Dietrich W. Koeppen University Duisburg-Essen, Germany Department of Economics

Desaggregation in Interegional Input-Output Models -

An Estimation Approach for Distortions

17th International Input-Output Conference 2009

in Sao Paulo, Brazil, 13th to 17th July

1. Introduction.

Problems of sectoral aggregation in input-output analysis have been broadly discussed during the last decades, especially themes like the classification and and formalization of aggregation types and procedures, the measurement of aggregation effects and the search for an unbiased or perfect aggregation (Kymn, 1990, gives an overview, see also Olson, 2002).

This paper is focussed on some further questions. Instead of upward changes, the opposite procedure, a desaggregation, shall be observed. Formally, this is only a reversed operation which seems to have no new analytical consequences, but there are at least practical differences: a given input-output data set allows for any kind of aggregation to condensate the original informations, whereas a desaggregation usually needs new data, which may be not, or only partly available, they have to be gained by additional empirical research or by estimation. The semi-quantitative approach, represented here, may be seen as a first step to get a look into unknown structural relations, trying to use a minimum of new material. The intention is, to evaluate the original results, especially the measurement of influences from exogeneos impacts on endogeneos variables, in the light of possible distortions. In this context, a perfect desaggregation, although reachable, is seen only as a special case.

A regional desaggregation does not differ, at least in a formal sense, from a sectoral one, as long as the regions (eventually also countries) are treated as single units. But the paper combines the spatial and the sectoral perspective by investigating multiregional-multisectoral models with so called full information. Such a system could be desaggregated either at the regional, or the sectoral or at both levels. The paper concentrates on the division into subregions and offers an estimation approach to measure the analytical consequences of such an extension. It may be seen as a pre-investigation, preceding a further, more exact and costly step down to a lower level of aggregation by evaluating, among others, the reliability of the given model solutions, by comparing the effects of different aggregation patterns, or by searching maximal or minimal distortions.¹

The following sections develop the approach. At first, a formal frame is given, then the conditions are derived for positive, negative or zero biases with help of the usual exponential approximation of a standard input-output model solution, followed by Some extensions and a conclusion with hints on the practical use of the proposal. 2. Estimating desaggregation effects.

2.1 Basic relations.

To discuss different possibilities for desaggregating a multiregional and multisectoral input-output model and to analyze the consequences, the following frame shall be used. Let there be n regions with s sectors or industries; at the original level, usually denoted as macro level, the basic relations can then be defined as

$$M^* = A^* x^* + y^* = x^*$$
 (1)

with a system of intra- and interregional s by s matrices of input coefficients

where the first m regions remain unchanged, the others shall be split into subregions; x^* and y^* refer to total output and final demand for each sector in each region.

Any desaggregation leads to the micro system

$$\mathbf{M} = \mathbf{A}\mathbf{x} + \mathbf{y} \tag{2}$$

•

with input matrices, total output and final demand vectors for the first m regions and all subregions.

M* is transformed into **M** by a diagonal grouping system **G** of unity matrices which correspond in the first rows and columns to the not desaggregated regions, the following rows show the number of subregions for each original region:

•

G =	$\begin{array}{ccc} \mathbf{I} & 0 \dots 0 \\ 0 & \mathbf{I} \dots 0 \end{array}$	$\begin{array}{ccc} 0 \dots 0 & 0 \\ 0 \dots 0 & 0 \end{array}$)0)0	$\begin{array}{c} 0 \dots 0 \\ 0 \dots 0 \end{array}$	
	 0 0 I			00	
	$\begin{array}{cccc} 0 & \ldots & 0 \\ 0 & \ldots & 0 \end{array}$	II 00	00 II	$\begin{array}{c} 0 \dots 0 \\ 0 \dots 0 \end{array}$	
	 0 0	···· 0 · · · 0 (· · · 0	 II	(3)

A weighting system W, built analogous to G, where the unity matrices are replaced by diagonalized weights derived from output shares, shows the distribution of the subregions.

The macro and the micro level are connected then by

$$A^* = GAW'$$

$$y^* = Gy$$

$$x^* = Gx$$
 (4)

Desaggregation effects can now be observed as differences between model results. Using a standard demand oriented static version, \mathbf{D} indicates if an output variation induced by exogenous variations of final demand differs at a deeper level from the original results at the macro level

$$\mathbf{D} = (\mathbf{A} - \mathbf{I})^{-1} \mathbf{y} \cdot (\mathbf{A}^* - \mathbf{I})^{-1} \mathbf{y}^*$$
(5)

which shows all effects, concerning each sector in each region in relation to any part of final demand, in detail. \mathbf{D} is of central meaning for aggregation theory, with special attention to conditions where all differences disappear. This may be, of course, also interesting in case of a desaggregation, but an analyst is possibly searching other types and effects and his main problem could be that data at a lower level are more or less unknown. The following section presents therefore an estimation procedure for all possible effects, including a zero bias desaggregation as special case.

2.2 A semi-quantitative estimation approach – first order effects.

The constellation which shall be discussed first is simple but, nevertheless, of some practical meaning. Let there be a two-regional model with intra- and interregional relations for a region 1, which may be called the observation region, whereas the second region r represents the aggregated rest of the economy. Searched are, under restricted information, possibilities to estimate the desaggregation effects if the rest of the economy r is split, in a first step, into two subregions.

The macro system is defined by

$$\mathbf{A}^{*} = \begin{matrix} \mathbf{A}_{11} & \mathbf{A}_{1r} & \mathbf{x}_{1} & \mathbf{y}_{1} \\ \mathbf{x}^{*} = & \mathbf{y}^{*} = & \mathbf{y}^{*} = \\ \mathbf{A}_{r1} & \mathbf{A}_{rr} & \mathbf{x}_{r} & \mathbf{y}_{r} \end{matrix}$$
(6)

The grouping and weighting matrices are

A special approximation of (5) shall be used to derive the differences between micro and macro results, the power series

$$\mathbf{D} = (\mathbf{G}\mathbf{A} - \mathbf{G}\mathbf{A}\mathbf{W}'\mathbf{G})\mathbf{y} + (\mathbf{G}\mathbf{A}^2 - \mathbf{G}\mathbf{A}^2\mathbf{W}'\mathbf{G})\mathbf{y} + \dots + (\mathbf{G}\mathbf{A}^z - \mathbf{G}\mathbf{A}^z\mathbf{W}'\mathbf{G})\mathbf{y}$$
(8)
$$\mathbf{D} = \mathbf{D}_{(1)} + \mathbf{D}_{(2)} + \dots + \mathbf{D}_{(z)}$$

which measures the desaggregation effects of so called first and higher order.

The concept to work with this approximation has a tradition in aggregation theory (see Theil, 1957; Kymn 1990) but, usually, only the first order effects are analyzed. It will be shown below that this restriction must be given up for certain cases.

From (6) and (7) follows

$$\mathbf{D}_{(1)} = \frac{(\mathbf{A}_{11} - \mathbf{A}_{11})\mathbf{y}_{1} + (\mathbf{W}_{2,1}\mathbf{A}_{12}\mathbf{y}_{2} + \mathbf{W}_{2,2}\mathbf{A}_{13}\mathbf{y}_{3}) - (\mathbf{A}_{12} + \mathbf{A}_{13})(\mathbf{y}_{2} + \mathbf{y}_{3})}{[(\mathbf{A}_{21} + \mathbf{A}_{31}) - (\mathbf{A}_{21} + \mathbf{A}_{31})]\mathbf{y}_{1} + \mathbf{W}_{2,1}(\mathbf{A}_{22} + \mathbf{A}_{32})\mathbf{y}_{2} + \mathbf{W}_{2,2}(\mathbf{A}_{23} + \mathbf{A}_{33})\mathbf{y}_{3} - (\mathbf{A}_{22} + \mathbf{A}_{32} + \mathbf{A}_{33})(\mathbf{y}_{2} + \mathbf{y}_{3})]}$$

$$\mathbf{D}_{(1)} = \frac{\mathbf{D}_{11(1)} + \mathbf{D}_{1r(1)}}{\mathbf{D}_{r1(1)} + \mathbf{D}_{1r(1)}}$$
(9b)

Four groups of differences can now be observed: D_{11} denotes the effects in region 1 if final demand in this region is changed, thus representing the intraregional effects in region 1; D_{1r} concerns output variations in region 1 coming from exogenous changes region R; D_{r1} belongs to impacts on R resulting from final demand in region 1; D_{rr} marks the intraregional effects in R. All **D** are sqare of rank s, the number of sectors, thus describing the desaggregation consequences for each sector and each part of final demand.

Apparently, the first order effects are zero in two cases:

$$\mathbf{D}_{11(1)} = (\mathbf{A}_{11} - \mathbf{A}_{11}) \mathbf{y}_1 = 0$$
(10)
$$\mathbf{D}_{R1(1)} = [(\mathbf{A}_{21} + \mathbf{A}_{31}) - (\mathbf{A}_{21} + \mathbf{A}_{31})] \mathbf{y}_1 = 0$$
(11)

a result which follows from the assumption that region 1 remains unchanged, but which has to be controlled later by observing the effects of higher order.

A next group of differences can be derived now from (10):

$$\mathbf{D}_{1r(1)} = \mathbf{W}_{2,1}\mathbf{A}_{12}\mathbf{y}_2 + \mathbf{W}_{2,2}\mathbf{A}_{13}\mathbf{y}_3 - (\mathbf{A}_{12} + \mathbf{A}_{13})(\mathbf{y}_2 + \mathbf{y}_3)$$
$$\mathbf{D}_{1r(1)} = (\mathbf{A}_{12} - \mathbf{A}_{13})\mathbf{W}_{2,2}\mathbf{y}_2 + (\mathbf{A}_{13} - \mathbf{A}_{12})\mathbf{W}_{2,1}\mathbf{y}_3$$
(12a)

and at the sectoral level

$$d_{1r,pq(1)} = (a_{12,pq} - a_{13,pq}) w_{2,2} y_{2p} - (a_{13,pq} - a_{12,pq}) w_{2,1} y_{3,p} \qquad (p,q = 1...s)$$
(12b)

According to (12), the desaggregation effects of first order for this group are given by the relation between two corresponding input coefficients and two weighted parts of final demand.

If these variables are known the differences between macro and micro results of the first order are given exactly by (12), but in case of unknow microstructures, at least the signs can be found in the following way

$$\begin{aligned} d_{1r,pq\,(1)} &> 0 \quad \text{if } a_{12,pq} &> a_{13,pq} \quad \text{together with } y_{2,p} &> y_{3,p} \\ \text{and also } \text{if } a_{12,pq} &< a_{13,pq} \quad \text{together with } y_{2,p} &< y_{3,p} \end{aligned}$$
 (13a)

$$d_{1r,pq(1)} < 0 \quad \text{if } a_{12,pq} > a_{13,pq} \text{ together with } y_{2,p} < y_{3,p}$$
(13b)
or if $a_{12,pq} < a_{13,pq} \text{ together with } y_{2,p} < y_{3,p}$

$$d_{1r,pq(1)} = 0$$
 if $a_{12,pq} = a_{13,pq}$ and/or if $y_{2,p} = y_{3,p}$ (13c)

An example may illustrate these conditions: let there be an industry 1 in region 1 which delivers (among others) to industry 2 in region 2 (the first subregion in R) and to industry 2 in region 3 (the second subregion in R); it shall be estimated that the input coefficient $a_{12,12}$ for the first subregion exceeds $a_{13,12}$ for the other subregion; at the same time, the estimated deliveries of industry 1 in region 2 to the final demand of region 2 $y_{2,1}$ shall be greater than the corresponding relations of industry 1 in region $y_{3,1}$. The consequences are, according to (14a), that there is a positive difference $d_{1r,12}$ between the impacts at the micro level, after an assumed desaggregation, and those at the macro level: the original results are downward biased, they underestimate the influence of final demand on sectoral output (it should be noted that a reversed ranking order for both variables does not change this derivation).

The opposite case is given in (13b) where the order between the input coefficients does not correspond with that of the final demand parts. The difference d is then negative, a potential desaggregation leads to an overestimation at the macro level.

The final constellation (13c) marks an unbiased desaggregation. The difference d disappears if either the input coefficients or the final demand deliveries (or both) are equal.

A last group of desaggregation effects, resulting from a variation of y_r for x_1 , can be derived again from (10)

$$\mathbf{D}_{rr(1)} = (\mathbf{A}_{22} + \mathbf{A}_{32}) \mathbf{W}_{2,2} \mathbf{y}_2 + (\mathbf{A}_{23} + \mathbf{A}_{33}) \mathbf{W}_{2,3} \mathbf{y}_3 - (\mathbf{A}_{22} + \mathbf{A}_{32} + \mathbf{A}_{23} + \mathbf{A}_{33}) (\mathbf{y}_2 + \mathbf{y}_3)$$

$$\mathbf{D}_{rr(1)} = [(\mathbf{A}_{22} + \mathbf{A}_{32}) - (\mathbf{A}_{23} + \mathbf{A}_{33})] \mathbf{W}_{2,3y2} + [(\mathbf{A}_{23} + \mathbf{A}_{33}) - (\mathbf{A}_{22} + \mathbf{A}_{23})] \mathbf{W}_{2,2} \mathbf{y}_3$$

(14)

This means for any sectoral combination (p,q=1...s)

$$d_{r,pq(1)} > 0$$
 if $(a_{22,pq} + a_{32,pq}) > (a_{23,pq} + a_{33,pq})$ and $w_{2,3} y_{2,p} > w_{2,2} y_{3,p}$ (15a)

but also if $(a_{22,pq} + a_{32,pq}) < (a_{23,pq} + a_{33,pq})$ and $y_{2,p} < y_{3p}$

A negative difference results, $d_{rr,pq(1)} < 0$

if either
$$(a_{22,pq} + a_{32,pq}) > (a_{23,pq} + a_{33,pq})$$
 together with $w_{2,3y_{2,p}} < w_{2,2} y_{3,p}$
or $(a_{22,pq} + a_{32,pq}) < (a_{23,pq} + a_{33,pq})$ and $w_{2,3}y_{2,p} > w_{2,2} y_{3,p}$ (15b)

There is no difference between macro and micro results

$$d_{rr,pq(1)} = 0$$
 if $(a_{22,pq} + a_{32,pq}) = (a_{23,pq} + a_{33,pq})$ and/or if $w_{2,3} y_{2,p} = w_{2,2} y_{3,p}$ (15c)

The interpretation of these conditions runs analog to (13a,b,c): a parallel order of input coefficients and final demand yields a positive effect, the opposite case is given by a reversed ranking; the effect is zero if the coefficients or the final demand deliveries are equal.

2.3 Effects of higher order.

To control the derivation of desaggregation effects of the first order, differences of the second and of higher order shall be observed now. A first group concerns the intraregional effects for region 1. They can be found from the second term in (8) together with (6) and (7):

$$\mathbf{D}_{11(2)} = (\mathbf{A}_{11}\mathbf{A}_{11} + \mathbf{W}_{2,1} \mathbf{A}_{12}\mathbf{A}_{21} + \mathbf{W}_{2,2} \mathbf{A}_{13}\mathbf{A}_{31}) \mathbf{y}_1 - [\mathbf{A}_{11}\mathbf{A}_{11} + (\mathbf{A}_{12} + \mathbf{A}_{13})(\mathbf{A}_{21} + \mathbf{A}_{31})] \mathbf{y}_1$$

$$\mathbf{D}_{11(2)} = \mathbf{W}_{2,1} \mathbf{A}_{12} (\mathbf{A}_{21} - \mathbf{A}_{31}) + \mathbf{W}_{2,2} \mathbf{A}_{13} (\mathbf{A}_{31} - \mathbf{A}_{21})$$
(16a)

and at the sectoral level

$$d_{11,pq(2)} = w_{2,2} a_{12,pq} (a_{21,pq} - a_{31,pq}) + w_{2,1} a_{13,pq} (a_{31,pq} - a_{21,pq})$$
(16b)

This result is remarkable: although region 1 is not desaggregated, the division of the rest of the economy r has an influence on the intraregional effects for 1.

These interdependencies do not show up as first order biases, (10) and (11) are zero, but they appear as second order effects.

We there are positive or negative distortions at the original level can be derived from (16) in the usual way. Deciding are, apparently, the relations between the weighted coefficients $a_{12,pq}$ and $a_{13,pq}$ at the one side, and $a_{21,pq}$ with $a_{31,pq}$ at the other side: parallel ranks mean underestimation at the macro level, opposite ranks show that the

original results are overestimated, the effects are zero if either the first pair of coefficients or the second is equal.

Second order effects for the next group concerning relations between r and y_1 , which also have been zero initially, are given by

 $\mathbf{D}_{r1(2)} = (\mathbf{A}_{21} - \mathbf{A}_{31}) [\mathbf{W}_{2,2}(\mathbf{A}_{22} + \mathbf{A}_{32}) - \mathbf{W}_{2,1}(\mathbf{A}_{23} + \mathbf{A}_{33})]$ (17) with all possibilities to find the conditions for desaggregation effects at the sectoral level in analogy to (16).

Slightly different is the situation for $\mathbf{D}_{1r(2)}$ and $\mathbf{D}_{rr(2)}$ because there are distortion already of first order. Again, the new conditions follow from (8), but it appears, after a comparison, That the signs, which signalize positive, negative or zero biases, are identical with those of the first derivations (13) and (15). Only the absolute values change, they become smaller.

What was found about the second order desaggregation effects, is confirmed concerning all effects of higher order: there are no changes of the signs but decreasing absolute values.

2.3 Some extensions.

The case which was discussed above, a bi-regional model with an observation region and one aggregated region, split into 2 subregions, represents a special constellation. So at first, the possibilities to extend the number of regions which are not desaggregated shall be regarded shortly.

Going back to (1) and (2), the system consists now of m such regions confronted with one region R which represents the rest of the economy. there are now m+1 by m+1 difference matrices $D_{i+1,j+1}$ showing the biases between macro and micro results.

It can be seen easily that, for example, $\mathbf{D}_{1r(1)}$ turns now into $\mathbf{D}_{ir(1)}$ which is for all m regions

$$\mathbf{D}_{ir(1)} = (\mathbf{A}_{i \ m+1} - \mathbf{A}_{i \ m+2}) \mathbf{y}_{n+1} + (\mathbf{A}_{i \ m+2} - \mathbf{A}_{i \ m+1}) \mathbf{y}_{n+2} \qquad (i = 1... \ m)$$
(18)

All condition for over- or underestimation as well as the zero bias case then can be found as shown above. Not only in this case but for all the other derivations of desaggregation effects at the regional and the sectoral level, the corresponding coefficients and the final demand deliveries have now to be compared between any of the m unchanged regions and the subregions in r. More difficulties arise, if the number of the subregions shall extended. To demonstrate the principle, the case of one not disaggregated region and 3 subregions may be regarded. (12) for example, then becomes

$$\mathbf{D}_{1r(1)} = \mathbf{W}_{2,1}\mathbf{A}_{12}\mathbf{y}_1 + \mathbf{W}_{2,2}\mathbf{A}_{13}\mathbf{y}_2 + \mathbf{W}_{2,3}\mathbf{A}_{14}\mathbf{y}_3 - (\mathbf{A}_{12} + \mathbf{A}_{13} + \mathbf{A}_{14})(\mathbf{y}_2 + \mathbf{y}_3 + \mathbf{y}_4)$$
(19a)

Formally, a generalization for more subregions would also be possible, but the identification of positive or negative differences can be problematic even with only 3 subregions because of the following conditions:

$$\mathbf{D}_{1r(1)} > 0$$
 if $\mathbf{A}_{12} > \mathbf{A}_{13} > \mathbf{A}_{14}$ together with $\mathbf{W}_{2,4} \mathbf{y}_2 > \mathbf{W}_{2,2} \mathbf{y}_3 > \mathbf{W}_{2,3} \mathbf{y}_4$ (19b)

$$\mathbf{D}_{1r(1)} < 0$$
 if $\mathbf{A}_{12} > \mathbf{A}_{13} > \mathbf{A}_{14}$ with $\mathbf{W}_{2,4} \mathbf{y}_2 < \mathbf{W}_{2,3} \mathbf{y}_3 < \mathbf{W}_{2,1} \mathbf{y}_4$ (19c)

$$\mathbf{D}_{1r(1)} = 0$$
 if $\mathbf{A}_{12} = \mathbf{A}_{13} = \mathbf{A}_{14}$ and/or $\mathbf{W}_{2,4} \mathbf{y}_2 = \mathbf{W}_{2,2} \mathbf{y}_3 = \mathbf{W}_{2,3} \mathbf{y}_4$ (19d)

Thus, the sign of **D** can be identified if the ranking orders are straight, but difficulties arise with mixed rankings, for example if $A_{12} > A_{13} > A_{14}$ (this can be reached by the numbering of the subregions) together with the weighted $y_{12} > y_{13} < y_{14}$. In this case, not only the ranks but the differences $y_{12} - y_{13}$ and $y_{13} - y_{14}$ are deciding, which needs additional estimations. For a larger number of subregions it could be helpful, in this case, to observe at least parameters like Spearmans rank correlation coefficient ς to get a hint on the direction of D_{1R} .

Similar problems arise for all the other conditions, that means a more complex estimation process than for two subregions, but depending on the given some results may be reachable.

It should be noted finally that the approach is not restricted to a standard demand driven model but could be applied directly to all dual models and those with similar solutions, for example base on capital coefficients.

3. Conclusions

This paper has sketched out, what should be regarded as a proposal rather than an elaborated method. So, the following remarks do not give a detailed evaluation but only some hints on a possible use.

Assuming that the need for regional studies which are based on input-output techniques will increase rather than decline, it seems plausible to improve the use of already existing data sets and models for instance by a further desaggregation. The possibilities for such a step downwards depend on the given material and on the costs of additional research, either by

survey, non-survey or hybrid methods. At the other hand, and in relation to these constraints, the intentions may reach from a loose estimation of desaggregation effects, restricted to selected regions or sectors, to full information about structural relations at a lower level.

It seems reasonable, to plan such an attempt as a stepwise investigation, beginning with more simple questions and results. The procedure, described above, could provide such a starting point. First estimations may have a wide range, signalizing only if results at the original level are too high or too low compared with those after a desaggregation, but the margins can be reduced by further informations. Other possibilities are given by simulating different data constellations within the limits of the original material. This may be of special interest, if alternative types of desaggregation shall be observed.

Advantages may also been seen in a certain flexibility of the procedure which can be focussed on a single sector and its relation to another sector and final demand, but also extended to row and column sums of effects, up to overall effects for whole regions or subregions.

It should be noted that empirical desaggregation effects tend to be small, staying in many cases below the limits of data accuracy. Nevertheless, the conditions which were derived allow to find the constellations where the effects reach a maximum or a minimum. All these possibilities may underline the preliminary but, in a certain sense, also heuristic character of the approach.

Notes:

¹ Besides a great number of non-survey techniques, special desaggregation approaches have been developed at the sectoral level, for example by Wolski (1984) and Gillen and Guccione (1990). They provide exact solutions, as well as estimations, but they seem to need more detailed informations, for instance about prices and price variations. Nevertheless, they may be very usefull in combination with the approach, sketched out in this paper

² There are possibilities not to go down to the level of single sectors as in (12b) but to use row and column sums of the coefficient matrices and final demand vectors of (12a). This re-aggregation which could be applied to all other equations at the level of the regional units may simplify the estimation procedure and provide more concentrated informations.

³ This may include, in a wider sense, not only the modelling of interdependencies but also all the methods to organize structural information as flows between units or activities in the typical way of input-output systems.

References:

- Gillen, William J. and Guccione, Antonio : "Disaggregating Input-Output Models; an Alternative to Wolsky's Method " Economic Systems Research, Vol. 2,1 (1990)
- Kymn, Kern O. : "Aggregation in Input-Output Models: a Comprehensive Review, 1946 – 71 " Economic Systems Research, Vol. 2,1 (1990)
- Olson, J. Asger : "Perfect Aggregation of Industries in Input-Output Models" In: Input-Output Analysis: Frontiers and Extensions. Ed. by Lahr, Michael L. and Dietzenbacher, Erik. Hampshire and New York (2001)
- Theil, Henry : "Linear Aggregation in Input-Output Analysis" Econometrica, Vol. 25 (1957)
- Wolsky, Alan M. : "Disaggregating Input-Output Models" The Review of Economics and Statistics, Vol. 61 (1984)