea when applying min(qλ,1). The second reason is that, when applied to the Finland data, min(q,1)λ seems to yield some general, though marginal, estimation improvement in relation to the conventional min(qλ,1). The results are not presented here as they digress from the central theme of this paper. However, they are available from the author on request.

[5] In relation to case 2, it should be noted that the constraint min(qλ,1) typically results in the adjusted regional matrix initially not scaling to *r,* the known level of regional intermediate transaction, but to something greater than *r*; effectively if *r* is the desired target, $\lambda = r/t$ is the 'wrong' scalar for the min(qλ,1) method of application.

[6] Regional matrices showing each region's input use irrespective of geographical source, or input use from Finland, were not available from the government statistics department for Finland when they were contacted.

[7] The difference between the conventional FLQ and the 'survey based' FLQ is not considered significant. The 'survey based' FLQ for example has a lower error than the conventional FLQ in 8/20 cases. Endnote 5 should also be taken into consideration: the constraint min(qλ,1) effectively makes the survey-based scalar somewhat 'inaccurate' relative to its application min(q,1) λ.