D1 Final report of methodology for the generation of P/C matrices for the Swedish National freight model system

Christer Anderstig, Siv Schéele (Inregia AB), Jan Eriksson (VTI), Lars Harald Vik, Arne Stokka (SINTEF)

on commission by SIKA on behalf of the Samgods group

Contents

1 Introduction

Background and purpose of the study

In connection with the development programme for a Swedish National Freight Model, it has been decided to develop a set of “base” commodity-specific “production-consumption” [P/C] matrices of flows to and from all Swedish municipalities. This will provide the input to a logistics module, which will then produce a set of consignment-based matrices [O-D] as input to a network model, with the aim of allocation to modes and routes.

The logistics module is being developed within a separate project together with the Norwegian planning authorities. There should be close working between the logistics and base matrices projects, since much of the data and insights will be common.

There is already a set of base matrices within the Samgods model, but these are of a “hybrid” nature, being partly P/C and partly O-D (origin-destination). In addition, there is considerable doubt as to their reliability.

The long term tasks for the project are a) to develop a methodology for integrating data from different sources, b) to develop procedures for implementing the methodology, and hence c) to produce the set of base P/C matrices.

Given the complexity of the process and the data sets involved, it will be necessary to proceed in an incremental way.

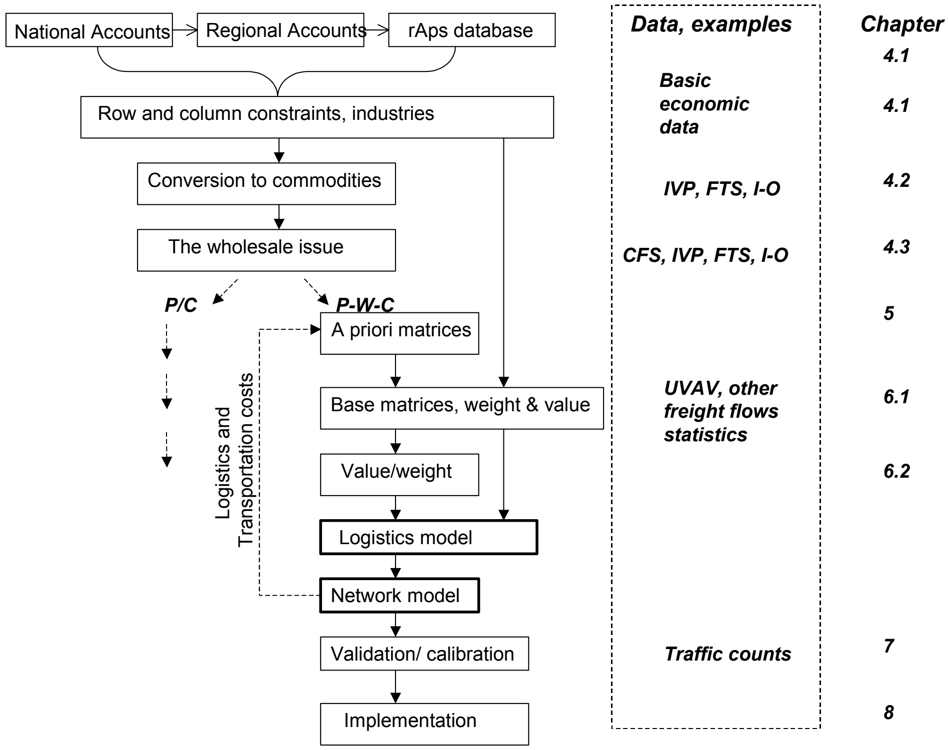
This report refers to the first six months of the base matrix development and implementation that would take about three years in total.

2 Contents of the report, an overview

The following figure gives an overview of the report. To the left the different modules are listed and their relations are indicated. For each module the main data sources used are listed and so is the chapter in the report where the subject is dealt with.

There is a crucial choice to be made with respect to a pure P/C versus a PWC matrix. The foundation of the PWC alternative is presented rather carefully. Some aspects of the wholesale issue are also discussed in Annex 4.

Details on basic economic and other data are found in Annex 1.



3 Methodology for generating base matrices – background and outline

3.1 Introduction

In the review of the Swedish National Freight Model, Samplan 2004:1, central functions and components of the future model structure are presented in the concluding chapter. As to the P/C matrices it is stated that the present mixed O-D and P/C matrices should be replaced by “pure” P/C matrices and that the P/C matrices could be regarded as the very heart of the whole model system or the pivot point around which everything else rotates. Further, “fundamentally, the P/C matrices for the Base are based on empirical data. This is a prerequisite for the matrices to live up to the general conditions of robustness and precision that were defined as essential for the base case.”

It goes without saying that estimations of cells in the P/C matrices based on empirical observations is a better approach than any model based estimation, provided necessary accuracy of observed data. This ‘empirical data approach’ is also the main approach according to our suggested methodology.

It may seem as if the preconditions for this approach is promising, since data now are becoming available through the Commodity Flow Survey, CFS. Actually, to our knowledge surveys of this kind are conducted only in Sweden and in the US[1].

However, in the Swedish context CFS is essentially a new data source and several questions related to its scope and design have to be solved before it can be seen as fully adapted to the purpose of estimating P/C flows. E.g., there is at present inadequate identification as to the type of activity generating and receiving shipments.

Therefore, the ‘empirical data approach’ for estimating base P/C matrices has to integrate other data sources than CFS, as well as making use of prudent model based estimation. It is not surprising, yet interesting, to find that a similar perspective with respect to data integration also is present in the US context[2].

Let us temporarily disregard how other data sources could complement CFS and focus on information gained from model based estimation. In agreement with Samplan 2004:1, we can assume “that some kind of probabilistic spatial allocation model will be required”.

---

[1] In US surveys have been conducted 1993, 1997 and 2002. The next CFS will be conducted 2007.

[2] See, e.g., Ambite, J.L et al (2002)

3.2 SCGE models and transportation costs

In the SCGE Pre-study, p.19, a specification of such a model is presented, where trade flows ("the amount of each commodity that is transported from each production zone to each consumption zone") are calculated by use of a standard logit type formulation:

- Given the consumption of commodity *m* in zone *c*, the probability that the commodity has been transported from zone *p* is affected by the capacity for production of *m* in zone *p* and the disutility (generalized cost) of the production of one unit of *m* in zone *p*, plus the disutility of transporting it from zone *p* to zone *c*.

This modeling approach is very close to what was presented by the same author in a study for the Department of Transportation in UK, ME&P (2002). There is, however, a difference with respect to context and preconditions for creating the base P/C matrix since, in UK, there is *no* data source available that provides direct observations of this matrix. (In contrast to CFS in Sweden.)

Therefore, in the UK context the corresponding matrix is built up indirectly "as an output from combining the observed data that is available within an appropriate model structure that explicitly includes the logistical stages" (observed O-D matrices of modal shipments by type of commodity).

The suggested procedure is complex. In brief, it means that the model results are compared (and calibrated) against O-D totals. The latter are constructed by merging intermodal shipments: "multiple movements of a commodity that are measured as separate modal shipment legs are collapsed into a single intermodal shipment from the factory to the port".

Turning to the SCGE Pre-study, one of its basic ideas is that the present Swedish national freight model could be called in question with respect to the "embedded assumption that cost of transport has **no** influence on patterns of goods movement". The suggested method for calculating P/C matrices is evidently a relaxation of this assumption. The question is to what extent it is possible to estimate elements in a P/C matrix as determined by disutility of production and transportation[3].

Going over to actual and operational SCGE models, Bröcker (1998) is a good representation. Bröcker uses a similar gravity model to describe trade flows between more than 800 regions in Europe, where transport cost is a function of distances (shortest routes). Based on information from reviews of logistic and transport costs, the calibrated transport cost intensity, i.e. share of sales value, is in the order of 5 to 10 percent. Only one average figure is needed, since the model includes only one tradable good.

[3] Further, variations in "disutility of production", i.e. variation in unit cost of production, is – to be brief – not observable, nor consistent with the (for the same reason, necessary) assumption of fixed input-output technology (i.e., the same production function) in all regions.

Further, only transport cost is included since logistic costs, which are not related to distance, cannot be included in the estimate.

In a more recent SCGE application, Bröcker (2002), the approach is extended by also including business and private passenger travel. The rationale for this extension is that “Most SCGE applications...have up to now only taken the impact of goods transport into consideration. To a large extent, however, welfare effects are due to time and cost savings in passenger transport.”

This extension is not so usual in the SCGE context, whereas the motive is uncontroversial. E.g., in the SCGE Pre-study, it is stated in the introduction that “there are indeed reasonable grounds for assuming that the influence of the cost of transport on the pattern of goods movements is not as powerful as the influence of cost on passenger travel patterns.”

The relative importance of goods transport cost versus passenger transport cost is an issue being highlighted in a recent paper by Glaeser and Kohlhase (2004). Their discussion is mainly concerning the implications of declining transportation costs for manufactured goods with respect to the theoretical framework of urban and regional economics. In brief, they claim that in the world of today it is essentially free to move goods, but expensive to move people. Certainly, this statement is being qualified in the paper. Yet, on basis on best available data for US they find that today the average transport costs share (of GDP) is in the (low) order around 3 to 6 percent.

More interesting is their analysis of the importance of transport costs across industries, using data from the 1997 Commodity Flow Survey in US. Not very surprisingly, they find quite important transport costs for some industries (e.g., wood products, basic chemicals). However, many bigger industries “all face trivial transportation costs”.

The paper by Glaeser and Kohlhase serves as an inspiring piece of analysis, with respect to a corresponding analysis on data from the Swedish CFS. Starting with preliminary explorations, e.g. the relationship between value per ton and average length of haul, more elaborate models could successively be developed. Celik and Guldmann (2002) is such an example of modeling interregional commodity flows, on basis of Commodity Flow Survey data.

In sum, there are suitable probabilistic spatial allocation models, whether termed logit, gravity or entropy, to be used for a careful estimation of the influence of transportation costs on trade flows, i.e. elements in the P/C matrices. In a final stage the cost matrices should comprise not only transportation costs but also all relevant cost elements, i.e. the generalized logistic costs provided by the logistics model together with the network model. As these generalized costs will be derived with the P/C matrices as input, it follows that an iterative procedure will be necessary, where cost matrices are successively refined.

3.3 Suggested methodology – in essence

According to the discussion above some central features and conditions for the construction of base matrices could be summarised as follows:

- The base matrices should be pure P/C,
- The P/C matrices should fundamentally be based on empirical data,
- The Commodity Flow Survey, CFS, provides empirical data of the requested kind,
- There are problems related to both the scope and the design of CFS which makes it necessary to use complementary matrix data,
- One appropriate type of complementary data makes use of model based estimation with respect to the influence of transportation costs.

However, taking all components into consideration, it turns out that the idea of pure P/C matrices is problematic, and the problem is basically related to the role of the wholesale sector.

In the pure P/C matrix concept it is assumed that all wholesale activities are excluded, since these activities do not concern production of commodities. (The distinction between the wholesale *sector* and wholesale *activities* is made for the simple reason that manufacturing activities within the wholesale sector are included in the P/C concept.)

From the logistics project perspective it is still a problem that wholesale activities cannot only be defined as 'logistic activities', i.e. as activities by consolidation and distribution centers.

From the base matrix project perspective another problem is related to CFS. More than half of all observations in CFS are representing shipments from the wholesale sector (W). Therefore, some strong assumptions have to be made to make use of these observations in the construction of pure P/C matrices [4].

In the report from the logistics project additional problems with the pure P/C concept are discussed. Taking the arguments altogether, instead of constructing pure P/C matrices,

- There is reason to start with constructing base matrices where wholesale is fully incorporated.
- This means that flows from W will be used as flows from P and flows to W as flows to C. The outcome can be described as PWC matrices.

In the presentation below it will generally be understood that this PWC approach is applied. However, since the option of pure P/C matrices is of great

[4] The same type of problem is noted in US when making use of CFS for estimating flow matrices; see Jackson et al (2004).

concern for further investigations, questions related to pure P/C matrices will also be given specific attention.

The main features of the suggested methodology can be described according to the illustration in figure 1.

Figure 1 Base matrix PWC for commodity k, in weight

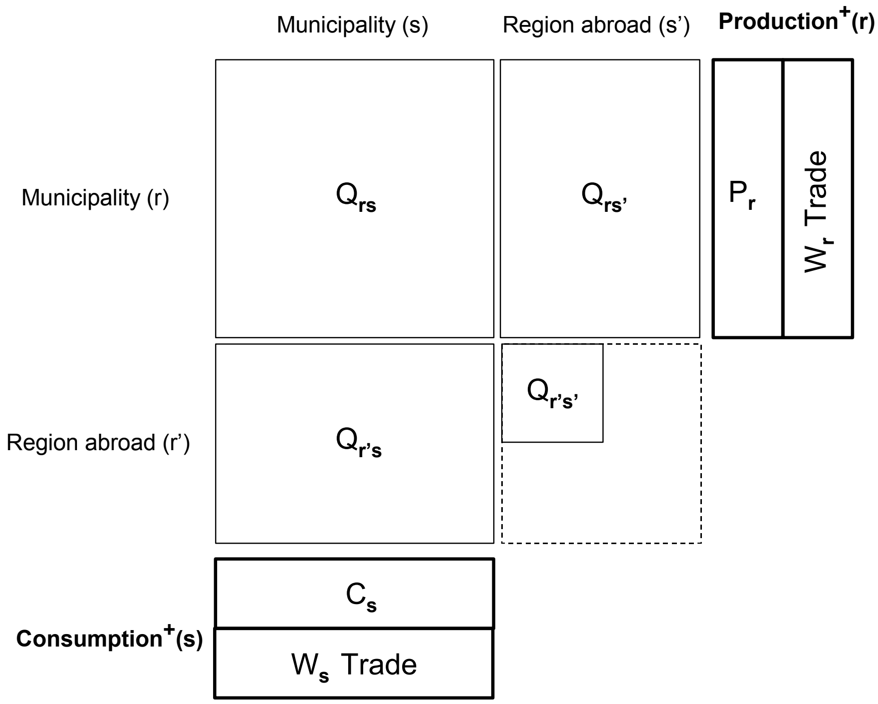


Figure 1 represents the final outcome, when base matrices in weight (Q) are provided. The corresponding (and preceding) matrices in value will be denoted (V).

The definition of regions refers to 290 municipalities (r, s) in Sweden and 174 STAN99 regions abroad (r', s'). We will later on use capital (R, S) and (R',S') respectively, to denote larger regions.

The row and column constraints are here termed Production$^+$ and Consumption$^+$, to indicate that trade is added to P and C in PWC matrices:

W_r is the commodity *k* sold by W in *r*; W_s is the commodity *k* purchased by W in *s*. By definition of trade it follows that $W_r \equiv W_s$ for r=s.

But, before coming to PWC matrices in weight, several preceding steps of construction work refer to value terms.

First, the construction of row and column constraints starts by using regional economic data, at the municipality level, derived from the rAps database and from using the rAps model system. This step gives the value of P and C at the industry level (h).

In the next step P(h) and C(h) are converted to commodities (k). Then data on wholesale trade, also available for commodity (k), can be added to get the row and column constraints, P^+ and C^+, still in value terms.

Next, an a priori matrix is constructed for each k, both in value and weight. This matrix is a combination of two different matrices:

A) Matrix elements based on direct observations from CFS;

B) Matrix elements based on a gravity model approach, using CFS, row and column constraints and data on transportation costs (generalised cost from STAN).

The B component cannot be calculated for the diagonal elements (r, r) for which transportation cost data are missing. These diagonal elements are calculated by a specific method, the regional purchase coefficient (RPC) technique.

Given the a priori matrices, by use of entropy maximisation the matrix elements are estimated subject to row and column constraints, P^+ and C^+. The estimation of matrices in weight is making use of value/weight relations to get the column (consumption) constraints.

Some comments should be made concerning the row and column constraints to be used. First, from foreign trade statistics (FTS) additional row and column constraints will be defined. As FTS is given both in value and weight this information can be described in weight terms.

FTS gives information on Swedish bilateral export and import of commodity k. Thus, flows from a region abroad should equal Swedish import from this region,

$$\Sigma_s Q_{r's} = Q_{r'.},$$

and flows to a region abroad should equal Swedish export to this region

$$\Sigma_r Q_{rs'} = Q_{.s'}.$$

In case regions abroad represent parts of a foreign country, a summation over $r' \in R'$, $s' \in S'$ is added to get the respective constraint.

Another comment is concerning differences in the constraint structure, with respect to differences in reliability of data. The row constraints for production P are more reliable than the column constraints for C, since the former are taken from data whereas the latter are based on model calculations. This difference in reliability will be handled by introducing ‘soft’ constraints for C.

Soft constraints are also suggested to make use of information from CFS, by aggregating observed flows in CFS to larger regions (NUTS II), the regional level at which CFS is expected to be reliable.

Finally, in figure 1 one part of the trade between foreign regions, $Q_{r's'}$, is indicated. This part is meant to illustrate possible transit movements through Sweden. Although trade flows between non-domestic regions in general will be ignored, transit traffic through Sweden is treated in the logistics project and the corresponding matrices must therefore be provided. Some part of the transit traffic will be represented in the PWC matrices, namely reexport of imported commodities by wholesalers.

4 Row and column constraints

The purpose of this chapter is to describe how economic variables will be used to construct the constraints for the base matrices that are the ‘row totals’ for production and the ‘column totals’ for consumption. We will start with P and C, as defined in the pure P/C concept. The additional constraints when wholesale is incorporated in the PWC case will be discussed in a separate section 4.3.

4.1 Estimation of row and column totals at the industry level, P and C

As the base matrices refer to *commodity* groups, at present the 12+ STAN groups, an initial question is why economic variables at the *industry level* should be addressed at all.

The basic reason is that the estimation of row and column totals, (P) and (C), needs model support, and that the appropriate rAps database and model system is operating at the industry level.

It is true that most data on production (P) is available from various data sources, without need of any model based calculations. This does not, however, apply to consumption data (C). There are no sources for regional data on, e.g., final private consumption or intermediate consumption in the production system. Thus, model based calculations (rAps) are necessary to get information on both (P) and (C). To get sufficient information for row and column totals, however, a transformation from industry (h) to commodity group (k) is needed, see section 4.2

The quality of this transformation depends, among other things, on the quality of the data at the industry level. Hence, the procedure for constructing row and column constraints should include a check and adjustment of rAps economic data, in order to get reliable estimates of P(h) and C(h)[5]. Although such adjustments will be part of the implementation, some expected cases will be noted in the presentation below.

4.1.1 The role of rAps, introductory notes

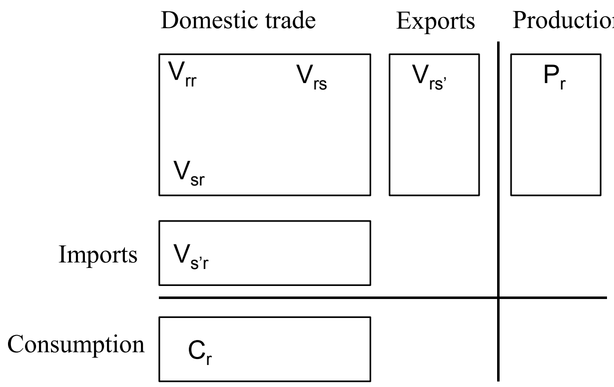
The macroeconomic sub-model in rAps contains for each region and industry a balance of resources of traditional type. This balance can be reduced to the following expression

Production - Exports + Imports = Consumption

[5] This check and adjustment refers both to National Accounts and to Regional Accounts, or corresponding data where regional accounts are missing. For a description of these basic economic data sources, see Annex 1.

where Exports and Imports refer to both domestic and foreign trade. By using the terms Exports and Imports exclusively to represent foreign trade, we can schematically illustrate how information from the balance of resources in rAps represents parts of the P/C relationships for region *r*, industry *h*.

Figure 2 Representation of P/C relationships in rAps, region *r* industry *h*



It should be noted from this figure that (in the rAps context) domestic region *s* is an aggregate of all other regions in Sweden, and in the same way *s'* is an aggregate of all regions abroad.

The same P/C relationships apply to any region *r*, i.e. a summation for all Swedish regions gives the domestic trade balance

$$\sum_r V_{rs}^h = \sum_r V_{sr}^h \quad .$$

As to domestic trade neither National Accounts nor the Regional Accounts provide any functional data. Domestic trade is therefore calculated within the rAps model system, where base year exports[6] to other domestic regions *s* is

$$V_{rs}^h = P_r^h - V_{rs'}^h - V_{rr}^h \,.$$

Similarly, base year imports from other domestic regions *s* can be calculated as

$$V_{sr}^h = C_r^h - V_{s'r}^h - V_{rr}^h \,,$$

although such calculation for the base year is not performed in the model. No base year calculation is needed since domestic imports, in the same way as foreign imports, are determined endogenously – by fixed parameters for the domestic imports shares. These domestic import shares are expressed as 1-RPC, where RPC denotes 'Regional Purchase Coefficient'. The construction of RPC is presented below.

---

[6] Given this base year calculation, exports to other regions in future years are calculated by use of parameters for rate of change per year. In the multiregional model in rAps these parameters are determined from balancing domestic trade.

The regional purchase coefficient, RPC

The RPC coefficient has the common magnitude (among others) as outlined by Stevens et al. (1989): It simply means the proportion of total demand or absorption, which has its origin from producing industries within the region itself. This is the crucial value to estimate and use when regionalising the account figures. The RPC may be expressed from a total or gross delivery perspective, or from a net perspective, depending on what kind of information is available.

The functional form of RPC is discussed in Annex 3. The chosen form is based on the assumption that the country consists of two regions that trade with each other, but not with the rest of the world. The following expression for RPC emerges as valid under some plausible restrictions (industry *h* suppressed):

$$RPC = \alpha Min\{1, P_r/C_r\}+(1-\alpha)(P_r/P_{tot})$$

The RPC is thus given by a weighted sum of the Supply Demand Ratio SDR (= P_r/C_r) and the share of national production, (P_r/P_{tot}). The RPC is given by the SDR, but scaled down with a factor α. The SDR ratio matters, but the relative size of the region matters as well. Note, it is here assumed that local supply is proportional to the production in the region.

The parameter α weights the SDR ratio with the size of the region, expressed as (P_r/P_{tot}). This parameter may vary with region. However, if no prior information exists, then one would be left by setting the parameter α to a fixed value. Even with a fixed α weight the expression has certain inbuilt properties that are advantageous compared to approaches based on either SDR or relative size.

Suppose two regions both produce more than they consume ($P_r/C_r > 1$ for r = 1, 2). If one region is larger than another ($P_1 > P_2$), then the smaller region would get a lower RPC value (as long $\alpha < 1$). The RPC value is therefore varying with the size of the region, with a lower RPC value for smaller region.

Regional purchase coefficients have as their advantage the extremely low data requirements. The RPC methods are best classed in line with model based approaches (SCGE). From limited information about the variables of interest, and from assumptions about trade patterns, a subset of the variables of interest is obtained. Production in the region is used to estimate the fraction of consumption that is satisfied by local production.

4.1.2 The row constraint, production P

As for row totals, (P), production per industry and municipality is in accordance to corresponding data in National Accounts and Regional Accounts. These data from rAps can thus be used directly as input to subsequent construction, without any adjustments.

4.1.3 Foreign trade

Before we turn to the column constraint C, the calculation of exports and imports should be presented since the determination of foreign trade is of importance for the final calculation of C.

At the industry level (h) validation of rAps base year data for foreign trade can only be made with respect to national totals. To make use of foreign trade statistics (FTS) a conversion to commodities (k) has to be made, see section 4.2

4.1.3.1 Exports

Since exports in rAps is a distribution of national totals, by use of production per industry, it goes without saying that the distribution by municipality is in agreement with national totals. For several reasons it seems, however, to be a very strong and uncertain assumption to make exports from industry (h) strictly proportional to production in industry (h).

The basis for an alternative distribution of exports has to be based on information at the commodity level, to be discussed below in section 4.2. Whatever options this offers we assume that an altered distribution of exports at the commodity level is fed back to the industry level.

4.1.3.2 Imports

There is at present a discrepancy between national totals for imports and the resulting totals from calculation in rAps. For one thing the fixed import shares determining imports by industry have to be updated.

But even if these parameters are corrected the same kind of uncertainty in using fixed parameters by industry (h) applies to imports, as well as for exports. Information at the commodity level must be used to get a more reliable regional distribution of imports. In the same way as for exports we assume that an altered distribution is fed back to the industry level.

4.1.4 The column constraint, consumption C

As stated above, there are no data sources for regional data on consumption, and therefore data on C must be based on model calculation. This model calculation is performed in two steps.

First, as consumption cannot be assessed as an aggregate, it but must be considered by its integral parts. Thus, calculation is made with respect to intermediate consumption in the production system, and private consumption and investment in final use[7]. Second, this resulting consumption total must be calibrated to achieve consistency in the regional accounts for all regions.

[7] Public consumption is part of final uses in National Accounts, but since it only concerns consumption of services it will be disregarded. We will, however, take the intermediate consumption for producing public services into account.

4.1.4.1 Calculation of consumption components

4.1.4.1.1 Intermediate Consumption

The calculation of intermediate consumption is based on the assumption of fixed input-output technology, i.e., the production functions do not vary by region. At present intermediate consumption is calculated from coefficients based on national input output tables for year 1995. These parameters should be adjusted, according to the new input output tables for year 2000.

4.1.4.1.2 Private Consumption

The calculation of private consumption is, in a similar way, based on the assumption that consumption functions do not vary by region. As to the regional distribution of total private consumption, various tests have compared micro and macro based approaches, based on data from National Accounts and the Household Budget Survey (HBS). These calculations indicate that the present consumption function, using disposable household income, is satisfactory as basis for regional distribution of *total* private consumption. Given the current purpose, however, a number of adjustments are called for.

First, there is some regional variation in the consumption pattern that ideally should be taken into account[8].

Second, the definition of regional private consumption is not exactly corresponding to the appropriate definition in this context. The definition of private consumption aimed at is concerning the region where households make their expenditures for consumption, quite irrespective of the household is living in the same region or not. In terms of regional economic variables this means that the net export of consumption goods should be added to the calculated private consumption. This net export in region r is defined as the consumption in r by residents living outside r, minus the consumption by residents in r taking place outside r. To get this adapted variable for private consumption it seems necessary to make use of auxiliary data, e.g. turnover figures from retail trade. It has to be further considered how a variable of this kind should be combined with the original consumption variable, with respect to necessary segmentation.

4.1.4.1.3 Investment

In rAps, base year data for gross fixed capital formation (investment) is based on a distribution of national totals by use of production per industry and municipality. At present there is an underestimation due to imported investment goods being underestimated, as mentioned in section 4.1.3.2 above.

---

[8] For example, according to the Household Budget Survey (HBS) Housing expenditure as share ot total household consumption is varying from around 25 percent in the metropolitan regions (Stockholm, Gothenburg, Malmö) to around 20 percent in the sparsely populated regions.

4.1.4.2 *Calibration of C for consistent regional accounts for industry h*

Assuming that production (P), export ($V_{rs'}$), import ($V_{s'r}$) and RPC are known (exogenous), the calibration of consumption (C) is made with respect to the domestic trade balance. Domestic (interregional) export and import are defined

$V_{rs}^{h} = P_{r}^{h} - V_{rs'}^{h} - RPC_{r}^{h} * C_{r}^{h}$, and

$V_{sr}^{h} = C_{r}^{h} - V_{s'r}^{h} - RPC_{r}^{h} * C_{r}^{h}$, respectively.

Starting from the calculated C_r as the a priori value, we are looking for the adjusted C_r* which satisfies the domestic trade balance

$\sum_{r} V_{rs}^{h} = \sum_{r} V_{sr}^{h}$.

This calibration procedure for individual industries is relatively simple. The adjusted C_r* should satisfy that the national total, $\Sigma_r C_r$*, equals national total of production minus national exports plus national imports. However, the adjustment of C_r implies an adjusment of imports as well, as imports is based on the level of consumption.

4.1.5 Constraints at the industry level, in sum

Production is based on data and can be used directly in the construction. Consumption is based on model calculation and the result from using the present source (rAps) should be modified. There are three consumption components:

(i) Intermediate consumption , based on the input-output tables. This component should be modified by updating the input-output tables to be consistent with the national input-output table from year 2000.

(ii) Private conumption, based on a consumption function. This component should be modified with respect to consumption as defined by retail trade.

(iii) Investment. This component should be modified to be consistent with survey-based data in Regional Accounts.

Given the model based estimate of consumption, a calibration can be made by use of the national total for consumption.

4.2 Conversion from industries to commodities

The National Accounts includes a *supply table* (also called *make table*) which gives information about product x industry, that is which products *k* being *supplied* by which industries *h*. This table thus provides a conversion from industries to commodities, at the national level.

However, the product mix within industries is varying between regions. To get P at the commodity level it is therefore necessary to get regional data on the production of commodities for enterprises and their local units. This data is provided by regional statistics on the production of commodities and industrial services, IVP. Foreign trade statistics, FTS, provides corresponding figures on export and import at the same commodity level.

To get regional consumption C at the commodity level there is no data source available (except for imports by using FTS). Thus, converting C(h) to C(k) we have to assume that the national conversion, from the national input-output table, is valid also at the regional level.

4.2.1 IVP

The conversion of production from industries (h) to commodities (k) will make use of the IVP, which gives yearly data on value of production on a very detailed product level (KN8). By using this kind of micro data a direct conversion between unit (h) and product (k) will be provided, which take the regional variation in product mix into account.

With respect to IVP the tasks to be dealt with concern, e.g., the regionalisation procedures in the case the survey concerns multi-plant companies, representation of units less than 10 employees (not included in the surveys/estimation).

As IVP covers commodities from manufacturing activities, additional information for other goods producing sectors, i.e. Agriculture, Forestry and Fishing, will be evaluated through the respective national authority, which also are in charge of the corresponding data.

4.2.2 Foreign trade statistics

Foreign trade statistics, FTS, provides information on bilateral exports and imports at a detailed commodity level. In the present Foreign trade model (FTM) in SAMGODS model system, the regional distribution of exports and imports within countries is made by general and simple principles. This regionalisation has not been applied. For example, exports and imports of commodities classified as goods for intermediate consumption and investment goods are distributed according to total regional employment in manufacturing industries. Exports and imports of commodities for private consumption are distributed according to disposable household income. This kind of regional distribution seems too simplified to be of general use.

It is appropriate to see the regional distribution of exports and imports as one task for Swedish regions and another task for regions abroad.

As for Swedish regions it is, according to Statistics Sweden, possible to make a distribution of exports from municipalities, by combining several registers of data.

For manufacturing commodities the approach means that data at a very detailed commodity level from IVP and FTS are linked to data on enterprises from Business Statistics.

For products from agriculture and forestry, data on production and exports for detailed product groups will in a corresponding way make it possible to distribute exports to municipalities within Sweden. The respective national authority already provides a distribution of exports from larger Swedish regions.

This approach was tried some years ago, but without paying notice to the problem of wholesalers being exporters[9].

The distribution of Swedish import of commodities to municipalities must handle the same problem as when converting consumption from industries to commodities. Therefore it seems natural to distribute import (k) according to calculated regional import (h), using the conversion (h) – (k) for imports at the national level.

As for regions abroad the distribution of exports and imports will be made in accordance with available (published) data, where the present suggestion in FTM is the default alternative.

A final comment is concerning foreign trade statistics for possible transit traffic through Sweden. Such matrices have been produced, in two alternatives: ”Transit freight matrices for Sweden”, LT Consultants and Matrex, March 2004.

We also assume that these matrices finally will be provided as a specific task, mainly external to the generation of base matrices. To that extent the transit traffic will be represented in the PWC matrices, cooperation is however called for.

---

[9] For example, as for export of cereals, enterprises registered as exporters are often located in other (metropolitan) regions than the regions from which the export originates. In this case, by using information and expertise from the department of agricultural statistics it is possible to locate the regions from which the export originates.

In this context it could be mentioned that the new computerised transit system (NCTS), a computerised follow-up system to be used in transit procedures, may be of interest as a data source for actual transit traffic.

4.3 Row and column constraints in the PWC case

In the following we will discuss the consequences for the construction of constraints in the PWC case. To see the implications clearly it may be helpful to first give some background information concerning different value measures used in National Accounts.

The supply and use tables produced by Statistics Sweden are an integrated part of the National Accounts, as a central tool for calculation of GDP from the production and use side, respectively. The *supply table* gives information about product x industry, that is which products being *supplied* by which industries. The total supply of product *k* is given by the sum of production of *k* in all industries and imports of *k*. This total supply is valued at *Basic price*, which means the price that the producer gets[10].

The *use table* also gives information about product x industry, but now meaning which products being *used* by which industries - used in intermediate consumption by industry *h* and used in final consumption. This total use is valued at *Purchasers price*, which means the price the consumer pays. The difference between basic price and purchasers price is the *trade margin*, which is the value of wholesale and retail trade production[11].

In addition to the supply and use tables Statistics Sweden also produces a symmetric *input-output table*, derived from the supply and use tables. This input-output table gives information about product x product, meaning the supply of *k* being used - as intermediate consumption in production of commodity *l* and in final consumption of product *k*. The input-output table is calculated for total supply, for domestic production and for imports. In this input-output table the valuation is at basic prices. This means that all trade margins have been moved from the products where they are included in the purchasers price to the product which refers to trade margins, namely wholesale and retail trade products.

The regional input-output tables in rAps are corresponding to the national input-output table with respect to valuation, i.e. also the regional tables are at basic prices[12]. Thus, from the input-output table in National Accounts and its regional

---

[10] In Swedish: Producentpris.

[11] It could be noted, however, that imports are valued cif (freight included). This means that for imported commodities a part of the trade margin, namely the part generated by trade from the foreign country to the Swedish border (= the importing enterprise), is already included in the basic price.

[12] However, the symmetric regional input-output tables are calculated as industry x industry.

counterpart in rAps the valuation of both P and C is the same, i.e. at basic prices. The fact that commodities do not change in value provides a good foundation for the construction of base matrices in weight, where the same condition (unchanged commodities) applies.

How does an incorporation of wholesale (W) into the base matrices fit in? Obviously it cannot be made in a meaningful way, if W is valued at basic prices. As we have seen the production of 'wholesale products' is the production of trade margins, i.e. production of services and not commodities.

To get a PWC matrix that is corresponding to P/C, with respect to flows of commodities, the approach could be as follows:

- The consumption in the wholesale sector is the flow of commodities from P to W, i.e. the value of commodities purchased by W from P, at basic prices.
- The production in the wholesale sector is the flow of commodities from W to C, i.e. the value of commodities sold by W to C, also at basic prices.

The information easily available to illustrate these flows are the following for the wholesale and commission trade, SNI 92=51, MSEK year 2001:

Net turnover	870 165
Production (trade margin)	214 847
Difference	655 318

The difference between net turnover and trade margin represents the value of traded commodities at basic prices. This value of traded commodities is the base for incorporating wholesale in the PWC matrices.

Using the same notation as above, incorporating wholesale means that the following constraints will be used in construction of the PWC matrices, commodity k:

Row constraint is equal to total production of commodity *k* in municipality *r*, plus the value of commodity *k* sold by wholesale in municipality r :

$$P_r^+ = \sum_s V_{rs} + \sum_s W_{rs}$$

Column constraint is equal to total consumption of commodity *k* in municipality *s*, plus the value of commodity *k* purchased by wholesale in municipality *s* :

$$C_s^+ = \sum_r V_{rs} + \sum_r W_{rs}$$

Since the value of traded commodities is at basic prices, the value of sold and purchased commodities is the same, and thus the following identity holds for a municipality where the wholesaler is located:

$$\sum_s W_{rs} \equiv \sum_r W_{rs}$$

From Business Statistics at SCB it seems possible to get data on traded commodities for enterprises, which should concern not only enterprises in the wholesale industry but other (primarily manufacturing companies) industries with trade activities as well.

Assuming that traded commodities refer to the same category as the NACE code of the trading enterprise, whether wholesale or manufacturing, it will be possible to expand the P and C constraints in the way described.

The main argument in favour of PWC is the fact that CFS can be used as it is defined in the survey 2001, i.e. without information of receiving activities. The main argument against PWC is that the location of wholesale is made exogenous.

5 A priori matrices

5.1 Introduction

The final purpose is to construct Base matrices of freight flows between all municipalities in Sweden and between municipalities in Sweden and the 174 STAN99 regions abroad. The matrices will be estimated in weight terms for a specific base year for the 12+ commodity groups in STAN. The base matrices could also be given in value terms, as the value/weight relation is important for the logistics modeling.

5.2 Data sources

For estimation of the base matrices the following main data sources are used: The commodity flow survey (CFS), the yearly produced statistics IVP, Foreign trade statistics (FTS) and national input-output matrix for the specific base year.

The Swedish commodity flow survey (CFS) is a key data source for infilling the individual cells in the P/C-matrix. At present there is a survey from 2001, CFS 2001. The plan is to make a new survey every third year and the CFS 2004/05 is now ongoing.

The present CFS 2001 is a sample survey of the manufacturing, mining and wholesale sectors. In addition data for transports of forestry and logging products, dairy-farming and sugar-beet were obtained from register sources. Reported data on individual shipments include industry sector, commodity type, origin and destination, value and weight.

The sample survey is a stratified three stage random sample of shipments, where strata are defined by STAN commodity group produced, size of local unit by number of employees and geographical location by NUTS II in Sweden[13]. The stratification is designed to get good estimates of totals on values and weights.

For outgoing shipments the municipalities for origin and destination of the shipment are given for national shipments and foreign destination by STAN99 regions for export. For incoming shipments, which are imports, the origin is given by STAN99 regions abroad and destination by municipality in Sweden.

The shipments are given by commodity at the CFS-level; 81 in total of which 79 are represented in the database. The shipments are also given by industry sector (SNI/NACE) for the sending firm but unfortunately not for the receiving one. This put limits on the use of the present CFS2001 for pure P/C construction.

[13] NUTS II = 'Riksområden' in Swedish

The base matrix for commodity *k* is denoted Q_{rs}^{k} for quantities in weight terms and V_{rs}^{k} in value terms, where *r* and *s* are municipalities (NUTS V regions) in Sweden and STAN99 regions abroad.

On an aggregate regional level, *R* and *S* are here notations for NUTS II regions in Sweden and countries abroad.

The matrices that are input to the logistics model are for STAN commodity groups, denoted *K*. When constructing the a priori matrices we deal with more disaggregated commodity groups, denoted *k*.

5.3 Constructing the a priori matrices

The a priori matrix for commodity k is according to the suggested approach a combination of two different matrices:

A) Matrix elements based on direct observations from CFS;

B) Matrix elements based on a gravity model approach, using CFS, row and column constraints and data on transportation costs (generalised cost from STAN). The diagonal elements are calculated by the regional purchase coefficient (RPC) technique. (For a discussion about applying RPC at the municipality level, see section 5.3.2.1 below.)

In the following the respective type of matrix will be denoted by top index A and B.

5.3.1 A: Direct observations from CFS

Among various sources of data suitable for making the basis of an á priori matrix, the observations from CFS is obviously of central importance, irrespective of the relatively poor representation at the municipality level. Since the pure P/C case and the PWC case put different requirements on CFS data, the two cases are discussed separately.

5.3.1.1 The pure P/C case

The matrices $Q_{rs}^{k,A}$ and $V_{rs}^{k,A}$ could be estimated quite straightforward from CFS data, provided information on industry sectors for incoming as well as outgoing shipments are available. A problem with the current CFS 2001 is that it includes information on receiving activity only for international incoming shipments.

To be able to use CFS directly in the construction of pure P/C-matrices some alterations in CFS thus have to be made.

One critical issue is to get the industry sector for the receiver; at least it should be possible to separate the wholesale sector. But even if this information on the receiving activity is available, the problem has to be solved as to making a flow from P to C via shipments to and from wholesale activities.

A way of dealing with this problem can be outlined as follows. Let us use some temporary notations for activities in different locations (municipalities):

p	production
c	consumption
w	wholesale

We are assuming that shipments *X* registered in CFS refer to a specific commodity *k*, which is implicit below.

Integration of shipments from/to wholesale in the construction of the flow P/C, where *X* represents either *Q* or *V*:

$$X^{A}_{pc} = X_{pc} + 0{,}5\sum_{w} X_{pw} \left(\frac{X_{wc}}{\sum_{c} X_{wc}}\right) + 0{,}5\sum_{w} X_{wc} \left(\frac{X_{pw}}{\sum_{p} X_{pw}}\right)$$

This solution is based on information on three kinds of shipments, 'pure' p-c, shipments w-c, and shipments p-w[14]. The p-w shipments are distributed forwards to c, in proportion to the share of c of all shipments received from w. In the same way, w-c shipments are distributed backwards to p, in proportion to the share of p of all shipments sended to w.

It is expected that forthcoming CFS will identify activities both for incoming and outgoing shipments, in order to serve as basis for 'pure' P/C[15].

It is still a survey, though, and the estimation of the a priori matrices is suggested to be sequential. The most reliable level is NUTS II, elements in the CFS matrices are weighted to get reliable values for NUTS II. This means, however, that some municipalities are weighted quite high, as there are many municipalities, which are not included in the sample.

Here we will make some comments on "sparseness" in the observed matrices and on reliability of data.

- There are 888 thousands outgoing shipments registered in the CFS 2001 database, of which 552 thousands from the wholesale sector.
- Shipments from wholesale are registered for 75 CFS-commodities, for other industries shipments are registered for 74 CFS commodities. This means that most of the commodities are registered both for wholesale and other sectors.

[14] This solution is based on the assumption that shipments from p to c can be treated as 'pure' p-c. However, it should be noted that wholesale activities are also taking place outside the wholesale industry.

[15] This kind of information will be available from the ongoing CFS 2004/05.

A full NUTS V matrix covering Sweden has 290*290 = 84100 elements. Looking at different commodities this number is relevant only in the case all municipalities are producing and consuming the commodity. For the 12 STAN groups the actual number of combinations of municipalities which have observations in CFS2001 is varying between 41 (Round timber) and 30129 (Manufactured articles). A small number of observed combinations is of course expected in cases where only few municipalities are producers and only a few are consumers.

But of course the relatively small number of observed combinations also reflects that for most commodity groups CFS is a sample survey.

As the sample is stratified to guarantee reliable values only at the NUTS II level we suggest that elements at the NUTS II level will be used as restrictions when using CFS-data as an apriori P/C matrix on the NUTS V level.

Over time there will be data from a number of CFS surveys; hence, the matrix will be less sparse. How to merge data from various surveys is described below.

5.3.1.2 The PWC case

The fundamental difference with the P/C case is that observations can be taken directly from CFS. There is no need to register the receiving industry sector.

5.3.2 B : Model based on CFS, transportation and logistic costs, and RPC

The second component in the combined a priori matrix is a synthetic matrix, based on CFS data and data from the logistics/network model on transportation and logistics costs. The elements of this á priori matrix would then be made up of exponential cost functions, served by data which initially could be provided from STAN, and successively refined by additional cost elements from the logistics module

The approach could be presented as follows[16]. Although the approach may be generalised to deal with international as well as domestic flows, only the latter is presented here.

Separately for each commodity, international flows are excluded from the weighted CFS matrix X (could be either V value or Q quantity).

Define row and column constraints from CFS:

$X_{r.} = \Sigma_s X_{rs}$ (ie production), $X_{.s} = \Sigma_r X_{rs}$ (ie consumption), total X_{**} .

Estimate *d* to obtain best fit to observed data X_{rs} ($r \neq s$) assuming model form

[16] The presentation follows suggestions made by John Bates in his comments on a previous draft.

$$\hat{X}_{rs} = p_r . c_s \exp(-d\, g_{rs}) \text{ and } \Sigma_s\, \hat{X}_{rs} = X_{r*} - X_{rr},\ \Sigma_r\, \hat{X}_{rs} = X_{*s} - X_{ss}$$

(Typically this is done by equating mean cost between observed and modelled flows.)

For each municipality, calculate

$x_r = \text{Min}\ \{1, X_{r*}/ X_{*r}\}$,

$z_r = X_{r*}/ X_{**}$, and

$y_r = X_{rr}/ X_{*r}$

estimate α in Eq $y_r = z_r + \alpha.[\, x_r - z_r] + \varepsilon_r$ (eg by minimising $\Sigma_r\, \varepsilon_r^2$).

Set $X^B_{rr} = X_{*r}.\ (z_r + \alpha.[\, x_r - z_r])$, i.e $X^B_{rr} = X_{*r}.RPC$.

Subtract X^B_{rr} from X_{*r} and X_{r*} to give new "reduced" row and column constraints X^B_{*r} and X^B_{r*}.

Using previous *d*, estimate X^B_{rs} $(r \neq s)$ assuming model form

$X^B_{rs} = p^B_r .\ c^B_s \exp(-d\, g_{rs})$ such that $\Sigma_s\, X^B_{rs} = X^B_{r*}$, $\Sigma_r\, X^B_{rs} = X^B_{*s}$

We think it is important to make these estimations at a commodity level as detailed as possible, in order to take different value/weight relations into account. There will thus be specific *d* parameters for weight and value.

A complication in case of pure P/C matrices is the fact that most of the shipments pass a warehouse, and thus we do not know exactly the true shipments between producing and consuming municipality. The flows have to be aggregated over wholesale as suggested in A.

5.3.2.1 RPC at municipality level

According to the approach B, the diagonal elements are estimated by use of the RPC technique. Since RPC is normally thought of as being applied to "regions", a comment is needed with respect to its intended application at the municipality level.

Generally, small regions will naturally be more open than bigger ones, and the level of self-sufficiency will be lower. At the same time, certain production may be part of value chains which are "clustered" in the area, and then the at the regional (or municipality) level, intermediate deliveries may have a high local concentration, more or less independent of the geographical demarcation. However, while this may occur for intermediate deliveries, or even investments, this will normally not be the case for consumption goods delivered in the area.

In any case, the RPC method is not a method specially developed for "big" areas. The core of the method is the focus on the municipalities demand for a certain commodity, as opposed to the municipality's ability to deliver the same commodity. For practical use, the difficulty lies in whether a sufficient detailed or disaggregation level of the traded commodity in question may be possible or not. In case not, we have the problem of non-homogeneity to face, and adopting RPC functional forms (not RPC values) from larger regions/municipalities on smaller geographical units may be inaccurate. However, available data from IVP on detailed commodities produced in each municipality should be a necessary condition for obtaining acceptable results.

Surveys on regional and interregional intermediate deliveries in small municipalities in Norway (from 1975) show large values of variance at the element level within the same municipality/region. In the case with the base matrices, the intermediate elements are not relevant, only the flows aggregated over all kinds of consumption in each municipality. When other sources for estimating such flows are dubious, our assessment is that the RPC method (RPC functional form) is suitable for establishing an a priori level of the intra-municipality flow for each commodity.

5.3.2.2 RPC in the PWC case

The RPC formulation is based on estimating the share of consumption that is covered by local production. The PWC formulation, on the other hand, replaces production with supply (where supply may be from production or from the wholesale sector) and demand (where demand may be from production or from the wholesale sector).

This reformulation does not alter the theoretical underpinnings for using the RPC framework as long as supply/demand is formulated on a goods by goods level. However, it may change the practical utility of the proposed framework.

The practical problem arises from strongly localized wholesale structures (distribution centers). These centers would get a large value for P_r/P_{tot} due to the wholesale unit located in the municipality. This would then "force" the diagonal element to be high, implying that a large fraction of goods demanded locally also are supplied from local suppliers (be they wholesale or producers).

This may, or may not, be a reasonable assumption, but one would expect a large amount of cross-hauling even if a distribution center is located in the municipality. It may therefore be necessary to estimate different values for α, one for municipalities with distribution centers, and one for those without a distribution center.

In short, including the wholesale sector exacerbates aggregation problems in estimation of α since supply now is a "mix" of production and supplies from the wholesale sector, while demand is a "mix" of consumption and demand from the wholesale sector. Thus even more care has to be taken in the estimation of α, especially since the PWC formulation is rather unorthodox, at least from a regional economics perspective.

5.4 Combining á priori matrices

The two a priori matrices can be used to estimate the final á priori Base Matrix, by using a two-step entropy maximisation procedure. See Eriksson (1988).

$$X_{rs}^{K} = \alpha * X_{rs}^{K,A} + \beta * X_{rs}^{K,B}$$

The coefficients α, β have the sum equal to 1 and are decided through the restrictions. The a priori matrix fitting the restrictions best gets the highest weight as described in the following.

If we have different candidates for á priori matrices, constructed from different statistical sources in such a way that we have statistical measures of the uncertainty of these á priori matrices, we can use the statistical measures to combine the candidates. This combination can be used as a new á priori matrix.

Another technique for combining the candidates is to simultaneously compute combination of candidates and the entropy solution instead of the two steps procedure see Eriksson (1988). This approach (the simultaneous two-step entropy model) has been successfully used for modelling housing consumption for Stockholm.

The method can be described as follows:

For a given number of candidates for á priori matrices *XO, YO,…,ZO*, define a new á priori matrix $X0^{**}$ as the linear combination

$$X0^{**} = \alpha_{XO} XO + \alpha_{YO} YO + \ldots + \alpha_{ZO} ZO$$

where $\alpha_{XO}, \alpha_{YO}, \ldots, \alpha_{ZO}$ are parameters, which should be computed to give the highest possible entropy value subject to the constraints to the original entropy problem. This could also be apprehended as that we try to find a á priori matrix the pattern of which coincides as close as possible with the pattern given by the constraints. The method is closely related to the average data model and the Plane model, see Eriksson (1988).

6 Estimating base matrices by entropy maximisation

6.1 Estimation of base matrices with data of different reliability

The solution of the entropy model is a smooth correction of the á priori matrix such that the corrected matrix (solution matrix) should satisfy a set of given constraints. Now the problem is: the constraint structure (left hand side of the constraints) but the data (the right hand side of the constraints) is not fully reliable. Indeed, the elements in the data have different reliability. The constraints for Production (P) are more reliable than the constraints for Consumption (C). We will here, by introducing the concept of soft constraints, show how this problem can be handled.

6.1.1 Soft constraints

We will here introduce the concept of *soft constraints*. (Confer e g Hovland, 2004). If we formally add the soft constraints $Cx \cong y^0$ (where y^0 is an á priori right hand side of the constraints $Cx = y$) to the equality constrained entropy problem we get

$$\min_{x} \sum_{i=1}^{N} x(i) \times \ln(x(i)/x^0(i)) \quad - (x(i) - x^0(i))$$

$$\textit{subject to } \; Ax = b, \quad x \geq 0$$

$$\textit{and } \; Cx \cong y^0, \quad y^0 \geq 0$$

(In this example the index *i* is used just as an example).

Mathematically, by this notation, we mean

$$\min_{x,y} \sum_{i=1}^{N} x(i) \times \ln(x(i)/x^0(i)) - (x(i) - x^0(i)) +$$

$$\sum_{i=1}^{M} w(i) \times \left[y(i) \times \ln(y(i)/y^0(i)) - (y(i) - y^0(i)) \right]$$

$$\textit{subject to } \; Ax = b, \; x \geq 0 \quad \textit{and } \; Cx = y, \; y \geq 0 \quad (w > 0)$$

Some remarks: The non-negative restriction $y^0 \geq 0$ is not a real restriction. If some $y^0(i) < 0$ we can multiply the *i*:th constraint by -1. By using individual weights *w(i)* it is possible to handle different constraints having different reliability. If we have an entropy problem with mixed equality and inequality constraints we define the soft constraints in the same way as for the equality constrained case.

6.1.2 Soft Constrained Base Matrix Model

The model is not only modified by introducing soft constraints. It is defined both for weights ($X \equiv Q$) and for values ($X \equiv V$). The model can now be defined as an entropy model with:

- The Production (P) as row constraints
- The Consumption (C) as soft column constraints
- CFS (the matrix) at the NUTS II level as soft CFS constraints
- Export /Import at the NUTS I level as export/import constraints[17]
- Individual lower (L) and upper (U) bounds for each element in the matrix X

The use of CFS directly is not possible without introducing soft constraints. The á priori matrix

$$X0 = \begin{pmatrix} X0^{r,s} & X0^{r,s'} \\ X0^{r',s} & 0 \end{pmatrix}$$

consists of three sub matrices, the 290×290 á priori matrix $X0^{r,s}$ describing domestic flows, the 290×174 á priori matrix $X0^{r,s'}$ describing export and the 174×290 á priori matrix $X0^{r',s}$ describing import.

The elements in the matrix $X0$ are denoted by $X0(r,s)$ and correspondingly for the sub matrices. E.g., this means that $X0^{r,s'}(r,s') \equiv X0(r,290+s')$.

Now, the entropy model can be formulated in **value terms** as:

$$\min_X \sum_r \sum_s X(r,s) \times \ln X(r,s) / XO(r,s) - (X(r,s) - XO(r,s))$$

$$\textit{subject to } \sum_{s=1}^{290+174} X(r,s) = P(r)\,,\ r=1,...,290$$

$$\sum_{r=1}^{290+174} X(r,s) \cong C(s)\,,\ s=1,...,290$$

$$\sum_{r\in R}\sum_{s\in S} X(r,s) \cong CFS_{II}(R,S)\,,\ R=1,...,9,\ S=1,...,9,\ (R,S) \neq (9,9)$$

$$\sum_{r=1}^{290} X(r,s') = X_{EX}(s')\ ,\ s'=1,...,174$$

$$\sum_{s=1}^{290} X(r',s) = X_M(r')\ ,\ r'=1,...,174$$

$$0 \le L \le X \le U \le \infty \quad \textit{(or any mix with } 0 \le L \approx< X \approx< U \le \infty\textit{)}$$

The total production in NUTS V region *r* is denoted *P(r)*. The right hand side of the linear row constraints (production) are considered as accurate information

---

[17] In reality, a modification of NUTS I: an element could be (besides a country) a part of a country or an aggregation.

and hence we let the row constraints be equality constraints. This holds in both value and weight terms.

The total consumption in NUTS V region *s* is denoted *C(s)*. This quantity, *C(s)*, has been computed by modelling and cannot be viewed as fully true information. Hence the right hand side of the linear column constraints (consumption) is considered as inaccurate information and we let the column constraints be soft equality constraints. ***C(s)* is only given in value terms**. In weight terms the restriction has to be reformulated, see below.

The total flow between the NUTS II region *R* and the NUTS II region *S* derived from CFS is denoted by $CFS_{II}(R,S)$ both in value and weight terms.. In our definition the NUTS II region number 9 consists of all foreign regions. Since the CFS is a sample survey the estimated value of $CFS_{II}(R,S)$ is only an approximation of the true value. Hence the right hand side of the CFS constraints (Commodity Flow Survey) is considered as approximate information and we let the column constraints be soft equality constraints.

There is also a possibility to put constraints on the individual flow between the NUTS V region *r* and the NUTS V region *s*. Such constraints could be derived from UVAV, CFS or from other sources as e.g. railway statistics. A lower bound is denoted $L(r,s)$ and an upper bound it is denoted $U(r,s)$. If it is considered as approximate information we let it be a soft constraint. Some restrictions are in weight terms whereas others are in value terms.

Note that the soft constraints can be weighted according to their reliability.

If information were available, it would also be possible to put reliability values on each element in the matrix.

In this context it could also be mentioned that there are other ways of dealing with the reliability problems. In a paper on estimation and accuracy of OD highway freight flows (Metaxatos, 2003), gravity models have been estimated for both weight and values and confidence intervals have been estimated using a Poisson distribution model.

However, in our case the confidence intervals could be quite difficult to handle in the base matrices. We will instead use the inverted values of the variances as weights for different data sources. If no statistical variances are given, subjective weights have to be given.

6.2 Value/weight and consumption restrictions in weight terms

The aim is to construct commodity flow weight (e.g. tonnes) matrices. Economic models and statistics are more often using value (e.g. Sw. Cr.). There is thus a need of knowledge about weight-value relations.

We can reasonably assume that there is an individual weight-value relation for each cell in the matrix, as the mix of specific commodities in the commodity group is different in each P/C relation. This means 290×290 domestic, 290×174 export and 174×290 import weight-value relations, totally 185 020 weight-value relations, for each commodity group.

First we will remark that the lack of data is considerable. Consumption data is poor compared to data for production, export and import. There is also some weakness about weights in production data (IVP), but this problem could probably be handled by imputing weight data from FTS into IVP.

The real problem is the lack of data about weight-value relations for each P/C relation.

In this context it should also be stated that we do not envisage an exclusion of high value goods (associated with high probability for air transport) from the P/C concept.

For the transportation model weight is needed, for the logistics model the relation between value and weight is also important. The values for both production and consumption are derived at basic prices. It would be desirable to have both weight and the relation between value and weight at the NUTS V level.

Value/weight at NUTS I

In the former VTI/TPR demand forecast model[18] in SAMGODS the value/weight relation, denoted *z,* was computed at the NUTS I level, i.e. one value within Sweden, one for export and one for import for each commodity group.

Value/weight at NUTS II

As we now have CFS we can, in the corresponding way, do the same at the NUTS II level:

$$z(R,S) = V^0(R,S)/Q^0(R,S), \quad R=1,\dots,8,\ S=1,\dots,8$$

For foreign trade the value/weight relation is taken from the foreign trade statistics and using the same value for all NUTS II regions in Sweden.

This will give more differentiated value/weight values.

[18] See Eriksson (2000)

With these value/weight parameters *z(R,S)* and *z(imp)* (value/weight for import) the consumption restriction in the entropy maximisation problem can be written

$$\sum_{r=1}^{290} Q(r,s)/z(R,S) \cong C_{national}(s), \; s=1,290$$

$$\sum_{r'=1}^{174} Q(r',s)/z(imp) \cong C_{imp}(s), \; s=1,290.$$

The subindex *national* means consumption of commodities that are produced in Sweden, *imp* means import.

To get the base matrix in weight terms the soft constraints of consumption have to be weighted constraints, but is easily handled in the entropy maximisation program.

6.3 Handling of data for different years

The problem is to construct a base matrix for a specific base year when data are missing, e.g. from the CFS, for this specific base year but when we have CFS data for a different year.

As described above Statistics Sweden makes every year a survey, IVP, of production of commodities and industrial services. For most products, quantities and values of the deliveries over the year are reported. For some products in the raw material based industry, the total production quantities are reported.

One alternative is to directly use the IVP for the base year to make an update of CFS from the different year to the base year. Probably we will lose a lot of CFS information by doing an update in this way, in principle; IVP would replace some CFS data and hence "destroy" the CFS. In other words, by doing the update as suggested here, the result will be a mixing of CFS and IVP data and not an updated CFS.

The alternative, as we prefer here, is to use the IVP to compute the changes (value and weight) from the different year to the base year and then apply the changes to the CFS. The shipments in CFS are given for 81 different commodity groups. The geographical area level in CFS is NUTS V. The quotients should be computed from IVP at a CFS commodity level and for a chosen geographical level.

For export and import in CFS there is an alternative to use FTS instead of IVP. The technique does not differ from the one described here for IVP.

For updating data for a base year by using data for earlier years, see the following section on Procedure for merging sample survey matrices. One important part in this section is geographical updating (updating is dependent of the P/C relations).

6.4 Procedure for merging sample survey matrices

There are two different sample surveys that are used here. Of major importance is CFS (Commodity Flow Survey) and of minor importance is UVAV (Swedish national and international road goods transport). We will here describe a method for updating CFS similar with one that has been used for UVAV (Swedish domestic road goods transport).

We suppose that the structure of CFS is given by the latest survey. We have to do adjustments in the old surveys such that they coincide with the latest one (e.g., how the wholesale sector is handled). When the necessary adjustments have been made we construct P/C matrices on the NUTS V level and for each commodity group for both value and weight for each CFS directly from CFS data.

Let CFS^t denote the P/C matrix for the (adjusted) old survey at the time *t* and let CFS^T denote the P/C matrix for the latest survey at the time *T*.

The solution to the entropy problem

$$\min_{CFS} \sum_r \sum_s CFS(r,s) \times \ln CFS(r,s) / CFS^t(r,s) - (CFS(r,s) - CFS^t(r,s))$$

$$\textit{subject to} \quad \sum_{r \in R} \sum_{s=1}^{290+174} CFS(r,s) = \sum_{r \in R} \sum_{s=1}^{290+174} CFS^T(r,s)\,, \quad R=1,\ldots,8$$

$$\sum_{r=1}^{290+174} \sum_{s \in S} CFS(r,s) = \sum_{r=1}^{290+174} \sum_{s \in S} CFS^T(r,s)\,, \quad S=1,\ldots,8,$$

will be denoted CFS_T^t and called a normalized *CFS* matrix from the time *t* to the time *T*. Applying this technique for all old *CFS* we get two sets (for value and weight, respectively) of normalized P/C matrices (normalized to the actual time for the last *CFS* at the time *T*).

To use all *CFS* we have to give each normalized CFS_T^t a weight β_t and then summarize $(\sum_{t=1}^{T} \beta_t = 1)$:

$$\overline{CFS_T} = \sum_{t=1}^{T} \beta_t \times CFS_T^t$$

There are some different possibilities to choose β_t , more or less correct and more or less simple. We will mention one here. Let $size(CFS^t)$ denote the number of observations (“hits”) for the original CFS at the time *t* for the selected commodity group. Then the weights will be

$$\beta_t = \frac{size(CFS^t)}{\sum_{t=1}^{T} size(CFS^t)}$$

The advantages with the technique to use several sample surveys, and put them together in a reasonable way, is that we get more dense matrices compared to if we only use one single CFS.

The same technique can be used for UVAV, but then we should apply it on a NUTS III level (counties) instead of a NUTS II level.

The third example would probably be the most interesting model. It gives an opportunity to vary production costs at the NUTS V level but still use a more aggregated figure on the consumption part, as the reliability of data is not so good for column sums. Maybe CFS could be of a good help here.

6.5 Iterations

The freight model starts with a base matrix, PWC or P/C. The logistics module transforms this matrix into an O-D matrix of physical freight flows. The O-D matrix is assigned to a traffic network by the STAN model giving link flows in weights and vehicles. STAN will also generate new generalised costs, which will influence the a priori matrix of type B, and the entropy problem should be resolved.

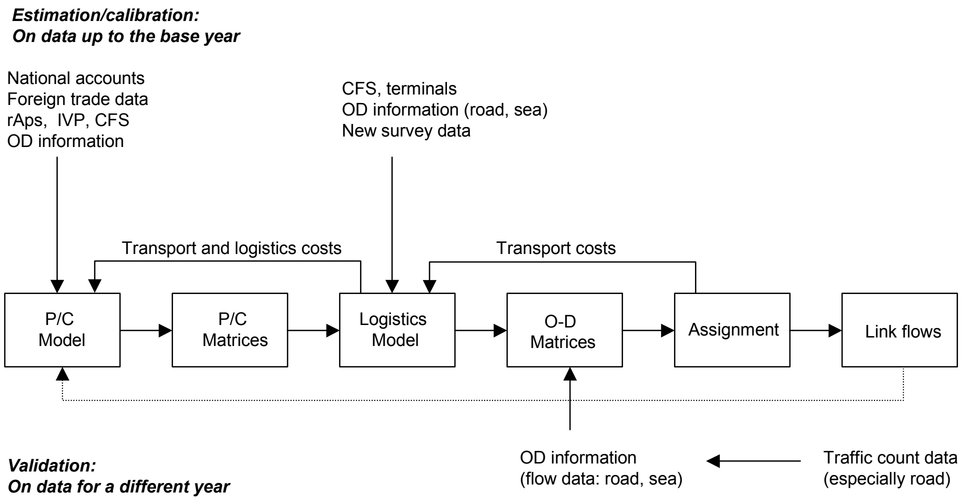
7 Validation and calibration

In Figure 3, the estimation, calibration and validation process for the Swedish and Norwegian national model systems for freight transport are depicted (the estimation data are above the boxes, the validation data are below the boxes). [19]

The P/C matrices are partly based on observations and partly synthetic (model-generated). The data used in this process come from the CFS, the regional economic system rAps (both only for Sweden), economic statistics from national accounts and foreign trade data, and from the questionnaire-based OD surveys (origin, destination, costs, commodity) for truck, rail and sea (in the current Norwegian procedure). The logistics model is estimated on the CFS (Sweden only), information on terminals, available OD information (e.g. on trucks, but also for other modes) and possibly on new survey data, and then transforms the initial P/C matrices into OD matrices.

When no disaggregate estimation data would be available or could be made available in the near future, there are some possibilities to use a normative logistics model, that optimises the logistics choices (see section 4.3) and to make this model more realistic by calibration it to aggregate data (OD information). One could use a normative logistic costs function, assume this works as a logit model with alternative-specific constants and a scale parameter, aggregate the initial results, compare with aggregate data and then adjust the constants and/or scale in an iterative process. This is clearly inferior to estimation on disaggregate data, but could provide a way out if these data would be unavailable.

Figure 3 Estimation and validation of the model systems



[19] This chapter is the same as in the report from the logistics project.

Programming and validation

The new logistics model needs to be programmed as part of the Samgods and NEMO systems, especially with interfaces to and from the P/C matrices and network models with definitions, segmentations and formats that match. This also includes the changes that might be needed in the network model and other model components to accommodate the new models.

In application of the logistics model, the P/C or PWC matrices are the starting point. Then for each commodity flow from the P zone to the C zone, one or more firms are generated to be the producer in the P zone and one or more other firms are generated to be the consumer (step A).

After that, in step B, for the total weight of the commodity-specific firm to firm flows, the shipment size is determined first. The logit models provide this in the form of probabilities for the different alternatives with regards to shipment size, which can be summed to give the expected frequencies. But the model application could be simplified by taking a draw from this distribution. This will give a shipment of a certain size of a certain commodity that is transported from firm m to firm n. In model option III, the shipment sizes are not produced by the model, but fixed shipment size distributions (by commodity type) are simply taken from existing data.

The model then assigns probabilities in terms of logistics chains, modes and vehicle sizes to this shipment. In model options II and III mode choice is not included here. These can also be summed or used to draw from. By repeating this procedure a sufficient number of times, we can cover all flows between the firms for the relevant commodities and in this way all flows in the P/C or PWC matrix.

In step C these are aggregated into flows of vehicles (options IA, IB) or tonnes (options II and III) by commodity type between the zones (in all options except III also by shipment size). There are passed on to the network model.

The model application process is iterative (see Figure 3, middle and upper part): after assignment, the new generalised costs matrices need to be used to adjust the P/C or PWC matrices, etc. This gives rise to an inner loop (the outer loop is described later in this section):

Inner loop:

1. The base matrix projects (Norway and Sweden) provide initial P/C (or PWC) matrices.

2. The logistics model transforms these into OD matrices, using transport cost provided by the network model.

3. The network model assigns the OD matrices to the networks.

4. The network model and the logistics model provide transport and logistics costs matrices to the base matrix projects.

5. The base matrix project calculates new base matrices on the basis of the new transport and logistics costs and provides these to the logistics model.

Etc.

This loop continues until equilibrium is reached between (in practice until a pre-set maximum distance from equilibrium is reached. The tasks at both ends consist of running the models as they are (no estimation required within this inner loop): the inner loop is about the adjustment of model variables (inputs and outputs), not model coefficients. What makes this model application loop time consuming is especially the transfer of matrices between the different models.

The validation process is depicted in Figure 3 as well (bottom part of the figure). As part of the validation, the predicted OD flows can be compared to the observed OD flows from OD surveys (by mode: road, sea, maybe rail), preferably for a different year (using the inputs of that year) than used in estimation, to make it an independent validation.

- To emphasise the independent nature of the model validation, it could be carried out by an independent model validation team: a group of potential model users and other experts.

The truck surveys provide information on domestic trucks. Information on foreign trucks can come from the traffic counts. Differences between observed and predicted can be due (apart from measurement errors in the observations) to both the P/C (PWC) model and the logistics model, and separating these is very difficult. Independent observed P/C (PWC) matrices would be needed for a separate validation of the P/C (PWC) matrices.

After the assignment of the OD flows to the networks, the predicted link flows can be compared to observed link flows from traffic counts (especially for road, maybe also for rail). If big discrepancies arise, these need to be analysed. Finally, parameters in all the models can be recalibrated (this requires an iterative procedure). This is the outer loop (the inner loop was described earlier in this section), which concerns different equilibrium situations for the inner loop. In the outer loop or model calibration loop, model coefficients in all constituent submodels are adjusted to reach a good match with aggregate data.

The final calibration (to counts and OD surveys) has to be done jointly by both teams. The methodology needs to be worked out, but it can be regarded as one big optimisation problem, that needs to be cut into several smaller sub-problems.

The objective of the overall optimisation is to minimise a function (e.g. the squared difference) of the distance between the observed and predicted values (traffic counts in particular). One parameter is modified at a time and then the minimisation function tells us which parameter change leads to the biggest reduction in the distance between observed and predicted. The parameters that will be changed first, are the ones that have the biggest effect on reducing the distances between the model predictions and the observations, which depends on the first derivatives of the overall objective function with respect to the coefficients (gradient search).

The overall optimisation problem is very complex and non-linear. It includes the P/C model, the logistics model and also the network model (e.g. calibrating a stochastic version of the network model, where the stochastics arise to represent omitted factors and measurement errors).

There is no simple recipe for such complex optimisation problems, they cannot be solved analytically, and iterative search procedures are needed. What often works reasonably well is a subdivision into smaller problems. This can for instance be done by extracting from the counts the number of crossings of some screenline (e.g. giving the observed amount of traffic between the North and the South of Sweden or Norway). These observations would not include the effect of route choice, so they can be used to calibrate the other models. Also information from OD surveys (as was described above) can be used to calibrate a problem that does not involve route choice. Observed fractions on specific routes can be used to calibrate the assignment. Observed fractions for the modal split and the use of consolidation and distribution centres can be used to calibrate the logistics model.

8 Design of implementation - summary

The presentation in this chapter mainly gives a summary and conclusions from previous chapters, as the suggested methodology for obvious reasons also has included a description of data needs, available data, and estimations procedures.

Decision on 'pure P/C' or PWC

Although the the report is based on the assumption that the PWC option will followed, a decision in this matter has to be taken before the development of base matrices can proceed.

Decision on base year

If the PWC option is followed it seems feasible to use CFS 2001 as a key data source, and thus 2001 as base year. The P/C option presupposes an amended CFS (information on receiving activities) and thus it seems as the base year has to be 2004, according to the ongoing CFS.

Defining new data and making adjustments of available data for row and column constraints

Since the rAps database and model system provide basic input to the construction of row and column constraints, adjustments are needed with respect to the demands for the specific purpose. For example, the regional input-output tables should be updated in agreement with the input-output national table for year 2000, the distribution of exports and calculation of imports should be revised in accordance to national accounts.

Further, the conversion from industries to commodities and reconciling foreign trade calculations in rAps with regionalised foregn trade statistics are tasks which will request a close cooperation with expertise at Statistics Sweden, concerning the production of new data.

Construction of a priori matrices

This part follows the suggestions made in the report and includes a detailed analysis of CFS, and estimations of the gravity model based matrix. In this context a thorough estimation of RPC is part of the work.

Estimating base matrices

In addition to arranging the specific programming framework for the estimations to be performed, this part also includes the collection of complementary constraint data.

Exporting base matrices to the logistics model

Validation and calibration of models

According to the description in chapter 7.

9 Glossary

rAps	Swedish national regional analysis and prognosis system for population, employment and economy
SAMGODS	Swedish national freight model system
STAN	network software program for freight transport, currently used in SAMGODS
P:	production of commodities in value (V) or weight (Q) terms
C:	consumption of commodities, including retail and further processing of raw materials and intermediate products
W:	wholesale sector
P/C matrix	table of goods flows between production and consumption locations denoted V for value terms and Q for weight terms
PWC matrix	table of goods flows between production locations, wholesalers and consumption locations in value and weight terms
OD matrix	origin-destination table of goods flows in weight terms (Q) from production/consolidation/distribution locations to consolidation/distribution/consumption locations
NUTS	(Nomenclature des Unités Territoriales Statistiques) divisions of geographical areas in 5 levels used by EU statistical data
	NUTS I — Sweden
	NUTS II — Riksområden
	NUTS III — Counties
	NUTS V — Municipalities
I-O	Input-output table for commodities
IVP	Production by commodity and industry sector (Industrins varuproduktion)
FTS	Foreign trade statistics
CFS	Commodity flow survey
UVAV	Freight transportation by Swedish lorries >=3,5 tonnes

10 References

Ambite, J.L et al (2002), “Integrating heterogeneous data sources for better freight flow analysis and planning”, http://www.isi.edu/~ambite/SGER-2002-4-30.pdf

Bröcker, J. (1998), “Spatial effects of Transeuropean Networks. Preliminary results from a spatial computable general equilibrium analysis, Tecnische Universität Dresden, mimeo.

Bröcker, J. (2002), “Passenger Flows in CGE Models for Transport Project Evalutation”, Paper presented at the ERSA Congress 2002, Dortmund.

Canning, P. and Wang, C. (2004), “A Flexible Modelling Framework to Estimate Interregional Trade Patterns and Input-Output Accounts”, World Bank Policy Research Working Paper 3359

Celik, H.M. and Guldmann, J-M, “Spatial interaction modeling of interregional commodity flows”, Paper presented at the ERSA Congress 2002, Dortmund.

Edwards, H. (2000), “Traffic Assignment Analysis and O/D-matrix Calibration”, KFB & VTI research 32

Eriksson, J. (1981), “Algorithms for Entropy and Mathematical Programming", Dissertation no 63, Linköping University, Sweden,.

Eriksson, J.(1988), "The simultaneous two-step entropy model”, Journal of Computational and applied Mathematics 22.

Eriksson, J. (2000), “VTI/TPR-modellen – Ett prognossystem för godstransporter”, KFB & VTI forskning/research 31

Eriksson, J., Björketun, U., Edwards, H. and Karlsson, R.(2000), ”Dokumentation av NÄTRA-modellen – En modellering av näringslivets transporter med bil i Stockholms län”, VTI notat 36-2000.

Flegg, A. T., C. D. Webber and M. V. Elliott (1995) On the Appropriate Use of Location Quotients in Generating Regional input-Output Tables. Regional Studies Vol 29 pp. 547-561.

Glaeser, E.L. and Kohlhase, J.E. (2004), “Cities, regions and the decline of transport costs”, Papers in Regional Science 83 (1). http://post.economics.harvard.edu/hier/2003papers/HIER2014.pdf

Hovland, Svein (2004), ”Soft Constraints in Explicit Model, Predictive Control” Department of Engineering Cybernetics, The Norwegian University of Science and Technology, Trondheim, Norway

Jackson, R.W, W.R. Schwarm, Y. Okuyama and S. Islam (2004), “A Method for Constructing Commodity by Industry Flow Matrices”, Research Paper 2004-5, Regional Research Insitute, West Virginia University.

Kapur, J.N. and Kevsavan, H.K.(1992), "Entropy Optimization Principles with Applications", Academic Press, Inc.

Kullback, S.(1959), "Information Theory and Statistics", John Wiley and Sons.

Lahr, M. L. and de Mesnard, L. (2004). “Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis”. Economic Systems Research.

Lien T. and K. Ø. Sørensen (1996) Regional Accounts in Norway. In Regionalregnskap i Norden, edited by Mønnesland, J. and K. Rönnbäck-Zetterström, Working Papers 1996:9. NordREFO Stockholm.

Mataxatos, P. (2003) “Estimation and Accuracy of Origin-Destination Highway Freight Weight and Value Flows”, Journal of Transportation and Statistics V6, N2/3, pp 67-80.

McMann, P., J. H. Dewhurst (1998) Regional Size, Industrial Location and Input-Output Expenditure Coefficients. Regional Studies Vol. 32 pp. 435-444.

ME&P (2002), “DfT Integrated Transport and Economic Appraisal – Review of Freight Modeling – Final Report”.

MVA (1999), “Report on the network monitoring and the network model architecture, UTMC 04 Deliverable 4 Part 1.

Round, J. I. (1978) An interregional Input-Output Approach to the Evaluation of Nonsurvey Methods. Journal of Regional Science Vol 18 pp. 179 - 194

Snickars, F. & Weibull, W. (1977), " A minimum information principle - theory and practice", Regional Science and Urban Economics 7.

Stevens, B. H., G. I. Treyz, D. J. Ehrlich and J. R. Bower (1983), “A New Technique for the Construction of Non-Survey Regional Input-Output Models”, International Regional Science Review Vol 8 pp. 271-286

Stevens, B. H., G. I. Treyz and M. L. Lahr (1989), “On the Comparative Accuracy of RPC Estimating Techniques”, In Frontiers of Input-Output analysis, edited by R. E. Miller, K. R. Polenske and A. Z. Rose. New York and Oxford: Oxford University Press.

Stone, R. (1961), “Input-Output and National Accounts”. Organization for European Economic Cooperation. Paris.

Tengblad, Å (1999) “Swedish regional accounts - some principles”, Unpublished note from Statistics Sweden (in Swedish)

TÖI (2004), “Construction of basis OD matrices for 1999 freight flows between municpalities in Norway, TÖI report 699/2004.

Annex 1 - Data sources

National accounts

The National Accounts (NA) is the primary data source for constructing the constraints, since most economic variables are derived from, or made consistent with NA. More specifically we make use of the supply, use and input-output tables from NA. The supply and use tables produced by Statistics Sweden are an integrated part of NA, as a central tool for calculation of GDP from the production and use side, respectively.

The *supply table* gives information about product x industry, that is which products *k* being *supplied* by which industries *h*. The total supply of product *k* is given by the sum of production of *k* in all industries and imports of *k*. This total supply is valued at *Basic price*, which means the price that the producer gets[20].

The *use table* also gives information about product x industry, but now meaning which products *k* being *used* by which industries *h* - used in intermediate consumption by industry *h* and used in final consumption. This total use is valued at *Purchasers price*, which means the price the consumer pays. The difference between basic price and purchasers price is the *trade margin*, which is the value of wholesale and retail trade production[21].

In addition to the supply and use tables Statistics Sweden also produces a symmetric *input-output table*, derived from the supply and use tables. This input-output table gives information about product x product, meaning the supply of *k* being used - as intermediate consumption in production of commodity *l* and in final consumption of product *k*. The input-output table is calculated for total supply, for domestic production and for imports.

In this input-output table the valuation is at basic prices. This means that all trade margins have been moved from the products where they are included in the purchasers price to the product which refers to trade margins, namely wholesale and retail trade products.

Regional Accounts

Statistics Sweden (SCB) has been working with Regional Accounts (RA) since 1992. The first time figures for Gross Domestic Product per Region (GDPR) were published was in 1996, concerning 1985-1994. Since then figures for GDPR have been constructed yearly.

[20] In Swedish: Producentpris.

[21] It could be noted, however, that imports are valued cif (freight included). This means that for imported commodities a part of the trade margin, namely the part generated by trade from the foreign country to the Swedish border (= the importing enterprise), is already included in the basic price.

NA for the period 1993 – 2000 has recently (2003) been revised and new statistical information has been included in the calculations. At the same time a corresponding revision of RA was made.

RA is basically a regional allocation of corresponding accounts for the whole Swedish economy, using the same concepts and definitions as in the NA. However, Statistics Sweden declares that there are big conceptual and practical difficulties when it comes to putting together complete accounts at the regional level.

Because of the stated difficulties the calculations are confined to, in summary,

- Production per industry
- Gross fixed capital formation per industry
- Disposable household income

Regional Accounts and the rAps database

The system for regional analysis and forecast, rAps, was launched in year 2000. The database in rAps is updated every year, and with respect to regional economic data RA is one source of information. However, RA is too limited to fulfil the data requirements in rAps, since calculations in rAps are performed in a supply and use framework. Therefore, one complementary body of data consists of input output calculations.

The regional input-output tables in rAps are corresponding to the national input-output table with respect to valuation, i.e. also the regional tables are at basic prices. However, the symmetric regional input-output tables are calculated as industry x industry (49 x 49). The regional tables are (at present) based on national input output tables for 1995 (which is not consistent with the most recently published National Accounts). New input output tables for year 2000 have been published in May 2004.

At the time when rAps was launched, regional economic data was based on specific regional input output tables for each county, reasonably consistent with the corresponding national input output tables for 1995[22]. The current version 1.6 of the rAps database refers to regional economic data for year 2001. With respect to the updated databases, the national input output tables for 1995 are still used to calculate intermediate consumption – as the new input output tables for year 2000 have not yet been incorporated.

With respect to other parts, totals from NA are distributed to regions in the following ways (mostly proportional).

---

[22] The multiregional model application that has been used for long-term regional economic forecasts (in year 2004) has also been based on this database, using 1995 as the base year in the calculations.

- Private Consumption, by use of Disposable household income
- Public consumption, by use of employment in public sectors
- Gross fixed capital formation, by use of gross output per industry
- Exports, by use of gross output per industry

It should be observed that Imports are excluded in this distribution of national totals to regions. Instead, Imports are calculated within the rAps model system. These calculations make use of fixed parameters for the imports share of intermediate consumption, private consumption and gross fixed capital formation. The estimations of these parameters are at present based on data from 1995.

Other data sources

In addition to NA, RA and rAps there are other data sources (often basic data to NA) at Statistics Sweden, which are central in this work:

Business Statistics[23] is providing basic economic information for enterprises, e.g. production, which is, input data to NA. In Business Statistics all registered enterprises are included, excluding financial enterprises. All enterprises with more tha 50 employees are surveyed by a questionnaire whereas remaining enterprises are surveyed by data from tax authorities.

Statistics on the production of commodities and industrial services, *IVP*[24], gives information on manufacturing production at a detailed commodity level (KN8), with association to enterprises in Business Statistics.

Foreign Trade Statistics, *FTS*, gives information on exports and imports at the same commodity level as IVP, with association to enterprises in Business Statistics.

[23] In Swedish: Företagsstatistiken
[24] In Swedish: Industrins varuproduktion

The following table presents a preliminary overview of different data sources to be used, defined by variables, geographical level, type and frequency.

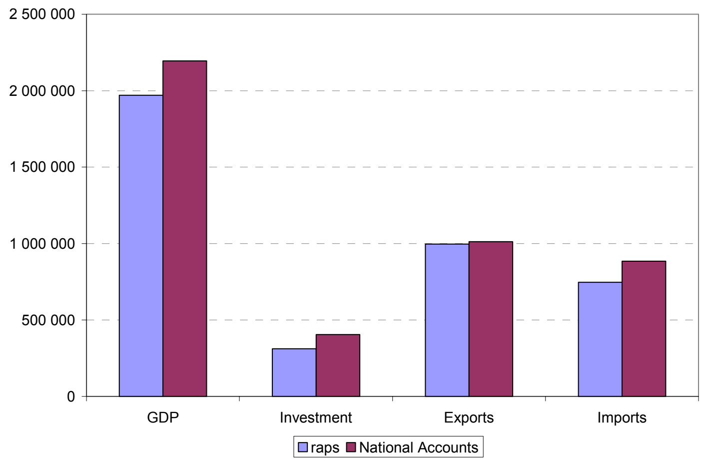
Data source	Variables	Geographical level, Sweden	Geographical level, abroad	Type	Frequency
rAps	Production etc. by industry (49), in value terms	NUTS V		Register data	Yearly
National input-output table	57 commodity groups in value terms	National			Intermittent
Foreign trade statistics	Commodity groups KN8 Value and weight	National	Country	Register data Cut-off	Yearly
IVP **(Corresponding data for primary sectors)**	Production by industry sector KN (10500 commodities) Value and weight	NUTS V		Register data Cut-off	Yearly
CFS	Commodity flows STAN groups 12 between industry sectors	NUTS V Reliable data NUTS II		Sample survey	2001, forthcoming 2004 and 2007
Freight transport Swedish lorries>= 3.5 tonnes	Commodity groups NST/R In weight terms	NUTS III (counties)	Country	Sample survey quarterly	Yearly since 2000
Traffic counts, TMS	Annual average day traffic	80 fixed points (TF) variable points (VS)		Sample of points at the national road system	Yearly

Annex 2 - rAps, CFS, National and Regional Accounts – some comparisons[25]

Economic data in the present version of rAps is partly based on national input output tables from 1995, which are not consistent with revised National Accounts. Further, current procedures for regionalisation of base year data for gross fixed capital formation (investment), exports and imports seem to be open to discussion. To get some idea about magnitudes, the following figures make some comparisons between data in rAps and data according to other sources. In Figure A2, data in rAps is compared to National Accounts (NA) with respect to GDP, Investment, Exports and Imports.

There is a difference between rAps and NA in GDP, amounting to more than 225 500 MSEK. This difference is, however, completely explained by that part of GDP in NA which is not allocated by industry (activity). (This means a discrepancy between GDP by expenditure and by industry).

Figure A2 Aggregate data year 2000 according to National Accounts and rAps, MSEK.



As for the difference in the main source seems to be an underestimation of imported investment goods, explained below. Otherwise, total Investment in rAps should be equal to that in NA, given the procedure for distributing figures from NA.

[25] When this paragraph was written in June 2004 the new version 1.6 of rAps, including economic data for year 2001, was not available. The comparisons therefore refer to rAps data for year 2000.

The difference in total Exports is only very marginal, which is also according to expectations, given that official figures for exports have been transformed to the very industries producing the same commodities and services.

On the other hand the difference in Imports is substantial. A basic reason for this has already been pointed at, namely that Imports in rAps are determined from parameters (import shares), originally estimated on data from 1995.

Turning to the regional distribution of Exports and Imports, it could be of some interest to compare rAps with figures from CFS. Such a comparison seems reasonably well-founded looking at aggregate figures at NUTS II level, i.e. Sweden divided into eight macro-regions. However, the comparison cannot be made for exactly the same year, since rAps refers to year 2000, whereas CFS refers to year 2001.

It should also be pointed out that CFS concerns trade of commodities, whereas rAps concerns trade of both commodities and services. To take this distinction into some consideration, the national average share of commodity exports/imports of total exports/imports, 80 percent and 75 percent respectively, has been applied to the figures from rAps.

Figure A3 Exports per NUTS II region, rAps year 2000 and CFS year 2001, MSEK.

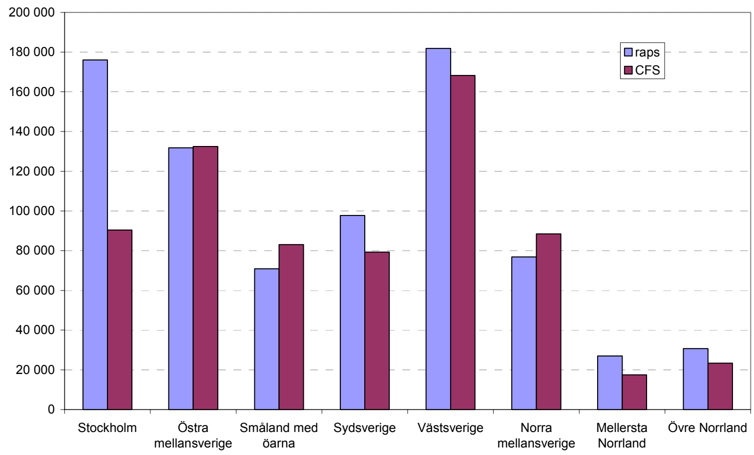
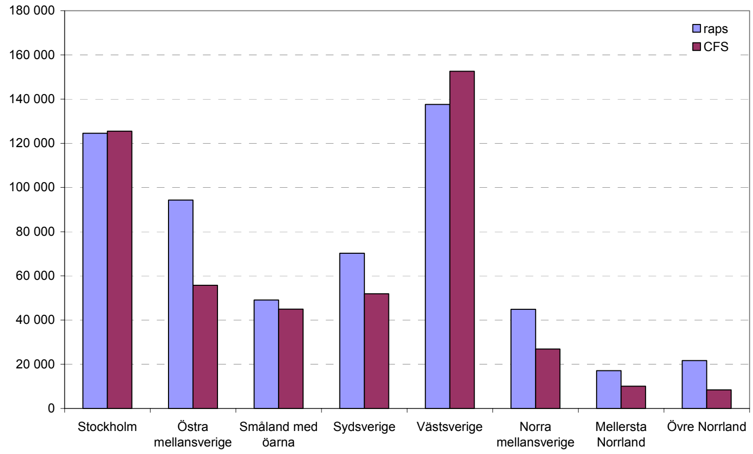


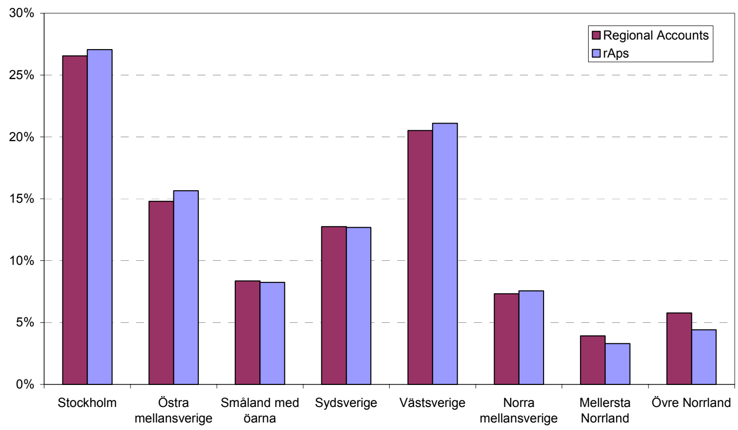
Figure A4 Imports per NUTS II region, rAps year 2000 and CFS year 2001, MSEK.



Taken as a whole, the regional distribution of Exports and Imports according to rAps and CFS is similar. One conspicuous exemption is Exports from Stockholm. This mainly reflects that Stockholm has a considerably lower share of commodity exports than the national average of 80 percent. According to rAps 40 percent of total Exports from Stockholm are generated in industries classified as services sectors. However, it is not possible to get detailed knowledge into this matter without further analysis where data for industries are converted into commodities.

Finally, rAps and Regional Accounts (RA) can be compared regarding the regional distribution of gross fixed capital formation. It was mentioned above that the national total in rAps is lower than in NA, probably because of imported investment goods being underestimated. However, according to figure 5 the regional distribution at the NUTS II level is very similar.

Figure 5 Gross fixed capital formation year 2000 per NUTS II according to Regional Accounts and rAps. Percent.



Annex 3 - RPC, theory and Norwegian experiences

Understanding the determinants of interregional commodity flows is critical not only for transportation infrastructure planning but also for regional development policies. Therefore, independent of the requirement for base matrices in freight modelling, the issue of estimating interregional trade flows has since long also occupied the regional economics society.

The perspective in regional economics has to large extent been concentrated on intraregional trade, i.e. the diagonal elements in the trade flows matrices, for several reasons. One reason is dearth of data; another reason is that the problem at hand often has made it sufficient to deal with interregional trade in terms of an aggregate pool.

The research in this field has been quite extensive during many years. On the assumption that methods derived from this field can serve as basis for estimating the diagonal elements in the P/C matrices a lot would be gained. We will therefore in the following give a short survey of the field.

Introduction

Regional accounts and regional trade can either be made bottom-up or top-down. A complete bottom-up approach is very cost demanding, and is normally avoided. Instead one can use a combination, consisting of a top-down approach on the accounts data, combined with bottom-up approach on trade data.

Supply and use tables (production and consumption) from national accounts are originally constructed without any geographical dimensions, except imports abroad (in the make table) and exports abroad (in the use table). One can obtain a description of production and consumption nationally by linking domestic production from the make table to domestic consumption in the use table.

When constructing regional accounts (from national accounts) it normally implies two types of adaptations: 1) scaling make and use tables to the chosen regional level, and 2) estimating/scaling exports and imports across the regional borders (both abroad and domestic). Instead of estimating the regional export/import or the aggregate interregional trade ratios, it is more efficient to estimate the intraregional trade. Regardless of the choice, one gets both intra- and (aggregated) interregional trade as a result.

So, the preferred way to develop regionalised input-output accounts (production-consumption matrices) is to add trade data from surveys to national accounts broken down to regions. Data should ideally be collected for distributions of both sales (production) and purchases, which makes it possible to use biproportional-fitting procedures.

However, non-survey methods must be considered as a necessary alternative since the data may be unavailable, or unreliable. The emphasis in the regional-economics literature has been on intraregional trade and the intraregional trade coefficient, expressed as the self-sufficiency ratio or the regional purchase coefficient (RPC).

Biproportional Fitting

Biproportional fitting is what the regional economics literature relies upon when data availability is limited to only totals (i.e. total production and total consumption of a good), but where some a priori representative distribution is available. Use of biproportional fitting in regional economics dates back to Stone (1961). Biproportional fitting methods, as formulated by Stone, are still one of the main methods for estimating (intra)regional flows in regional economics. Data demand is reasonably low – only sum production and sum consumption of each region is needed together with an a priori distribution – and the method is easy to apply. This does not imply that better methods are unavailable, or that the data demands are, in some instances too high.

Biproportional fitting has as its base the observation of the variables of interest at an aggregated basis. Total make (production) and total use (consumption) in different regions is observed. The distribution of the produced goods to different regions is then estimated through biproportional fitting on an a priori distribution. Its advantage is that it is very easy to apply and relates directly to the observed data. However, the method does not take other restrictions into account, or other information, than the information supplied by the data aggregates (although the method can be modified to take into account some restrictions).

Since these methods turn out to be a subset of entropy maximization procedures and entropy maximization methods are outlined elsewhere in the report, we will not discuss this further here[26].

The problem

There would have been very few statistical problems if the commodity flow survey had been a survey designed to sample commodity flows at a municipality × municipality level. However at this level the survey is a sample for different goods. Not all flows from a municipality to another municipality is registered, and just as important, not all flows of a good within a municipality is observed in the commodity flow survey sample.

---

[26] A state of the art for biproportional techniques is given in Lahr and Mesnard (2004), and a discussion of biproportional fitting versus mathematical programming methods is given in Canning and Wang (2004).

This problem is obvious at the municipality (NUTS V) level, but is also, indirectly, present at a higher geographical level (NUTS II) since all commodity flows within a county will not have been registered (i.e. let $Q_{R,R}$ be all flows from county R to county R, then these flows will not be estimated consistently if some of the $Q_{r,r}$ flows are missing from the sample, where the $Q_{r,r}$ are flows from sub-region r to sub-region r).

Ideally all flows, i.e. all flows of the form $Q_{r,s}$ where $Q_{r,s}$ is the flow of a category of good from region r to s, should be estimated. As noted in the earlier deliveries, regional economics has a set of tools for estimating $Q_{r,r}$ flows (i.e. the amount of goods consumed in region r that are also produced in region s). The estimates of the $Q_{r,r}$ flows are usually given in relative terms (i.e. relative to consumption of the good in the region) and denoted RPC (regional purchase coefficient).

Several attempts has been made to find better non-survey based measures of the regional purchase coefficient[27], in some instances data for validation of the measures has been lacking. The use of the regional purchase coefficient, and biproportional fitting, as proposed here is on the conservative side of this literature. A fairly basic extension of a common technique (estimation of RPC by location quotients) is combined with biproportional fitting. These tools can be regarded as fairly standard tools in the regional economics literature, with the one exception that the estimation of the RPC is more general than what is used in many instances. The tools are used here to generate a prior matrix that is more consistent than that obtained directly from the commodity flow survey. The steps are as follows:

I. Estimate the $Q_{r,r}$ flows by the use of techniques used in regional economics (i.e. use the ideas behind Regional Purchase Coefficients in the estimation procedure)
II. Use biproportional fitting to estimate the $Q_{r,s}$ ($r \neq s$) flows

This matrix would then give a more even pattern of flows of goods than what the sampled commodity flow survey data would give. The information used in the construction of this matrix, in addition to the data themselves, is an underlying assumption about the functional form of the Regional Purchase Coefficient.

Estimation: Functional Form

The proposed RPC estimation is based on the following functional form:

(1) $Q_{r,r} = \alpha \mathrm{Min}\{P_r, C_r\} + (1-\alpha)(P_r/P_{tot})C_r$.

---

[27] For one attempt, see Stevens et al (1983).

$Q_{r,r}$ is the amount of good k produced in region r that is also consumed in region r. P_r, is the total production of good k in region r net of export of the good from the region from abroad, C_r the total consumption of good k in region r net of import to the region from abroad. P_{tot} is the national production of the good net of total exports. As can seen from the expression above:

a. A region not producing the good ($P_r = 0$) is covering all consumption by imports from other regions ($Q_{r,r} = 0$)

b. If the region in question is the country as a whole ($P_r = P_{tot}$), then all consumption (net of imports) are covered by national production (net of exports) ($Q_{r,r} = P_{tot}$) (see g.-j. below)

c. The function is homogenous of degree 1, measurement units do not play any role so the expression is just as valid in value terms as in (different) weight terms.

d. The extreme case $\alpha = 0$ corresponds to the case where the region consumes own production of the good k in proportion to the region's share of national production (P_r/P_{tot}), where national production is net of import and export (implicitly this is a zero transport cost assumption and/or goods produced in different regions are entirely non-homogenous and not substitutes for each other)

e. The extreme case $\alpha = 1$ corresponds to the case where the region covers, as far as possible ($Min\{P_r, C_r\}$), consumption from own production of good k (implicitly, an infinite transport cost assumption and/or a goods produced in different regions are completely homogenous and substitutes for each other so that the closest producer is chosen)

f. Assume that α and $Q_{r,r}$ is given for a region of a particular size (with P_r, C_r as measures of size), and that the same α is used to estimate $Q_{s,s}$ for a smaller region. This smaller region has production $P_s = \lambda P_r$ and consumption $C_s = \lambda C_r$. $Q_{s,s}$ for the smaller region is then given by $Q_{s,s} = \alpha Min\{\lambda P_r, \lambda C_r\} + (1-\alpha)(\lambda P_r/P_{tot})\lambda C_r = \alpha\lambda Min\{P_r, C_r\} + (1-\alpha)\lambda^2(P_r/P_R)C_r = \lambda Q_{r,r} - \lambda(1-\lambda)(1-\alpha)(P_r/P_{tot})C_r$. In relative terms, let $RPC_s = Q_{s,s}/C_s$ and $RPC_r = Q_{r,r}/C_r$ then $RPC_s = RPC_r - (1-\lambda)(1-\alpha)(P_r/P_{tot})$. Any transport costs less than infinity (or non-substitutability between goods produced in different regions, i.e. $\alpha < 1$) make a smaller region cover, proportionally less of its consumption from own production.

The proposed function has a set of desirable properties. Consistency: (a), (b), homogeneity of degree one (thus the irrelevance of measurement units): (c), interpretability: (d), (e), and certain scale properties: (f). Furthermore, it extends a very commonly used assumption about RPC estimation that is based on estimating RPC through the use of location quotients (this corresponds to $\alpha = 0$ in (1)).

Estimation: Data and methods

The data for $Q_{r,r}$ can be obtained from the commodity flow survey for the cells (i.e. r,r) where this information has been sampled. The formula (1) can be used in several different ways:

g. P_r is total production in the region; while C_r is net of net exports (i.e. C_r is given by local consumption + exports from the region – imports to the region). P_{tot} is total production in the country. The consistency requirement (b) is then fulfilled

h. P_r includes import (i.e. P_r is given by local production + imports); while C_r includes exports (i.e. C_r is given by local consumption + exports from the region). P_R is total production in the country plus imports. The consistency requirement (b) is then fulfilled

i. P_r is net of net exports (i.e. P_r is given by local production + imports – exports); while C_r is given by local consumption. P_{tot} total production in the country net of export. The consistency requirement (b) is then fulfilled

The derivation of the functional form (1) was based on the fulfilment of the consistency requirement (b). Each of the different choices of data sources would give rise to different interpretations of α. However, the functional form (1) can also be regarded as a descriptive tool. In that case it is possible to use (1) as follows:

j. P_r is total production in the region (and thus includes export); while C_r is local consumption (do not include import). P_R is total production in the country.

The functional form of $Q_{r,r} = \alpha Min\{P_r, C_r\} + (1-\alpha)(P_r/P_R)C_r$ has only one parameter α to be estimated and this invites to perform an informal Hausman test of the validity of the model through the use of two different estimators: (1) the OLS (ordinary leas squares) estimator, and (2) an "average value of α" estimator. The average value estimator is computed from:

$$(2)\quad \frac{1}{n}\sum_{r=1}^{n} Q_{r,r} - \frac{1}{n}\sum_{r=1}^{n}\frac{P_r}{P_{tot}}C_r = \alpha\left[\frac{1}{n}\sum_{r=1}^{n} Min\{P_r, C_r\} - \frac{1}{n}\sum_{r=1}^{n}\frac{P_r}{P_{tot}}C_r\right]$$

Large discrepancies between this estimate of α and the OLS estimate indicate that the functional form is unreliable at the chosen level of aggregation.

Note that the proposed functional form has certain scaling properties that are intuitively plausible in that RPC estimated at a lower geographical level (as long as $\alpha < 1$) make a smaller region cover, proportionally less of its consumption from own production (cf. (f)). This makes it reasonable to use values for α estimated at county level (NUTS III) at a smaller geographical level (such as NUTS V).

However, the commodity flow survey gives data for $Q_{r,r}$ at the municipality level (NUTS V) albeit in a sampled form. This makes it possible to estimate α at the appropriate geographical level. As noted from (1) $Q_{r,r}$ is the dependent variable in the estimation, so sampling errors in $Q_{r,r}$ is allowed by the estimation of α. The sampling plan of the commodity flow survey should be examined in order to ensure that $Q_{r,r}$ is unbiased, but, as can bee seen from (1), correction for any bias is simple through the use of weighted averages. In short, the data availability makes it possible to estimate the α coefficient at the municipality level and this would then be preferable to estimating the α coefficient at county/region level. However, one would naturally then expect larger errors.

Estimation: Experiments with Norwegian Data

Attempts to use (1) in the estimation of regional purchase coefficients (i.e. RPC values and thus indirectly $Q_{r,r}$ flows) have been done at county level with Norwegian data on a data set that differentiated between 45 different classes in goods.

In general: Some instabilities (divergence between the OLS and the "average α" value estimator) were observed, in other words the functional form is not the be-all or end-all of RPC estimation (which is neither to be expected). Observed $Q_{r,r}$ flows were, in general, within ±20% of the estimated values given by (1), (more precisely, on average within -22% and +24% for all counties when the "average α estimator" was used, which was also found to be the most robust estimator).

However, there were some exceptions to these results. These occurred in industries which are heavily concentrated and/or dependent upon natural resources, i.e. fish processing, tobacco (where major production is localized in one county in Norway), oil/oil drilling and oil pipes, and in non-ferrous metals. In these cases estimates could deviate grossly from observed RPC (up to 65% between estimated $Q_{r,r}$ and observed $Q_{r,r}$). This is not unexpected since, for example the category oil, oil drilling and oil pipes, could be said to include a collection of very different goods and would thus not really be amenable to the proposed method. A similar argument could be said to be valid for non-ferrous metals (really a set of different products with a very concentrated localisation) and tobacco (on dominant producer and a few specialists).

In short the proposed RPC method seemed to work reasonably well, except in cases were the product category mixed a set of different products where each has a geographically concentrated production. In contrast, deviance between observed $Q_{r,r}$ flows and estimated $Q_{r,r}$ flows was low in the case of production of machinery (estimated and observed $Q_{r,r}$ flows differed by from -1,5% to 9%). Note that the data used was not without its own flaws though, and that commodity flows at one time period was fitted to production (and consumption) at another time period in order to obtain the results. Furthermore, export and import data were not handled correctly (i.e. it was not ensured that aggregation would give total production equal to total consumption). Nevertheless, the method seemed to work reasonably well for most industries.

Estimation: Consequences for Swedish data

The experiences with the Norwegian data indicate that:

I. The proposed method for estimating RPC will give reasonable results, except possibly when

II. A category of goods includes several goods that really are highly differentiated goods whose production is geographically very concentrated

The Norwegian RPC estimates for oil/oil drilling and oil pipes may have a parallel in (part) of the mechanical industry of Sweden. Some mechanical industries may be large and produce specialised goods (say airplanes) in a very concentrated location, while other producers are producing goods with a very local demand. Such situations may arise in the data set, care and judgment has to be used in order to get good RPC (or equivalently $Q_{r,r}$ flow data) in this case. It is a general aggregation problem that would arise whatever methods one used. Unstable estimates may be due to too aggregated categories of goods and as in all estimations common sense has to be used.

Annex 4 –Definitions of wholesale

Wholesale in the base matrices – the P-W-C matrix

The essence of a pure P/C matrix for a specific commodity could be defined as follows: The P/C matrix shows the flows (in tonnes) from the production location to the location where the same commodity is consumed. From this definition two features should be stressed: The matrix is only concerning *commodities*, and the commodities produced are the same as the commodities consumed, i.e. the commodities are *unchanged*. The meaning of a commodity being unchanged is here to be understood literally, i.e., no physical transformation of the commodity is taking place.

Is the product mix a problem?

From the point of view of constructing base matrices an ideal situation would mean that activities performed in different industries are exclusively such activities that are typical for the respective industry. This means that manufacturing is performed only in enterprises classified as manufacturing industries, and these enterprises are not performing other activities. It also means that enterprises within the wholesale industry perform all wholesale activities, and this industry is not performing other activities. In this ideal situation the wholesale industry should thus be excluded from the P/C matrices.

The present situation is, however, not ideal since there is a product mix within industries. This mix is of special concern with respect to manufacturing (commodity production) within the wholesale industry and wholesale activities within manufacturing industries. Let us for the moment represent the wholesale industry by an aggregate including also retail trade and some other activities[28]. According to data from the Supply table in National Accounts[29] for year 2001 the product mix in the wholesale industry can be described as follows:

- For this aggregate sector the Gross Output is 324 667 Million SEK, of which 292 612 MSEK, or 90 percent, refer to output of "wholesale products".

[28] This means that the industry code for Wholesale, SNI 92 = 51, is represented by SNI 92 = 50-52

[29] http://www.scb.se/statistik/NR/NR0102/2002A01/Supply and Use Tables 1995-2001.xls. See also the discussion about valuation issues (The value of P, C and W) below.

- The output of “non wholesale products”, 32 055 MSEK, is to two-thirds referring to various kind of services. Only 1 273 MSEK refer to commodity production, yet not manufacturing but Products of Agriculture.

Assuming that the aggregate sector is reasonably representative for the genuine wholesale industry, the figures according to National Accounts demonstrate a rather small product mix, especially with respect to commodity products.

As to wholesale activities within manufacturing industries the product mix can be described as follows:

- The total output of “wholesale products” amounts to 364 429 MSEK, of which 292 612 MSEK, or 80 percent is produced within the (aggregate) wholesale industry.
- The output of “wholesale products” outside the wholesale industry, 71 817 MSEK, is mainly taking place within manufacturing industries.

Judging from these figures the product mix problem, with respect to wholesale, essentially means that a significant and probably increasing part of wholesale activities are performed within manufacturing enterprises[30]. The question is whether this product mix problem, as such, obstructs the construction of pure P/C-matrices?

First, as to the product mix within the wholesale industry: Excluding all enterprises within the wholesale industry from the construction of P/C-matrices would mean that only a very small fraction of manufacturing is (erroneously) excluded. But even such a small error will not occur since the construction of P/C-matrices is not based on industry data but on commodity data.

These commodity data will be taken from IVP (statistics of commodities and industrial services), as described in the previous paragraph on conversion from industries to commodities. This survey is directed to all local units[31] (with more than 20 employees) in manufacturing *and* service enterprises. Thus, if manufacturing is taking place within an enterprise classified as wholesale, it will most probably be represented in IVP.

Next, turning to the product mix within manufacturing industries: The IVP survey gives data on commodity production *and* other activities, e.g. freights, wholesale and business services. These other activities will be excluded when making the conversion from output at the industry level to production of specific commodities.

---

[30] Comparing the Supply table for 2001 with the table for 1995 shows that the share of wholesale activities performed outside the wholesale industry has increased from 14 percent to 20 percent. The share of “non wholesale activities” performed by the wholesale sector is, on the other hand, the same in 1995 and 2001.

[31] Local unit is a term used for a unit within enterprise, “arbetsställe” in Swedish.

The conclusion is therefore that the product mix problem does not seem to obstruct the construction of pure P/C matrices.

However, some additional comments should be made concerning the definition of activities. In the discussion above an aggregate sector is used to represent the wholesale industry. The reason to use this aggregate sector, SNI 92 = 50-52, is that this definition is used in National Accounts when making input-output tables and the supply and use tables.

But there is of course more detailed information available from Statistics Sweden's Business Register[32]. According to this data the wholesale and commission trade, SNI 92=51, represents around 55 percent of the Gross Output for the aggregate, SNI 92 = 50-52.

'Genuine' wholesale trade, SNI 92 = 512-519, is defined as sales of commodities to producers and retailers; it also includes activities that usually are belonging to wholesale, as assembling and packing etc. The definition is clear but there are some imprecisions with respect to the distinction with retail trade. For example, builders merchants are classified as retail trade even if most part of the sales refer to producers[33].

However, even if there are ambiguities of this kind, the distinction with manufacturing is obvious: If there are manufacturing activities within a wholesale enterprise, these will be found in the data from IVP. Thus, as far as the construction of pure P/C matrices is concerned, ambiguities in the definition of wholesale do not seem to be a big problem.

Finally, a minor comment is concerning possible 'measurement errors', as to the classification of activities within enterprises. At present there is a significant difference in electricity taxes between manufacturing enterprises and other enterprises, and a total tax exemption for manufacturing enterprises has been ruling for 10 years. One could therefore assume that some enterprises in non-manufacturing, e.g. wholesale enterprises, would be inclined to untruthfully report some of their activities as manufacturing. According to officials at Statistics Sweden enterprises now and then make inquiries to get classified as a manufacturing enterprise, but it is not very common. Even if various measurement errors certainly exist it is difficult to see how they could be circumvented in this context.

[32] In Swedish: SCB´s Företagsregister

[33] The fact that the sales takes place in a shop is what makes the activity defined as retail trade. It could also be mentioned that trade of motor vehicles is part of SNI=50.