
**Global Trade Analysis Project**
https://www.gtap.agecon.purdue.edu/

This paper is from the
GTAP Annual Conference on Global Economic Analysis
https://www.gtap.agecon.purdue.edu/events/conferences/default.asp

# Disaggregating the dairy sector in the GTAP database[1]

Marian Mraz and Alan Matthews
Department of Economics, Trinity College Dublin

Draft version 14 April 2007

## *1. Introduction*

Social Accounting Matrices (SAM's) are a very useful way of organizing sectoral economic data in a consistent economy-wide framework. They link the production cost structures of the various commodities with final demand as well as capture the creation and distribution of income. They are increasingly used for policy support analysis, either directly or as an input to computable general equilibrium models. Many Statistical Offices now regularly produce balanced input-output tables or social accounting matrices on an annual or bi-annual basis. The available matrices, however, often differ among regions in terms of their commodity coverage, their precision in capturing the creation and distribution of income and their consistency with complementary data sources such as the economic accounts for agriculture, international trade statistics, labour statistics, household budget surveys etc. Often the sectoral breakdown is insufficiently detailed to allow their use for industry-specific analysis. In such situations estimating the missing data is often the only way to obtain the required SAM. `

The GTAP database and associated modeling suite have been widely applied for multi-regional and multi-sectoral analysis of agricultural and trade policies (Hertel, 1997).[2] The GTAP database is a very useful source of economic data in particular due to its substantial coverage and world-wide consistency. The current GTAP database version 6 distinguishes 57 sectors across 87 of the world's countries and regions. Technically, the GTAP database is a composition of all countries SAM's, mutually interlinked by bilateral trade flow matrixes. The SAM's typically contain the input-output (IO) technological (cost) data in the desired sectoral split, the

[1] Financial support from the Food Industry Research Measure of the Irish Department of Agriculture and Food is gratefully acknowledged. This project is also supported by TradeAg. TradeAg is a Specific Targeted Research Project financed by the European Commission within its VIth Research Framework.
[2] GTAP is an abbreviation for Global Trade Analysis Project, hosted by the Purdue University. Apart of the modeling tools they provide the world's most complex and publicly available database suitable for CGE analysis. GTAP database Version 6 has been used in this study.

composition of the sectoral value added, tax revenues, final demand and foreign trade flows. Trade matrices are built in a consistent fashion implying that the world's total trade sums to zero.

Nonetheless, there are policy issues where a more disaggregated and detailed database would be desirable. Elbehri et al. (2001), in evaluating the impact of trade liberalisation in the global oilseeds sector, suggest a number of critical reasons why the GTAP aggregate 'oilseed products' should be separated into meals and vegetable oils in this instance, including differences in the determinants of demand, different trade policies, and different trade status of countries as importers or exporters. Sue Wing (2006) attempted to extend the SAM by a bottom up break down of the electricity sector explicitly taking account of various technologies of electricity production. Horridge (2005) provides a GEMPACK utility called SPLITCOM for disaggregating the GTAP database to any desired level of regional and commodity aggregation. Our paper is in spirit closer to the work of Sue Wing, although we focus on the dairy sector rather then electricity production. The work can also be seen as complementary to Horridge's efforts in providing the necessary estimates for the splitting shares, which are a major input into the SPLITCOM utility. The SPLITCOM utility is a useful and flexible resource to derive a consistent disaggregation of the accounts in the GTAP database. However, it requires the user to provide the desired splitting shares, estimated from exogenous prior information. Putting these splitting shares together is not a trivial problem. A methodology is required which efficiently uses existing information yet remains flexible enough to incorporate any additional information which may become available (Robinson, 1998).

In this paper, we describe the process of disaggregating the dairy sector composite account in the most recent GTAP database. Our purpose is to develop a modelling tool to allow us to simulate future dairy policy reforms, particularly in the EU. Our major aim is to disaggregate the GTAP account for dairy products to capture explicitly the major dairy commodities. In terms of the coverage of the dairy industry, GTAP captures the dairy sector by means of two sectors i) raw milk (*RMK*) and ii) dairy products (*MIL*). However, when conducting in-depth analysis of agricultural policies it is clearly desirable to start with the highest possible disaggregated description of the benchmark sectoral and product inter-linkages. This is particularly relevant where different products included within a single sector are subject to different policies, which require a specific treatment in the model. In the case of the dairy industry in the EU and elsewhere, specific dairy products differ in terms of their market pricing, border protection, administrative policies as well as in their production technologies.

Disaggregating a sectoral account within GTAP requires additional information on the products' sub-sectoral row and column totals, estimates of the different production technologies

(cost shares) and value added components within each product sub-sector, as well as information on the different utilization of each product between intermediate use and various components of final demand (see Appendix 1 for a graphical depiction). This information is required for each region in the database. In our case, we work with a 14 region aggregation of the GTAP database although, in principle, our methodology would allow the disaggregation of the database for all 87 countries and regions distinguished in GTAP version 6 (see Appendix 2).[3] The commodity aggregation distinguishes 20 commodity sectors of which six are dairy products (Appendix 3). We provide a disaggregation of the composite dairy account into six product sub-sectors : butter; cheese; skim milk powder; whole milk powder; whey powder and fresh milk products. Information on the utilization accounts for these products (in volume terms) is drawn from the FAO Supply and Utilization Accounts supplemented, for the EU countries, by EUROSTAT data. Information on production technologies, cost and value added shares for the dairy industry is much less widely available, as this is generally considered proprietary information by the industry. One source which does provide this information in a usable form is the detailed US input output table from 1992. We have used these US cost shares as the starting point for estimating the detailed sub-accounts for our 14 regions. It is clearly a strong assumption that cost shares world-wide follow these 1992 US cost shares. It can be argued that production technologies in use will be influenced by the relative prices of the different cost elements in production. An increase in the relative price of capital, for instance, would initiate a move of the production point along the sectoral isocost curve and result in some substitution of capital by labor.[4] Therefore one could expect that cheese production, for example, might be more capital intensive in the US than in developing countries. However, in defence of the assumption, it can be pointed out that what is important in the US data are the relative shares of different cost elements. For example, is cheese production more or less capital-intensive than skim milk powder production? Is more milk required to produce a unit (in value terms) of fresh milk products compared to butter? It is not unreasonable to suggest that the answers to these questions provided by the US data are more generally applicable. Of course, if better data on individual countries were available, the methodology developed in this paper would allow the estimation of improved SAM's based on these data.

The paper is organized as follows. In the second section we describe the available prior information from the US IO tables. The procedure to compile a square symmetric IO table from

[3] We cannot guarantee that the constraints imposed within the sectoral breakdown would not turn out to be too restrictive and over-constrain the whole problem and therefore fail to find a solution.

[4] The magnitude of the substitution, given by the curvature of the industry isocost curve, is typically measured by the elasticity of substitution.

the US table is outlined in the third section. In the fourth section we present the information on the desired sectoral shares collected from various sources such as FAO and EUROSTAT databases. The fifth section introduces a number of formulations of the cross entropy estimation procedures applied within the study and presents the final estimates of the dairy sub-accounts. Section 6 evaluates the strengths and weaknesses of the approach adopted. The appendices include all analytical derivations and GAMS implementation codes of the mathematical programs.

## *2. The US input output tables*

This section describes the available US IO data from 1992 and outlines the necessary adjustments undertaken for the construction of the US SAM. The US SAM has been constructed to serve as prior information on the cost structure of the various dairy production sub-sectors. The US Bureau of Economic Analysis provides the US make and use tables for selected years. They are particularly rich in terms of data availability as the sectoral breakdown involves 492 industries including five dairy composite commodities (for the methodology see *US Department of Commerce 1998*).[5] The most recent US IO tables with the required sectoral breakdown were available for the year of 1992 (see Appendix 4 for the detailed product coverage of each sector).

The dairy sectors included in the US IO tables are[6] :

(i) "Fluid milk manufacturing" (NAICS 311511),

(ii) "Creamery butter manufacturing" (NAICS 311512),

(iii) "Cheese manufacturing" (NAICS 311513),

(iv) "Dry, condensed, and evaporated dairy products" (NAICS 311514),

(v) "Ice cream and frozen dessert"[7] (NAICS 311520).

The commodity and sectoral classifications employed in the US IO tables do not exactly correspond with the GTAP database and therefore some initial adjustments and aggregations are

[5] The expression "composite commodity" is used throughout the paper to denote the aggregate of similar products, e.g,. a composite commodity cheese is used as a representation of the total cheese production.

[6] The correspondences between HS and NAICS can be found on the web site of the Bureau of Economic Analysis (http://bea.gov).

[7] In our database we have introduced six dairy sectors. The sector of ice cream and frozen deserts has been aggregated with fluid milk manufacturing. For details on aggregation and correspondences to other major classifications please see Appendix 4.

necessary. The account for cattle ranching and farming (NAICS 1121), which is a single account in the US IO table, has been split between cattle ranching and farming *(CTL)* and dairy cattle and milk production *(RMK)* to match the *GTAP* classification. Dry, condensed and evaporated dairy products are treated as a composite good in the US data. Our analysis requires a more detailed treatment of dry dairy products and therefore this account has been further disaggregated to explicitly capture three main dry dairy products, i.e., skim milk powder *(SMP)*, whole milk powder *(WMP)* and dry whey *(WHP)*. As no publicly available information on the cost shares describing the production technologies of the dry dairy products has been found they were all estimated by a maximum entropy procedure. The values of totals entering the relevant constraints in the maximum entropy procedure were calculated by using the value shares from Table 2.1. The value shares were obtained by a simple division of the corresponding figures on production values by the total value production of dry dairy products.

Table 2.1 Production of dry dairy products in USA, 1992

| | SMP | WMP | WHP | total |
|---|---|---|---|---|
| Production value FAO | 1115260 | 287970 | 511400 | 1914630 |
| Value share | 0.5825 | 0.1504 | 0.2671 | 1 |
| Calculated production value in GTAP | 5374.3203 | 1387.6971 | 2464.3827 | 9226.4 |

Source : http://faostat.fao.org and authors' calculations

The next section outlines the general idea of the maximum entropy approach and mathematical programs used for the estimation.

### *2.1. Maximum entropy principle*

The concept of entropy, which was initially introduced as a measure of the lack of order of a system, has found applications virtually across all disciplines. Its applications in economics draw mainly on the information entropy revealed by Shannon (1948). The entropy function itself is derived by the combinatorial approach from the multinomial distribution giving the probability of the realizations of the particular set of numbers.[8] Its monotone transformation followed by the

[8] For a better understanding of the entropy principle an illustrative example can be found in *Jaynes (1963)*. He attempts to estimate the probability distribution of the following dice roll, where the only information available is the average value over a large number of N rolls. The problem is clearly ill posed as the number of estimated parameters is larger than the amount of "information" available.

Stirling approximation[9] leads to the definition of the entropy function 2.1.1. The maximum entropy approach developed by Jaynes (1957) maximizes the entropy function 2.1.1 subject to the constraints implied by the system.

Denote $N_{ij}$ the values of the particular matrix elements, $N_j$ and $N_i$ are the column and row totals respectively, and $a_{ij}$ the probabilities defined as $a_{ij} = \frac{N_{ij}}{N_j}$. Our aim is to find:

2.1.1. $$\max_a H(a) = -\sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij} \ln a_{ij}$$

subject to :
$$\sum_{j=1}^{K} a_{ij} N_j^1 = N_i^1 \qquad \forall i$$

$$\sum_{i=1}^{K} a_{ij} = 1 \qquad \forall j$$

The analytical solution of this problem reads as follows.[10] Note that the exact solution has to be calculated numerically as no explicit closed form solution of this problem exists.

$$\hat{a}_{ij} = \frac{1}{\Omega(\lambda_1, \lambda_2, ..., \lambda_K)} \exp\left(-\lambda_i N_j^1\right)$$

$$\Omega(\lambda_1, \lambda_2, ..., \lambda_K) = \sum_{i=1}^{K} \exp\left(-\lambda_i N_j^1\right)$$

Note that $\lambda_i$ identifies the Lagrangian multiplier. Table 2.1 provides the values of the desired column and row totals $N_j$ and $N_i$ respectively. The fully operational model has been written in GAMS suite software (Brooke et al 1998) and solved by the CONOPT solver (Drud 1985, 1994, 1998). The estimated US break down of dairy powder accounts is provided in Appendix 8. By assumption at this stage we refrain from considering any additional constraints imposing any particular production patterns. The estimates are purely determined by the maximum entropy principle.

---

[9] For $x \to \infty$, it holds that $\ln(x!) \approx x * \ln(x) - x$ see (Jaynes 1988). The accuracy is within 1% of $\ln(x!)$ for x>90. A more precise solution can be obtained from the formula $\ln(x!) \approx x * \ln(x) - x + \frac{1}{2}\ln(2\pi x)$ with an accuracy to within 1 % of $\ln(x!)$ for x>4 (Feller 1957).

[10] The solution can be easily derived from the problem's first order conditions by a differentiation of the Langrangian function, see Appendix 5.

### *2.2. Compilation of the symmetric US IO table*

The input-output accounting framework is highly flexible and allows for easy adoption of the most suitable format to capture the underlying economy. Apart from the additivity constraints no other specific restrictions apply on the actual shape of the IO sub-matrices.[11] Common IO tables are usually provided in the form of make and use matrices, although some statistical offices aim also to publish symmetric tables of Leontief coefficients. The US IO tables are available in the form of the make, use, value added and final demand matrices. In contrast to the inter-sectoral flows, provided at a highly detailed level, the value added section of the US IO tables is rather aggregated. Only sectoral information on labor remuneration, other value added, tax revenues and imports is provided. Final demand is given in the standard format distinguishing between private and government demand, investments and exports. To secure the consistency of the US data with the GTAP classification a symmetric US IO table has to be compiled.[12] Table 3.1 illustrates a simple IO framework for the US IO tables as released in 1992.[13] More detailed explanation on the sub-matrices is provided in Appendix 6.

***Table 3.1.*** US IO table, 1992

| | Commodities | Industries | Final demand | Totals |
|---|---|---|---|---|
| Commodities | | **U** | **E** | **Q** |
| Industries | **V** | | | **G** |
| Value added + Taxes + Imports | | **X'** | | |
| Totals | **Q'** | **G'** | | |

Source : *USDA (1992)*

A number of methods have been developed to convert the use and make tables into a symmetric square matrix of the input-output coefficients (a survey is provided by ten Raa and Cantuche, 2003). The methods differ by the particular additional constraints imposed on the shapes of the underlying matrices, e.g., non-singularity, non-rectangularity, as well as by the numerical performance in terms of the economic rationale of the generated results. Jansen and ten

[11] For example, dimensionality, number of commodities, industries etc.
[12] Other strategies could have been adopted. For example, separately estimating the make and use matrices would better allow for the explicit representation of the joint products in the dairy industry.
[13] Note that it is a particular feature of the US IO statistics to distinguish between the competitive and non-competitive imports and this is not to be taken as a generally applied principle. In this exercise we focus only on the prior information obtained from the use and final demand tables. As the data on imports are not of importance for us the derived symmetric IO table does not take this difference into an account.

Raa (1990) provide a set of desirability axioms allowing for ranking of the available methods. In practice, despite its rather poor economics justification, square IO tables have often been constructed on the basis of industry technology assumption mainly because it does not require consecutive adjustments of the IO estimates.[14] The estimated US 1992 IO table obtained from the industry technology model is presented in Appendix 7.

## *3. Macroeconomic framework for dairy industry*

The US data as constructed in the previous section are used to proxy information on production cost shares describing the production technology of the six individual dairy commodities. To complete the overall macroeconomic framework within the dairy industry, the relative magnitudes of the components of final demand and the composition of trade flows remain to be established. Given the constraints on the total values of the components of final demand and trade accounts implied by the initial GTAP database, the resulting estimates cannot correspond to the figures reported in the official statistics. Official statistics are used to determine the relative contributions of the respective components. Relativities within final demand, such as the value share of SMP in the total value of dairy private demand, or the value share of SMP in the total demand for dairy products, are calculated from the supply and utilization databases maintained by FAO and EUROSTAT. The separate steps of the procedure are illustrated using Irish data.

### 3.1. Supply and utilization of dairy commodities

The FAO and Eurostat databases provide a collection of time series of international statistics on agricultural production, food balances, international trade and others.[15] In addition to the main macroeconomic indicators such as production, import, export and domestic demand, a detailed record on commodity utilization is provided. Most of the commodities are either consumed or utilized for further processing into the higher value added products. The amount entering further utilization is determined by the given fixed extraction rate. A fixed share of the total production of each commodity is assumed to be wasted within the production process. The remainder, defined as total domestic demand, is then allocated between final use, feeding and further processing. The FAO database does not necessarily report all the product components for all the

[14] The industry technology model violates the cost minimization premise in economics. However despite this drawback its numerical tractability still provides sufficiently strong motivation for further use of the method, e.g., Sue Wing (2006).
[15] Data are available on the web site http://faostat.fao.org/ accessed in May 2006.

commodities therefore some commodity accounts are not complete. The missing values were calculated from the underlying identity:

production + imports + stocks – export – domestic supply = 0

The commodity supply and utilization balances typically report the data in metric tons, while data on international trade are available in the form of both quantities as well as in value terms. As we are seeking to determine the relative magnitudes among the demand and cost items of the dairy commodities in value terms, all indicators have to be expressed in value terms using the appropriate price indices. The conversion requires information on consumer and producer prices (or unit values).[16] The available data on dairy prices are incomplete and often inconsistent. Moreover, the quoted prices include various tax instruments, margins and rents, on which information is mostly not available. Ideally, the calculated value of production of each dairy commodity would equal the value of domestic production delivered to the domestic market plus the value of exports. In a similar vein, on the demand side the value of the consumer composite has to be equal to the value of imports and the value of the domestic production allocated to the domestic market. The price data collected from the diverse sources clearly do not ensure that the desired equalities hold and therefore have to be adjusted. The other source of adjustments is the treatment of the main dairy products as composite products, e.g., cheese is treated as a composite of the various cheese types with different prices and fresh dairy products is also a composite product across a number of HS lines.[17] The composite products face a theoretical single composite market price, which has to be established in line with the above mentioned identities.[18] The definitions of the various composite commodities as well as statistical enquiry methods differ widely among countries, making the statistics difficult to compare. For example, the fresh milk product composite in Germany differs in terms of the HS line product composition from the fresh milk products reported for Ireland. The prices of dairy commodities used are given in Table 3.1.

[16] Data on prices were collected from USDA, Eurostat (available only for some EU countries), and ZMP. The FAO database provides data only on the raw milk producer price. An alternative source for fresh milk prices is LTO Netherlands.

[17] An aggregate of the selected products from the CN 2000 codes 0401, 0403 and 2202 are reported as fresh dairy products. For further information consult EUROSTAT classifications.

[18] Within the category of fresh milk products there is in fact no public database on prices and no actual price quotation. In the case of cheese, prices of various types of cheese are available, but the composite price used for the modeling purposes is in fact a theoretical concept.

Table 3.1 Dairy commodity consumer prices in 2001 USD per metric tonne (EUR / USD exchange rate 1.12)

| | France | Benelux | Germany | UK | Europe | World | USA | Oceania |
|---|---|---|---|---|---|---|---|---|
| Smp | 2133.9 | 2232.1 | 2026.8 | 2581.3 | 2243.5 | 1975.0 | 2173.3 | 2043.3 |
| Wmp | 2419.6 | 2500.0 | 2562.5 | 2494.0 | 2494.0 | 1954.0 | 3111.8 | 1975.5 |
| Whp | 473.2 | 491.1 | 491.1 | 485.1 | 485.1 | 572.0 | 424.0 | 424.0 |
| Chs | 4482.1 | 3071.4 | 3517.9 | 3348.2 | 3604.9 | 2170.0 | 3171.3 | 2119.4 |
| But | 2776.8 | 2848.2 | 2875.0 | 2726.8 | 2806.7 | 1393.0 | 3664.0 | 1293.3 |
| Fmk | 283.28 | 334.7 | 305.0 | 271.23 | 298.6 | 267.0 | 267.0 | 267.0 |

Source : ZMP (2004), Eurostat

Data on production and consumption of most dairy commodities are well covered in the statistics. Nearly all data on production of dairy commodities were collected from the FAO database. Some difficulty was experienced while collecting the data on the production and consumption of fresh milk products. In EU member countries, these were taken from *ZMP (2004)* and Eurostat. Most of the available data are presented on a very aggregated level, without any reference to the HS product lines included in the aggregation, which constrains the comparability and interpretation of the figures. Very little data were available on the production and consumption of fresh dairy products in regions outside the EU.[19]

The calculation of the shares portraying the demand and cost relativities within each regional SAM is now straightforward. Table 3.2 shows that, for Ireland, more than 50% of the production of most dairy commodities is allocated to export markets, apart from exports of the largely nontradable fresh milk products which account for only 5% of total demand. Table 3.2 illustrates the calculated shares of imports in total production costs in Ireland. For example, 10.55 % of the total production costs of skim milk powder can be attributed to imports.

Table 3.2 Initial cost shares in supply of dairy products

| | Domestic supply | Exports |
|---|---|---|
| Smp | 0.529 | 0.470 |
| Wmp | 0.135 | 0.865 |
| Whp | 0.002 | 0.998 |
| Chs | 0.264 | 0.736 |
| But | 0.225 | 0.775 |
| Fmk | 0.950 | 0.050 |

Source: FAOSTAT, EUROSTAT, ZMP and authors' calculations.

[19] For some regions the total domestic supply is obtained by the multiplication of the per capita consumption levels by the population figures. In order to maintain the supply-demand equilibrium the resulting figures were adjusted. The EU member states are relatively well covered by the Eurostat data.

Table 3.3 Initial shares in final demand for dairy products

| | smp | wmp | Whp | Chs | but | fmk |
|---|---|---|---|---|---|---|
| Production | 0.895 | 0.862 | 0.859 | 0.6209 | 0.924 | 0.962 |
| Imports | 0.106 | 0.138 | 0.1408 | 0.379 | 0.0762 | 0.038 |

Source: FAOSTAT, EUROSTAT, ZMP and authors' calculations.

This completes the exposition of the calculation of shares. Our next step is to construct each region's SAM, ensuring that their estimated elements satisfy the underlying GTAP zero profit and market clearance conditions, while taking values reasonably close to the empirically-determined benchmark cost and expenditure shares respectively. We must also ensure that the trade elements of each SAM are consistent with each other in this process. A number of cross entropy type of programs have been employed to estimate the adjusted shares. These are introduced in the following section.

### *4. Estimation methods*

The ideal technique to estimate detailed economic-technological data requires a combination of econometric estimation methods and expert judgments. The information from the expert judgments enters the econometric procedures in the form of side constraints and upper or lower bounds on the particular variables. The availability of the required data on dairy production in practice is very limited. Consequently, the problem of estimating the missing parameters then often turns out to be underdetermined, i.e., ill-posed.[20] Various methods and techniques have been developed and applied to the estimation of ill-posed problems such as RAS[21] algorithms and cross entropy procedures. Their key idea is to define the mathematical program minimizing the Kullback - Leibler cross entropy "distance" (Kullback and Leibler, (1992) between any available prior information and the estimated matrix subject to the required constraints. The cross entropy procedure employs the specifications of the objective functions axiomatically introduced by Shannon (1948)[22] who defines a unique function (entropy function[23]) to measure the probability

[20] Note that we are attempting to identify $n^2$ unknowns, but have 2n-1 independent row and column adding up restrictions.

[21] The RAS method takes its name from the notation in Stone's (1962) original equations. The RAS method is used to update existing I-O tables to relate to a year for which intermediate input (column) sums are known but not the intermediate deliveries themselves. The simple RAS method consists of finding a set of multipliers to adjust the rows of the existing matrix, and a set of multipliers to adjust the columns so that the cells in the adjusted matrix will add up to the given row and column totals relating to the chosen update year.

[22] For a survey of procedures to balance square matrices see Schneider and Zenios (1989). The main difference among the methods is the actual definition of the objective function and therefore the actual

of a collection of events. Golan et al. (1994) provide a comparison of different formulations of the entropy procedures. They start with the basic maximum entropy formulation and elaborate their framework into the cross entropy problem by adding the relevant prior information into the objective function. At a later stage, by replacing the initial point prior estimates with a prior permitting a discrete probability distribution to be attached to each a(i,j), they reformulate the pure inverse maximum entropy problem into generalized maximum and cross entropy formulations. While the value of the entropy itself represents a measure of the remaining uncertainty, the generalized formulation delivers estimates of the information on the probability distribution and uncertainty measure for each a(i,j). Further extensions have been worked out by Robinson et al. (1998), who deal with the particular application of the cross entropy to estimate a social accounting matrix. In addition to the "standard" additivity constraints they add all available information into the program. They distinguish among i) *priors* which is essentially the matrix $A_{ij}$ ii) moment constraints presented usually in the form of all or some of the row and column totals iii) economic aggregates and iv) inequality constraints.

### *4.1 Formulation of the cross entropy method*

In this section we outline the formulations of the cross entropy problem applied in the study. The Shannon entropy defined in Section 2.1 is a subset of the Kullback-Leibler directed divergence or cross-entropy.[24] Technically the cross entropy procedure selects an n - tuple from all available n - tuples satisfying the additivity constraints with the highest probability $\pi\left(n\middle|\beta^0\right)$ of the realization by the given prior information.[25] Denote a set $n \equiv \left(N_{11}, N_{22}, ..., N_{ij}, ..., N_{KK}\right)$ to be a vector of the estimated SAM cells, $n^0 \equiv \left(N_{11}^0, N_{22}^0, ..., N_{ij}^0, ..., N_{KK}^0\right)$ to be a vector of the prior SAM cells, $N_j^0 = \sum_{i=1}^{K} N_{ij}^0$ and $\beta_{ij}^0 = \frac{N_{ij}^0}{N_j^0}$.

---

theoretical roots of the various approaches. Robinson et al. (1998) argue that the important advantage of the cross entropy approach is its substantial theoretical background rooted in information theory in contrast to the free application of the various penalty functions as suggested by the other approaches.

[23] The entropy concept has its origins in the work of Boltzman (1870) as well as Bernoulli.

[24] Note that the cross-entropy "distance" is not to be confused with a norm.

[25] The sign $\left|\right.$ refers to the Bayesian "subject to".

Our aim is to find:[26] $\arg\max_{n} \pi\left(n|\beta^0\right)$

where $$\pi\left(n|\beta^0\right) = \prod_{j=1}^{K} \frac{N_j!}{\prod_{i=1}^{K} N_{ij}!} \prod_{i=1}^{K} \left(\beta^0_{ij}\right)^{N_{ij}}$$

After log-transformation, Stirling approximation[27] and some reformulation we obtain:[28]

$$\ln \pi\left(n|\beta^0\right) \approx -\sum_{j=1}^{K} N_j \sum_{i=1}^{K} \beta_{ij} \ln\left(\frac{\beta_{ij}}{\beta^0_{ij}}\right)$$

Dividing by N and holding $w_j = \frac{N_j}{N}$ we obtain the following formulation of the cross entropy problem :

$$\min_{\beta} \sum_{j=1}^{K} w_j \sum_{i=1}^{K} \beta_{ij} \ln\left(\frac{\beta_{ij}}{\beta^0_{ij}}\right)$$

subject to
$$\beta_{ij} * y_j = x_i$$
$$\sum_i \beta_{ij} = 1$$
$$0 \le \beta_{ij} \le 1$$

The solution of the cross entropy obtained from the first order conditions reads as follows:[29]

$$\beta_{ij} = \frac{\beta^0_{ij} \exp\left(-\lambda_i N^1_j\right)}{\sum_{i=1}^{K} \beta^0_{ij} \exp\left(-\lambda_i N^1_j\right)}$$

[26] Note that the standard formulation of the cross entropy procedure would not deliver the "optimal" solution when the estimated matrix is not initially balanced i.e. the row and column totals are not equal, as it is usual the case with SAM matrices. The formulation of the objective function has to be accordingly adjusted to take this into the account. The "standard" definition of the cross entropy problem as defined in terms of probabilities i.e. $p_{ij} = N_{ij} / N$. The application of the cross entropy to SAM however requires that the estimated coefficients are defined as $p_{ij} = N_{ij} / N_j$.

[27] See footnote 7.

[28] See appendix 8 for the derivation.

[29] See appendix 8 for derivation.

### *4.2 Alternative methods*

In addition to the standard cross-entropy procedure, experiments with a number of alternative methods were carried out.[30] The purpose in applying alternative methods was twofold. First, the comparison of various methods allowed us to select the best available estimate. Second, in some cases different procedures were needed to solve some particular sub-problems where the cross entropy procedure either could not be applied or led to unsatisfactory results. In fact, meeting additional constraints was often impossible with a prior benchmark dataset. This led to the failure of the standard cross-entropy program to find a feasible solution, while other methods turned out to be more successful. This section outlines the actual procedure to disaggregate the dairy sector. We start by splitting the bilateral trade flows matrix. The total imports and exports of dairy commodities establish the trade vectors in each regional SAM. Applying the shares for cost and final demand presented in earlier sections allows us to estimate the total production of the six dairy commodities and finally estimate the regional SAM's.

#### *4.2.1 Break down of the trade flow arrays*

The GTAP database accommodates international trade flows by means of two pairs of arrays denoting the values of imports and exports respectively among all regions on a bilateral basis. The pairwise exposition allows all trade flows to be expressed both in world and internal market prices. The difference between the values traded expressed in world and market prices is accounted for by revenues acquired from tariffs, other trade protection measures and export taxes. In addition, the GTAP database provides two arrays of data on international trade services. The trade benchmark equilibrium is established when export of the commodity i from region s to region r valued in world prices plus transport margin charged on shipping services between the regions is equal to the import supply of the commodity i from region s in region r. The major difficulty in estimating consistent bilateral trade flow matrices arises from the scarcity of available prior information in particular on trade margins. In addition, different statistical and

[30] For example, see Sue Wing (2006) who has undertaken a similar exercise to ours, where he attempted to break down the energy sectors in the GTAP database and linked the aggregate numbers with bottom up information on the particular energy producing technology. His approach goes back to the positive mathematical programming due to Howitt (1995) where the objective function given by the square distance between the prior and estimated cost shares is minimized subject to adding up constraints.

trade administration practices among countries are other reasons for diverging indices on trade flows.[31]

The COMTRADE database, a major statistical resource for the GTAP trade data, has been used as prior information for the break down of the GTAP dairy trade accounts. COMTRADE is a fundamental database on bilateral world trade maintained by the United Nations.[32] The COMTRADE database allows users to extract bilateral data on CIF import and FOB export values. The data on exports are reported with and without adjustment for re-exports. First, the extracted data were aggregated to the required level.[33] Starting with a GTAP array giving the bilateral trade of each commodity in world prices we obtain a bilateral trade matrix for the GTAP dairy composite commodity. The row entries report each country's total dairy exports to the other trading regions, while column cells represent their value of imports. As there is no available information on the composition of trade with dairy products, shares splitting the total dairy exports from each region had to be established. The trade data extracted from COMTRADE were employed to determine the value shares of imports and exports of each dairy commodity in total imports and exports. Subsequently, the regional export shares were applied to break down the GTAP total values of dairy exports (i.e. row totals of the bilateral trade matrix) into six dairy commodities for each region.[34] The resulting export shares indicating the value share of the exports of the six specific dairy commodities in all regions in the total value of dairy exports are reported in the Table 4.2.1.1.

Table 4.2.1.1 Export shares of individual dairy commodities in total dairy exports

| | aus | xas | Ben | eun | Eus | xee | cee | row | usa | xsm | fra | deu | Gbr | irl | Xef | rus | xeu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fmk | 0.03 | 0.24 | 0.15 | 0.18 | 0.19 | 0.11 | 0.10 | 0.11 | 0.08 | 0.15 | 0.19 | 0.35 | 0.26 | 0.05 | 0.00 | 0.09 | 0.47 |
| But | 0.13 | 0.03 | 0.12 | 0.16 | 0.07 | 0.18 | 0.09 | 0.03 | 0.05 | 0.04 | 0.06 | 0.04 | 0.15 | 0.30 | 0.04 | 0.14 | 0.02 |
| Smp | 0.19 | 0.14 | 0.06 | 0.06 | 0.05 | 0.07 | 0.26 | 0.73 | 0.28 | 0.09 | 0.04 | 0.12 | 0.06 | 0.12 | 0.00 | 0.21 | 0.03 |
| wmp | 0.37 | 0.51 | 0.10 | 0.16 | 0.05 | 0.10 | 0.11 | 0.10 | 0.12 | 0.47 | 0.14 | 0.06 | 0.20 | 0.12 | 0.01 | 0.08 | 0.06 |
| Whp | 0.02 | 0.02 | 0.04 | 0.03 | 0.02 | 0.03 | 0.05 | 0.01 | 0.21 | 0.02 | 0.07 | 0.05 | 0.05 | 0.06 | 0.00 | 0.01 | 0.03 |
| Chs | 0.27 | 0.06 | 0.53 | 0.42 | 0.61 | 0.50 | 0.38 | 0.02 | 0.26 | 0.22 | 0.50 | 0.39 | 0.28 | 0.36 | 0.95 | 0.48 | 0.39 |

Source : own calculation using COMTRADE database

[31] Time lags between the records on imports and exports, trade via third countries, length of the customs warehousing periods etc.
[32] Extensive and detailed description of the COMTRADE statistics can be found in UN (2004).
[33] COMTRADE reports trade flows on the HS 6 digit basis. For the aggregation see Appendix 9.
[34] Both shares expressing the commodity composition of dairy imports and exports for each region were calculated. For the breakdown of the GTAP bilateral trade matrix, in view of the numerical performance of the cross entropy program, only the export shares were used to establish each region's total exports of each of the dairy commodities.

At this stage the cross entropy program set out in Section 4.2.1.2 below was employed to estimate trade flows denominated in world prices. The formulation of the problem ensures that the overall balance in international trade is met, i.e., that for each product the relevant import and export totals across all regions are equal to the trade aggregates in the regional SAM's. The programming problem minimizing the entropy objective subject to the underlying identities and adding-up constraints reads as follows:[35]

**4.2.1.2 Program to estimate the trade flows expressed in the world prices**

$$\min\left(\sum_{q=1}^{Q}\sum_{s=1}^{S} viws_{q,s}\ln\left(\frac{viws_{q,s}}{viws0_{q,s}}\right)\right)$$

$$\sum_{s} viws(q,s) = share(q,r)*GTAP\_trade\_total(r) \quad \forall r \in (regions)$$

$$\sum_{q} viws(q,s) = 1$$

The estimated matrix expressing Irish trade of dairy commodities denominated in world prices is depicted in table 4.2.1.3.

Table 4.2.1.3 Irish trade with dairy commodities (in billion 1997 USD)

| | aus | xas | ben | eun | eus | xee | cee | row | usa | xsm | fra | deu | gbr | xeu | xef | rus |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fmk | | 0.27 | 0.56 | 0.68 | 0.90 | | 0.50 | | | | 0.55 | 2.07 | 48.40 | | 0.03 | |
| But | | 3.66 | 67.74 | 7.62 | 6.76 | 0.78 | | 1.16 | 1.19 | 5.65 | 58.93 | 111.49 | 71.49 | 0.14 | 1.44 | |
| Smp | | 62.72 | 19.05 | 0.16 | 5.85 | | | 25.47 | 0.57 | 7.25 | 3.14 | 6.00 | 3.91 | | 0.34 | |
| wmp | 0.26 | 53.71 | 36.50 | 1.87 | 3.46 | | | 15.07 | 0.00 | 21.26 | 1.00 | 2.55 | 0.89 | | 0.08 | |
| Whp | | 18.90 | 22.12 | 0.53 | 5.19 | | | 0.93 | 0.14 | | 3.21 | 4.12 | 6.49 | 0.78 | 0.09 | 0.92 |
| Chs | | 19.84 | 9.73 | 2.08 | 26.66 | 0.67 | 0.38 | 2.32 | 25.17 | 1.79 | 15.01 | 13.90 | 294.07 | 0.74 | 0.71 | |

A similar procedure described in Section 4.2.1.3 has been applied to estimate the trade margins. By assumption the distribution of margins follows the trade patterns and the COMTRADE trade data was repeatedly used as prior information. A minor adjustment in the formulation of the cross entropy problem introduced in the form of an additional constraint ensures that the margins expressed in value terms do not exceed the value of trade flows denominated in world prices.

[35] The notation here should not be confused with the original notation in GTAP. VIWS in GTAP is an array for world imports denominated in world prices and it is three dimensional. Here we use the same acronym for the estimated shares of dairy commodity trade in total dairy trade as given in the GTAP database.

#### 4.2.1.3 Program to estimate the trade margins

$$\min\left(\sum_{q=1}^{Q}\sum_{s=1}^{S} vtwr_{q,s}\ \ln\left(\frac{vtwr_{q,s}}{vtwr0_{q,s}}\right)\right)$$

$$\sum_{s} vtwr(q,s) * GTAP\_mil\_vtwr(s) = share_{q} * GTAP\_m\arg in\_total$$

$$\sum_{q} vtwr(q,s) = 1$$

$$\sum_{s} vtwr(q,s) * GTAP\_mil\_vtwr(s) \le viws(q,s)$$

To complete the estimation of the balanced bilateral trade flows and margin matrices all that remains is to apply the relevant tariffs and calculate the remaining arrays denoting trade flows in market prices. In this version of the extended dairy database, tariffs are assumed to be equal for all dairy commodities with the same effective rate as implicitly given by the GTAP aggregate dairy sector. This is clearly a drawback for policy simulations and this assumption will be relaxed in subsequent versions of the database. The resulting matrix of trade and transport margins associated with Irish trade with dairy commodities is presented in Table 4.2.1.4.

Table 4.2.1.4 Irish trade and transport margins associated with dairy commodities trade (in billion 1997 USD)

| | aus | xas | ben | eun | eus | xee | cee | row | usa | xsm | fra | deu | gbr | xeu | xef | rus |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| fmk | | 0.06 | 0.01 | 0.02 | 0.02 | | 0.03 | | | | 0.01 | 0.05 | 1.48 | | 0.00 | |
| but | | 0.89 | 1.72 | 0.24 | 0.17 | 0.03 | | 0.09 | 0.05 | 0.37 | 1.37 | 2.68 | 2.83 | 0.00 | 0.07 | |
| smp | | 1.98 | 0.33 | 0.00 | 0.13 | | | 1.20 | 0.02 | 0.32 | 0.06 | 0.10 | 0.03 | | 0.02 | |
| wmp | 0.00 | 1.68 | 0.70 | 0.06 | 0.08 | | | 0.71 | 0.00 | 0.95 | 0.02 | 0.05 | 0.01 | | 0.00 | |
| whp | | 1.11 | 0.42 | 0.02 | 0.12 | | | 0.05 | 0.00 | | 0.06 | 0.07 | 0.08 | 0.02 | 0.00 | 0.03 |
| chs | | 3.23 | 0.21 | 0.07 | 0.64 | 0.03 | 0.02 | 0.16 | 0.96 | 0.11 | 0.32 | 0.29 | 6.79 | 0.02 | 0.04 | |

Source : own calculations

### *4.2.2 Positive mathematical programming*

Having estimated the bilateral trade flows, the regional vectors of dairy imports and exports can be easily calculated as row and column sums of the trade matrix respectively. Hereby we have obtained the initial vectors of the final regional SAM's consistent with the trade matrix interlinking all regional SAM's. The next step is to establish the totals for each of the dairy

accounts. It is desirable that the value shares, e.g., the estimated value of imports of cheese to the value of total dairy imports or the value share of imports of cheese to total cheese production, reflect reasonably closely the observed relativities for the dairy production of each region. While the former has been determined by the estimated trade matrix, the latter proportions were obtained from the FAO database and may now be employed to break down the GTAP value of total dairy production.

Following the approach suggested by Sue Wing (2006) developed along the lines of Howitt (1995) we define a criterion function capturing the weighted square divergence between the calculated and benchmark cost shares. The objective function is minimized subject to constraints imposing the desired trade-production patterns in each of the regions. The total value of dairy production remains unchanged and equal to the GTAP value. The formulation of the estimation program to calculate the total production of each dairy commodity in each region reads as follows:

#### 4.2.1.1 Program to calculate the SAM totals for dairy accounts

$$\min \sum_{i=1}^{K} col\_share0_i * \left( \frac{col\_share_i}{col\_share0_i} - 1 \right)^2 + \sum_{i=1}^{K} row\_share0_i * \left( \frac{row\_share_i}{row\_share0_i} - 1 \right)^2$$

$$col\_share_i * total_i = import_i$$
$$row\_share_i * total_i = \exp ort_i$$
$$\sum_i total_i = GTAP\_total$$

The results of this sub-program show the greatest divergences from the prior data within the entire procedure, indicating the double constraining nature of the estimation problem. As can be seen from 4.2.1.1, constraints on both the row as well as on the column totals have to be fulfilled. This clearly increases the difference between the benchmark and estimated coefficients. Despite a major effort to keep the shares of imports and exports relative to total cost and total demand, respectively, reasonably close to their benchmark values, the differences are large for some regions. The calculated totals for each region are presented in Table 4.2.1.1.

**Table 4.2.1.1 Estimated total value of supply of dairy commodities, US million dollars, 2001**

| | Fmk | But | Smp | Wmp | Whp | Chs | Total |
|---|---|---|---|---|---|---|---|
| Aus | 795.78 | 1782.27 | 1942.17 | 1812.69 | 85.19 | 2781.58 | 9199.68 |
| Ben | 3245.18 | 912.50 | 1015.00 | 1699.57 | 903.69 | 5465.83 | 13241.77 |
| Cee | 1721.04 | 2007.08 | 837.10 | 785.78 | 736.16 | 1332.17 | 7419.34 |
| Deu | 6379.55 | 3684.08 | 3623.19 | 3025.40 | 2651.51 | 8992.36 | 28356.07 |
| Eun | 1881.57 | 1719.40 | 953.70 | 1557.38 | 750.05 | 2908.48 | 9770.58 |
| Eus | 7440.66 | 3371.44 | 3606.53 | 2762.77 | 1783.67 | 9547.37 | 28512.44 |
| Fra | 4220.30 | 3461.39 | 2393.13 | 3174.75 | 2536.00 | 6986.13 | 22771.69 |
| Gbr | 4222.50 | 3668.46 | 1927.98 | 3132.11 | 1763.52 | 7227.79 | 21942.36 |
| Irl | 474.10 | 749.49 | 484.54 | 474.53 | 331.39 | 886.57 | 3400.63 |
| Rus | 1642.66 | 2415.01 | 1798.41 | 1699.12 | 564.90 | 4142.08 | 12262.18 |
| Usa | 9848.85 | 12171.99 | 15266.75 | 11506.52 | 14908.48 | 30626.53 | 94329.13 |
| Xas | 6653.06 | 5212.57 | 9057.92 | 12545.20 | 4841.91 | 8703.71 | 47014.36 |
| Xeu | 793.50 | 281.61 | 241.34 | 397.41 | 228.85 | 1015.36 | 2958.06 |
| Row | 1381.21 | 704.66 | 1357.48 | 1565.69 | 1337.99 | 302.90 | 6649.93 |

### *4.2.2 Cross entropy with additional macro constraints*

The final step in the estimation of the regional SAM's is to combine all the now available information and to perform the final adjustment of the values to achieve consistency for each regional SAM. This step could in principle be omitted and easily resolved by the above mentioned SPLITCOM utility (Horridge, 2005). Its use, however, requires separate knowledge of the splitting shares for final demand between the domestic market and imports. Our prior data do not provide such a detailed insight and only mediate the splitting shares for the Armington composite. Therefore the overall consistency of the database has been achieved by a so called "flexible" cross entropy approach proposed by Robinson et al. (1998). Unlike the traditional RAS, the flexible procedure does not require starting from a consistent SAM. It allows to estimate a consistent SAM from inconsistent data estimated with error. In addition, the method allows for incorporating additional inequality constraints as well as any sort of prior knowledge about any particular cell of the SAM. The definition of the error terms follows Golan et al (1994) who defined the error term as a weighted sum of elements of the support set. The three sigma rule applies to the benchmark value of the support vector and the weights can subsequently be calculated to yield the prior standard error $\sigma$.[36] The resulting regional SAMs and trade flows are made available in the form of HAR/Excel/GAMS files.[37]

[36] A rule of thumb indicating that the values of a normally distributed random variable would not differ from their expected values by a magnitude exceeding $3\sigma$. See Pukelsheim (1994) and for application in cross entropy related estimation Golan et al (1996).

[37] These files can be downloaded from the website www.tcd.ie/iiis/pages/people/researchfellows_mraz.php.

#### 4.2.2.1 Cross entropy program to estimate the disaggregated regional SAM's

$$\min_{\alpha} \sum_{i=1}^{K}\sum_{j=1}^{K} \alpha_{ij} \ln\left(\frac{\alpha_{ij}}{\alpha_{ij}^{0}}\right) + \sum_{i=1}^{K}\sum_{j=1}^{K} W_{ij} \ln\left(\frac{W_{ij}}{W_{ij}^{0}}\right)$$

$$Y_i = X_i + ERR_i$$

$$SAM_{ij} = \alpha_{ij}\left(X_j + ERR_j\right)$$

$$\sum_j SAM_{ij} = Y_i$$

$$\sum_i SAM_{ij} = X_j + ERR_j$$

$$\sum_i SAM_{ij} = 1$$

$$ERR_i = \sum_{jwt} W_{i,jwt} * vbar_{i,jwt}$$

$$\sum_{jwt} W_{i,jwt} = 1$$

## *5. Conclusions*

This paper develops a procedure to estimate the splitting shares to obtain a consistent disaggregation of the dairy industry account in the GTAP database. The disaggregation provides a detailed treatment of the six main dairy commodities. Having outlined the estimation procedure in detail in the previous sections, we briefly summarize the efforts behind the database breakdown, assess the "goodness" of the available estimates and the general applicability of the method which we have used.

Estimating the shares to break down a particular sector in the global database is a very constrained exercise.[38] For studies requiring a single country or regional focus it might well be justified to develop a new database from scratch rather than trying to impose prior relativities on existing SAM sub-totals which are themselves the result of a balancing process and which thus often deviate from or even conflict with the observed evidence. To address global issues such as world trade liberalization scenarios, the use of a detailed global database is inevitable and

[38] The constraining nature of the problem follows the initial setting of the problem. We aim to estimate splitting shares reasonable close to the collected prior information (1st constraint), reasonably reflecting the country specific production and trade patterns imposed by the side constraints (2nd constraint) and complying with the entire GTAP database i.e. totals in both directions have to match the GTAP total values from the dairy account (3rd constraint).

disaggregating an available database is the only feasible solution. In assessing the outcome from such an exercise it is important to bear in mind the constraints involved in creating a consistent global database.

The principal source of bias in the estimates is a general scarcity of statistics describing the dairy industry in sufficient detail. The model treatment of the dairy commodities as composites requires the prior supply and utilization data to be aggregated across a number of HS product lines. Fresh milk products gave rise to particular problems due to the persistent differences among the available statistics or their absence. Average producer and consumer prices for these dairy commodities are also required and their computation involves some heroic assumptions. The average prices used for the conversion of the supply and utilization balances into monetary value terms can only be regarded as rather crude estimates. Regarding the quality and accuracy of the prior IO data, the 1992 US IO tables were the only available source of information describing the cost shares for the major dairy commodities in sufficient detail. Even so, specific information distinguishing between milk powders, for example, is not available in this source. Also, the US dry dairy production composite includes evaporated and condensed products, while in our database these are components of the fresh dairy commodity composite.

Appendix 11 presents some descriptive characteristics assessing the goodness of the fit of the resulting shares. The "flexible" formulation of the cross entropy procedure allowing for error in the prior data was able to solve for all regions. The results are mainly determined by the degree of constraint implied by the adding up constraints of the problem and the absolute value of the column and row totals. Because the totals given in the GTAP database often conflict with the observed evidence, this frequently leads to reversal in the relative importance of cell elements among the estimates.

The paper also contributes to the discussion on the appropriate size and detail of the global database for policy modeling. It provides a procedure to disaggregate selected accounts within a global database, while utilizing the maximum of the information collected from other sources. The procedure may well serve as a complementary utility for the generation of the partial or full rank splitting weights for the sectoral break down as required by the SPLITCOM utility. The two procedures approach the disaggregation procedure from different starting points. SPLITCOM initially assumes a partial or full knowledge of the splitting weights and proceeds with the break down of the selected sectors. The final splitting shares are calculated ex post after the consistency of the extended database has been restored by the RAS type of adjustments. The prior information on the cost and demand shares in the dairy industry computed in the first half of the paper could have been used to construct the partial rank splitters for SPLITCOM. In our approach, however,

we have aimed to elaborate on the estimation method and by means of the more flexible extended cross entropy procedure to obtain splitting weights as close as possible to the relativities within the cost and demand structure of the dairy industry. A comparison between the results estimated by the method introduced in this paper and SPLITCOM, while using the same prior data are presented in the Appendix 12.

**References**


Boltzman, L., 1877, Wien. Ber. 76, 373, English translation : J. Le Roux 2002

Brooke, A., Kendrick D., Meeraus A., 1998. GAMS a User's guide. San Francisco, The scientific press

Drud, A., 1985. CONOPT: a GRG code for large sparse dynamic nonlinear optimization problems, Mathematical programming 31, 153-191.

Drud, A., 1994. CONOPT: a large scale GRG code, ORSA. Journal of computing 6, 207-216

Drud, A., 1997. Interactions between nonlinear programming and modeling systems. Mathematical programming 79, 99-123

Elbehri, A., Hoffman, L., Ash, M. and Dohlman, E., 2001. Global impacts of zero-for-zero trade policy in the world oilseed market: a quantitative assessment, Paper presented at the Fourth Conference on Global Economic Analysis, Purdue University, West Lafayette, Indiana, June 26-29, 2001.

Eurostat database accesed at http://faostat.fao.org

Faostat database accesed at http://faostat.fao.org

Feller, W., 1957. An introduction to probability theory and its applications, 2nd ed., John Wiley, NY.

Golan, A., Judge, G., Robinson, S., 1994. Recovering information from incomplete or partial multisectoral economic data. Review of economics and statistics 76, 541-9

Golan, A., Judge, G., Miller, D., 1996. Maximum entropy econometrics, robust estimation with limited data. John Wiley & Sons.

Hertel, T.W., 1997. Global trade analysis : Modeling and applications. Cambridge university press

Horridge, M. 2005. SplitCom, Programs to disaggregate a GTAP sector, CPS Monash University, Australia, available at http://www.monash.edu.au/policy/splitcom.htm.

Howitt, R.E., 1995. Positive mathematical programming, American journal of agricultural economics 77 : 329-342.

Jansen P., ten Raa, Th., 1990. The choice of model in the construction of input-output coefficient matrices. International economic review 31,1, 213-227

Jaynes, E. T., 1957. Physics review, 106, 620

Jaynes, E. T., 1988. in Erickson G.J, Smith C.R. (eds), Maximum entropy and Bayesian Physics review, 106, 620

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 4, 99-111

Pukelsheim, F., 1994. The three sigma rule. The American Statistician, Vol. 48, No. 2 (May), pp. 88-91.

Robinson, S., Cattaneo, A., El-Said, M., 1998. Estimating a social accounting matrix using cross entropy methods, TMD discussion paper No.33, Washington, International Food Policy Research Institute.

Schneider, M., Zenios, S., 1989. A comparative study of algorithms for matrix balancing, Operations Research, May – June 1990; 38, 3

Shannon, C. E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379-423

Stone, R., 1961. Input-output and national accounts. Paris OECD

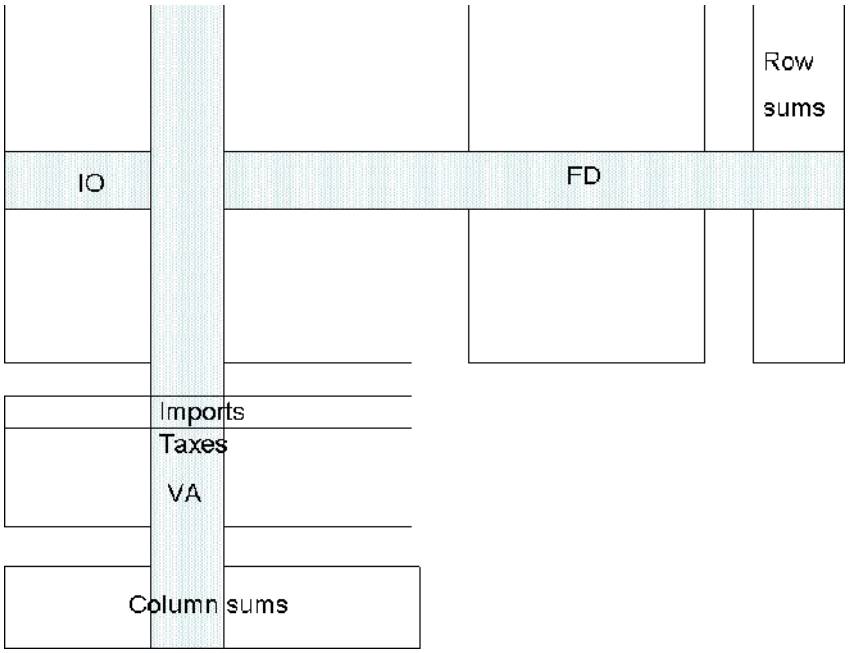
Sue Wing, I., (2006). The synthesis of bottom-up and top-down approaches to climate policy modelling : Electric power technology detail in social accounting matrix. Downloaded from thtp:// people.bu.edu/isw/papers/top-down_bottom-up_sam.pdf accessed 25/03/2007

ten Raa, T., Rueda Cantuche, J. M., 2003. The construction on input-output coefficients matrices in an axiomatic context : Some further considerations. Fundación Centro de Estudios Andaluces, Documento de trabajo E2003 / 30

US Department of Commerce 1998, Benchmark input-output accounts of the United States 1992

ZMP, 2004. Milchproduktion, ZMP, Bonn

## Appendix 1. Graphical exposition of splitting the dairy account in the regional IO tables



# Appendix 2

Regional structure of the model and benchmark prices of dairy commodities applied in the construction of the database.

| | | |
|---|---|---|
| Benelux | (ben) | Belgium, Luxembourg, The Netherlands |
| France | (fra) | |
| Germany | (deu) | |
| Ireland | (irl) | |
| United Kingdom | (gbr) | |
| Europe North | (eun) | Sweden, Finland, Denmark |
| Europe South | (eus) | Spain, Portugal, Italy, Greece |
| Rest of EU | (xeu) | |
| Central and Eastern Europe | (cee) | |
| Australia | (aus) | Australia and New Zeland |
| Russia | (rus) | |
| United States of America | (usa) | |
| Rest of Asia | (xas) | |
| Rest of the World | (row) | |

Consumer (Armington) prices of dairy commodities used in the study.

ADD TABLE

Source : Eurostat and FAO. Prices of wmp and whp in the UK are given as averages of the remaining EU countries. For the rest of the Europe a common EU price is applied, calculated as an average from the available prices. The US prices are the year averages published by USDA. WHP price for USA is taken from the OECD statistics. The world and Oceania price of WHP has been set to be equal to the US price. FMP price for US, Oceania and World has been taken as the price of US 2002 from FAPRI 2003.

# Appendix 3

GTAP sectoral classification

| | |
|---|---|
| OCR | Crops |
| CTL | Bovine, cattle and livestock |
| *rmk* | *Raw milk* |
| FRS | Forestry and fishing |
| ENE | Energy |
| OMT | Other meat products |
| VOL | Other plant products |
| *fmk* | *Fluid milk manufacturing* |
| *but* | *Creamery butter manufacturing* |
| *chs* | *Cheese manufacturing* |
| *pow* | *Dry, condensed, and evaporated dairy products* |
| OFD | Food products |
| LIN | Light industry |
| HIN | Heavy industry |
| MAN | Manufacturing |
| SCS | Services |
| TRD | Trade |
| OTP | Transport nec |

GTAP production factors

| | |
|---|---|
| fLand | Land |
| fUnSkLab | Unskilled labour |
| fSkLab | Skilled labour |
| fCapital | Capital |
| fNatlRes | Natural Resources |

GTAp relevant taxes

| | |
|---|---|
| tffLand | tax on land |
| tffUnSkLab | tax on unskilled labour |
| tffSkLab | tax on skilled labour |
| tffCapital | tax on capital |
| tffNatlRes | tax on natural resources |
| tinv | tax on investments |
| ty | tax on production |
| tc | tax on private consumption |
| ti | tax on intermediate demand |

# Appendix 4

| GTAP | dairy sectors | CPC classification |
|---|---|---|
| RMK | Raw milk | 0291 |
| MIL | Dairy products | 22 |

US dairy sectors applied for the split of the GTAP dairy sector (MIL)

| NAICS | | SIC | ISIC |
|---|---|---|---|
| 311511 | Fluid milk manufacturing | 2026 | 1520 |

- Acidophilus milk manufacturing
- Beverages, milk based (except dietary) manufacturing
- Buttermilk manufacturing
- Cheese, cottage manufacturing
- Chocolate drink (milk based) manufacturing
- Cottage cheese manufacturing
- Cream manufacturing
- Dips, sour cream based manufacturing
- Drink, chocolate milk manufacturing
- Eggnog fresh nonalcoholic manufacturing
- Eggnog nonalcoholic (except canned) manufacturing
- Flavored milk drinks manufacturing
- Fluid milk substitutes processing
- Homogenizing milk
- Milk based drinks (except dietary) manufacturing
- Milk drink chocolate manufacturing
- Milk pasteurizing
- Milk processing
- Milk substitutes manufacturing
- Fluid milk manufacturing
- Nondairy creamers liquid manufacturing
- Pasteurizing milk

- Sour cream manufacturing
- Sour cream substitutes manufacturing
- Whipped topping (except dry mix, frozen) manufacturing
- Whipping cream manufacturing
- Yogurt (except frozen) manufacturing

| NAICS | | SIC | ISIC |
|---|---|---|---|
| 311512 | Creamer butter manufacturing | 2021 | 1520 |

- Anhydrous butterfat manufacturing
- Butter manufacturing
- Butter, creamery and whey manufacturing
- Creamery butter manufacturing
- Whey butter manufacturing

| NAICS | | SIC | ISIC |
|---|---|---|---|
| 311513 | Cheese manufacturing | 2022 | 1520 |

- Cheese (except cottage cheese) manufacturing
- Cheese analogs manufacturing
- Cheese products imitation or substitute manufacturing
- Cheese spreads manufacturing
- Cheese imitation or substitute manufacturing
- Cheese natural (except cottage cheese) manufacturing
- Curds cheese made in a cheese plant manufacturing
- Dips cheese based manufacturing
- Processed cheeses manufacturing
- Spreads cheese manufacturing
- Whey raw liquid manufacturing

| NAICS | | SIC | ISIC |
|---|---|---|---|
| 311514 | Dry, condensed and evaporated dairy product manufacturing | 2023 | 1520 |

- Baby formula, fresh, processed and bottled manufacturing
- Beverages, dietary, dairy and nondairy based
- Casein, dry and wet manufacturing
- Condensed milk manufacturing
- Condensed evaporated or powdered whey manufacturing
- Cream, dried and powdered manufacturing
- Dairy food canning
- Dehydrated milk manufacturing
- Dietary drinks, dairy and nondairy based manufacturing
- Dry milk manufacturing
- Dry milk products and mixture manufacturing
- Dry milk products for animal feed manufacturing
- Eggnog canned nonalcoholic manufacturing
- Evaporated milk manufacturing
- Feed grade dry milk product manufacturing
- Ice cream mix manufacturing
- Infant's formulas manufacturing
- Lactose manufacturing
- Malted milk manufacturing
- Milk based drinks dietary manufacturing
- Milk concentrated, condensed, dried, evaporated and powdered manufacturing
- Milk malted manufacturing
- Milk powdered manufacturing
- Milk ultra-high temperature manufacturing
- Milkshake mixes manufacturing
- Mix ice cream manufacturing
- Nondairy creamers dry manufacturing
- Nonfat dry milk manufacturing

- Powdered milk manufacturing
- UHT milk manufacturing
- Whey condensed, dried, evaporated and powdered manufacturing
- Whipped topping, dry mix, manufacturing
- Yogurt mix manufacturing

| NAICS | | SIC | ISIC |
|---|---|---|---|
| 311520 | Ice cream and frozen dessert manufacturing | 2024 | 1520 |

- Custard frozen manufacturing
- Desserts frozen (except bakery) manufacturing
- Frozen custard manufacturing
- Frozen desserts (except bakery) manufacturing
- Fruit pops, frozen manufacturing
- Ice cream manufacturing
- Ice cream specialties manufacturing
- Ices flavored sherbets manufacturing
- Juice pops frozen manufacturing
- Pudding pops frozen manufacturing
- Sherbets manufacturing
- Tofu frozen desserts manufacturing
- Yogurt frozen manufacturing

## Appendix 5 Derivation of the maximum entropy solution

$$\max_a H(a) = -\sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij} \ln a_{ij}$$

subject to :

$$\sum_{j=1}^{K} a_{ij} N_j^1 = N_i^1 \qquad \forall i$$

$$\sum_{i=1}^{K} a_{ij} = 1 \qquad \forall j$$

$$L(a_{ij}, \lambda_i, \mu) = -\sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij} \ln a_{ij} + \lambda_i\left(N_i^1 - \sum_{j=1}^{K} a_{ij} N_j^1\right) + \mu\left(1 - \sum_{i=1}^{K} a_{ij}\right)$$

$$\frac{\partial L}{\partial a_{ij}} = -\left(\ln a_{ij} + a_{ij}\frac{1}{a_{ij}}\right) - \lambda_i N_j^1 - \mu = 0$$

$$\frac{\partial L}{\partial \lambda_i} = N_i^1 - \sum_{j=1}^{K} a_{ij} N_j^1 = 0$$

$$\frac{\partial L}{\partial \mu} = 1 - \sum_{i=1}^{K} a_{ij} = 0$$

$$\ln a_{ij} = -1 - \lambda_i N_j^1 - \mu$$

$$a_{ij} = \exp\left(-1 - \lambda_i N_j^1 - \mu\right)$$

$$1 = \sum_{i=1}^{K} \exp\left(-1 - \lambda_i N_j^1 - \mu\right)$$

$$1 = \sum_{i=1}^{K} \exp(-1-\mu)\left(-\lambda_i N_j^1\right)$$

$$\exp(-1-\mu) = \frac{1}{\sum_{i=1}^{K} \exp\left(-\lambda_i N_j^1\right)}$$

Now remains to plug the right hand side formula into the expression for $a_{ij}$ given on the left hand side and we obtain the solution for $a_{ij}$

$$a_{ij} = \frac{\exp\left(-\lambda_i N_j^1\right)}{\sum_{i=1}^{K} \exp\left(-\lambda_i N_j^1\right)}$$

# Appendix 6

Make and use tables are symmetric with an equal number of products and industries. Adopting the notation from Table 3.1 $V = v_{ji}$ is a *make matrix* of dimension industry *j* by commodity *i*. Entries in each cell of the *make matrix* show the monetary value of each commodity produced by each industry. The production of the primary commodities is reported in the cell on the diagonals, leaving off-diagonals for secondary products in the particular industry.[39] Taken together the row entries show the product mix of the industry. The entries in the columns show the different types of industries that produce the particular commodity regardless whether it is a primary or secondary product. The row totals indicate the total value of the industry output, while column totals reflect the commodity output.

The *use matrix* consists of two parts. The first part of the *use matrix* $U = u_{ij}$ of dimension commodity *i* by industry *j* denotes the monetary value of each commodity used by each industry, i.e., the absorption of the given commodities by the particular industry in producer prices. The second part provides the value added components used by each industry. The US 1992 tables provide the data on the components of the value added in a highly aggregated manner. The information is provided only on the compensation of employees, other value added and indirect tax revenues. The final demand denoted by *E* reports the allocation of the commodities within its main components such as private household consumption, public consumption and investments. Q stands for the total demand within the economy.

There are four established constructs of the symmetric IO table in the literature i) the commodity technology model, ii) by-product technology model, iii) industry technology model, and iv) and a combination of commodity and technology models, the so-called mixed technology model developed by Gigantes (1970). Here we outline only the two major methods applied in our study, i.e., the commodity and industry technology models, while referring readers to the relevant literature dealing with the related technical issues.

The *commodity technology model* as proposed by the United Nations (1967) assumes that the input structure in the production of a particular commodity is fixed regardless in which industry it is being produced. Thus $u_{ij}$ can be written as follows:

[39] Each classification provides an extensive set of definitions and conventions for the classification of industries.

2.1.3 $$u_{ij} = \sum_{k=1}^{n} a_{ik} v_{jk} \qquad i, j = 1,...,n$$

Note that direct requirements per unit of output of commodity k still remains $a_{ik}$ regardless of whether the commodity is produced by the industry *j* or *t*. The analytical expression for the matrix of the $a_{ik}$ coefficients can be obtained by the reformulation of the initial IO table identities. The formulas can be derived straightforwardly from the initial identities by replacing the use matrix by the formula implied by the initial assumption. Denote final demand by FD, q and g the commodity and activity totals respectively.

2.1.4. a/

$$q = Ui + FD$$
$$q = Bg + FD$$
$$q = BC^{-1}q + FD$$
$$q = (I - BC^{-1})^{-1} FD$$

2.1.4. b/

$$q = Ui + FD$$
$$Cg = Bg + FD$$
$$g = C^{-1}Bq + C^{-1}FD$$
$$g = (I - C^{-1}B)^{-1} C^{-1}FD$$

To restore the initial equilibrium (i.e. the equality between the row and column totals of the SAM) the value added sub-matrix requires additional adjustment. It turns out that the initial sub-matrix is multiplied by the inverse of the C matrix and the ratio of make matrix totals and use matrix totals. In a similar vein further adjustments are necessary also for the matrices of the dimension industry by industry. The final demand and tax sub-matrices are obtained by multiplying their original values by the C inverse. The commodity technology assumption has earned rather wide ranging support in particular due to its sound underlying economics. It is the only construct of the input-output coefficients which also fulfils all desirable analytical properties of the input-output coefficients established by the Jansen and ten Raa (1990) axioms.[40] Equation 2.1.4 indicates however the restrictiveness of this assumption. Both formulas require non-singularity of the C matrix to secure a well defined own inverse. This implies that the commodity technology model cannot be applied when the underlying make and use matrices are rectangular. Its major drawback relates to its numerical performance as this model tends to generate meaningless negative input-output coefficients of low magnitude.[41] The existence of the negative figures is commonly explained by the over-specification of the secondary products produced by

[40] The desirability axioms require the following equations to hold in the input-output system i) material balance, ii) financial balance, iii) invariance of the $a_{ij}$ matrix with respect to the unit of measurement, iv) scale invariance.

[41] In our calculations the negative values are in some cases of a relatively higher magnitude.

the particular industry (*UN, 1999)*. Consequently additional methods have to be employed to obtain a square matrix of positive coefficients[42] (*ten Raa and van der Ploeg, 2002*). The entire table is presented, however for this research only the IO coefficients of the dairy sectors are relevant. The cells with negative values were simply set equal to zero and the generated IO has been rebalanced by a cross entropy program.[43]

The *industry technology model* proposed by the United Nations (1967) is based on the assumption that each industry *j* has the same input requirements for any unit of output. This implies that each industry produces commodities by its own specific technology. Furthermore the assumptions imply i) the input structure of the industries is proportional to their output, ii) market shares are held fixed and independent on any level of output.

Before the actual derivation we define the following relationships :

2.1.5
$$U = Bg$$
$$g = Mi = Dq$$

We derive formulas for both possible dimensions commodity by commodity and industry by industry. In both cases we start with the underlying IO accounting balance. Expression 2.1.6 gives the relationships for the dimension commodity by commodity. By replacing the use matrix through the support matrices B and D as implied by the initial assumptions 2.1.5 the final formula is obtained. In the case of industry by industry a multiplication of the initial IO identity leads straight to the final formula.

2.1.6.a/

$$q = Ui + FD$$
$$q = Bg + FD$$
$$q = BDq + FD$$
$$q = (I - BD)^{-1} FD$$

2.1.6.b/

$$q = Ui + FD$$
$$q = Bg + FD$$
$$g = DBq + D * FD$$
$$g = (I - DB)^{-1} D * FD$$

The interpretation of the technical coefficients (dimension industry by industry) $a_{ij} = \sum_{k=1}^{n} b_{ik} d_{kj}$ follows from the main assumption of the industry technology model. The technical coefficients $a_{ij}$ denote the amount of input *i* required for one unit of output *j* resulting from a market share weighted average over industries *k*.

[42] The methods involve i) setting negative values equal to zero and using the RAS procedure to balance the table, ii) minimization of variances under constraints to generate the positive values, although this can be questioned due to the economic justification of the particular objective function.
[43] Treatment of the existing negative values would deserve more attention.

## Appendix 7. US dairy powder breakdown

| | ocr | ctl | ofd | rmk | frs | ene | omt | vol | fmk | but | smp | wmp | whp | chs | lin | hin | man | scs | trd | otp | vpm | vgm | inv | vxm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ocr | | | | | | | | | | | | 0.00 | 0.00 | | | | | | | | | | | |
| ctl | | | | | | | | | | | | 0.00 | | | | | | | | | | | | |
| ofd | | | | | | | | | | | 0.01 | 0.05 | 0.04 | | | | | | | | | | | |
| rmk | | | | | | | | | | | 0.32 | 0.10 | 0.16 | | | | | | | | | | | |
| frs | | | | | | | | | | | | 0.00 | | | | | | | | | | | | |
| ene | | | | | | | | | | | | 0.00 | 0.00 | | | | | | | | | | | |
| omt | | | | | | | | | | | | 0.00 | | | | | | | | | | | | |
| vol | | | | | | | | | | | 0.01 | 0.04 | 0.02 | | | | | | | | | | | |
| fmk | | | | | | | | | | | 0.02 | 0.05 | 0.04 | | | | | | | | | | | |
| but | | | | | | | | | | | 0.00 | 0.03 | 0.01 | | | | | | | | | | | |
| smp | 0.33 | 0.33 | 0.37 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.41 | 0.34 | 0.03 | 0.06 | 0.05 | 0.40 | 0.33 | 0.34 | 0.34 | 0.36 | 0.46 | 0.35 | 0.75 | 0.34 | 0.33 | 0.42 |
| wmp | 0.33 | 0.33 | 0.30 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.26 | 0.33 | 0.00 | 0.02 | 0.01 | 0.27 | 0.33 | 0.33 | 0.33 | 0.31 | 0.22 | 0.31 | 0.05 | 0.32 | 0.33 | 0.25 |
| whp | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.01 | 0.04 | 0.03 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.32 | 0.33 | 0.21 | 0.33 | 0.33 | 0.33 |
| chs | | | | | | | | | | | 0.02 | 0.05 | 0.04 | | | | | | | | | | | |
| lin | | | | | | | | | | | 0.01 | 0.04 | 0.03 | | | | | | | | | | | |
| hin | | | | | | | | | | | 0.02 | 0.05 | 0.04 | | | | | | | | | | | |
| man | | | | | | | | | | | 0.01 | 0.04 | 0.03 | | | | | | | | | | | |
| scs | | | | | | | | | | | 0.09 | 0.07 | 0.09 | | | | | | | | | | | |
| trd | | | | | | | | | | | 0.05 | 0.07 | 0.07 | | | | | | | | | | | |
| otp | | | | | | | | | | | 0.01 | 0.04 | 0.03 | | | | | | | | | | | |
| tax | | | | | | | | | | | 0.00 | 0.02 | 0.01 | | | | | | | | | | | |
| lab | | | | | | | | | | | 0.08 | 0.07 | 0.08 | | | | | | | | | | | |
| ova | | | | | | | | | | | 0.28 | 0.10 | 0.15 | | | | | | | | | | | |
| vimd | | | | | | | | | | | 0.05 | 0.06 | 0.06 | | | | | | | | | | | |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

## Appendix 8. Derivation of the cross entropy solution

$$\min_{\beta} \sum_{j=1}^{K} w_j \sum_{i=1}^{K} \beta_{ij} \ln\left(\frac{\beta_{ij}}{\beta_{ij}^0}\right)$$

subject to

$$\beta_{ij} * y_j = x_i$$
$$\sum_i \beta_{ij} = 1$$
$$0 \le \beta_{ij} \le 1$$

$$L(\beta_{ij}, \lambda_i, \mu) = -\sum_{j=1}^{K} w_j \sum_{i=1}^{K} \beta_{ij} \ln\frac{\beta_{ij}}{\beta_{ij}^0} + \lambda_i\left(N_i^1 - \sum_{j=1}^{K} \beta_{ij} N_j^1\right) + \mu\left(1 - \sum_{i=1}^{K} \beta_{ij}\right)$$

$$\frac{\partial L}{\partial \beta_{ij}} = \mathrm{w}_{\mathrm{j}} - \left(\ln\frac{\beta_{ij}}{\beta_{ij}^0} + \frac{\beta_{ij}}{\beta_{ij}^0}\frac{\beta_{ij}^0}{\beta_{ij}}\right) - \lambda_i N_j^1 - \mu = 0$$

$$\frac{\partial L}{\partial \lambda_i} = N_i^1 - \sum_{j=1}^{K} \beta_{ij} N_j^1 = 0$$

$$\frac{\partial L}{\partial \mu} = 1 - \sum_{i=1}^{K} \beta_{ij} = 0$$

$$\ln\frac{\beta_{ij}}{\beta_{ij}^0} = -1 - \lambda_i N_j^1 - \mu - w_j$$

$$\beta_{ij} = \beta_{ij}^0 \exp\left(-1 - \lambda_i N_j^1 - \mu - w_j\right)$$

$$1 = \sum_{i=1}^{K} \beta_{ij}^0 \exp\left(-1 - \lambda_i N_j^1 - \mu - w_j\right)$$

$$1 = \sum_{i=1}^{K} \exp\left(-1 - \mu - w_j\right)\beta_{ij}^0\left(-\lambda_i N_j^1\right)$$

$$\exp\left(-1 - \mu - w_j\right) = \frac{1}{\sum_{i=1}^{K} \beta_{ij}^0 \exp\left(-\lambda_i N_j^1\right)}$$

Now remains to plug the right hand side formula into the expression for $\beta_{ij}$ given on the left hand side and we obtain the solution for $\beta_{ij}$

$$\beta_{ij} = \frac{\beta_{ij}^0 \exp\left(-\lambda_i N_j^1\right)}{\sum_{i=1}^{K} \beta_{ij}^0 \exp\left(-\lambda_i N_j^1\right)}$$

# Appendix 9 Concordance with COMTRADE database

To be added

## Appendix 10. List of tables

a) US SAM 1992 taken as prior information for the cross entropy programs.
b) Bilateral trade matrix extracted from the COMTRADE database.
c) Initial value cost shares collected from FAO and EUROSTAT.
d) Collected information on prices of dairy commodities.
e) Regional SAM estimates.
f) Estimate of the matrix of bilateral trade.
g) Estimate of the matrix of margins on a bilateral basis.

## Appendix 11. Goodness of fit indicators

Our aim in this study was to obtain the reasonably "closest" possible estimates of the cost shares of dairy commodities relative to the prior information from the US 1992 IO tables. On space grounds and to aid in the comprehension of the exposition we limit ourselves to an assessment of the results obtained for Ireland. The following indicators were used to assess the robustness of the estimates.

i) correlation coefficients

ii) mean absolute deviation $$MAD = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left|\hat{a}_{ij} - a_{ij}\right|$$

iii) squared error measure $$SEM = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left(\hat{a}_{ij} - a_{ij}\right)^2$$

iv) maximum proportionate error $$MaxPE = \max_{i,j}\frac{\left|\hat{a}_{ij} - a_{ij}\right|}{a_{ij}}$$

v) mean absolute proportionate error $$MAPE = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{\left|\hat{a}_{ij} - a_{ij}\right|}{a_{ij}}$$

vi) goodness of fit $$GOF = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\left|\frac{\left(\hat{a}_{ij} - a_{ij}\right)^2}{a_{ij}}\right|$$

## Appendix 12. Comparison of the cross-entropy and Splitcom splitting weight estimates

To be completed