Spatial-economic-ecological model for the assessment of sustainability policies of the Russian Federation

Project 213091

D2.1
Description of the constructed database, data quality and data collection methods

Contract No.	SUST-RUS 213091
Workpackage	WP2 – Description of the constructed database, data quality and data collection methods
Date of delivery	M15
Actual Date of Delivery	M16
Dissemination level	Public
Responsible	CEFIR
Authors	Natalia Tourdyeva, CEFIR Christophe Heyndrickx, TML
Status of the Document	Draft
Version	1.0

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement No. 213091.

Table of contents

Tables

Figures

1. Database for the SUST-RUS project

The SUST-RUS project is aimed on providing sustainable development policy assessment for the Russian Federation. The proposed assessment method is based on the use of the SUST-RUS applied general equilibrium model.

An applied general equilibrium (AGE) model is a structured mathematical representation of an economy in question. Structure of an AGE model is replicated in a benchmark dataset, which should include all economic agents specified in the model as well as satisfy model's balancing constraints.

Usually the benchmark dataset is organized in a form of a social accounting matrix (SAM). The SAM format ensures that all material and financial flows are balanced. This report focuses on construction of the benchmark dataset for the SUST-RUS project in the social accounting matrix format. Given spatial nature of the SUST-RUS project, the benchmark dataset is a multiregional social accounting matrix, where each regional SAM represents economy of a federal district of the Russian Federation. All regional SAMs (RSAMs) are interconnected by trade and income flows. All RSAMs sum up to the country social accounting matrix and all RSAMs have structure implied by the SUST-RUS model. The creation of system of RSAMs is the task of the WP2 of the SUST-RUS project and the subject of this report.

2. Social Accounting Matrix Format

A social accounting matrix (SAM) is "… a square matrix in which each account is represented by a row and a column. Each cell shows the payment from the account of its column to the account of its row. Thus, the incomes of an account appear along its row and its expenditures along its column. The underlying principle of double-entry accounting requires that, for each account in the SAM, total revenue (row total) equals total expenditure (column total)." (Lofgren et al., 2002).

The SUST-RUS social accounting matrix aggregated across regions is presented in the Appendix A. (Table A.1).

Table 1. Structure of the SUST-RUS SAM

Receipts	Expenditures								
	Commodities	**Activities**	**Factors**	**Taxes**	**Households**	**Government**	**Investment**	**ROW**	**Totals**
Commodities	-	Intermediate inputs	-	-	Private consumption	Government Consumption	Investment	Exports	Demand
Activities	Maketed outputs	-			-	-	-	-	Activity Income
Factors	-	Value-added	-	-	-	-	-	-	Factor Income
Taxes	-	Producer & value-added taxes	-	-	Income & Commodity Taxes	Commodity Taxes	Commodity Taxes	-	Taxes Collected
Households	-	-	Factor income to households	-	-	-	-	-	Household Income
Government	-	-	-	Taxes to Government Budget	-	-	-	-	Governemnt Income
Savings	-	-	-	-	Private Savings	Public Savings	-	Foreign Savings	Savings
ROW	Imports	-	-	-	-	-	Investments in ROW	-	Foreign exchange outflow
Totals	Supply expenditures	Activity	Factor expenditures	Total Taxes	Household Expenditures	Government Expenditures	Investment	Foreign exchange inflow	

3. Source data

3.1 Russian input-output tables

There are a number of input-output tables at our disposal. Unfortunately, there is no published input-output table which we could use as a basis of the SUST-RUS database. Thus, we collected several tables for estimations. Among these tables are:

The 1995 Russian symmetric input-output table (SIOT). This table consists of 110 sectoral groups. Classification of the sectoral groups is based on the old Soviet industrial classification system called OKONH (Общероссийский классификатор отраслей народного хозяйства, All-Russia Industrial Classification, (Obsherossiiskii klassifikator otraslei narodnogo hozaistva, OKONH[1])), which was official Russian statistical classification system until 2004.

The System of Russian input-output tables for 2003. This is an official Rosstat publication[2], also based on OKONH format. The 2003 SIOT consists of 22 industial and services groups. The source table is a symmetric input-output table in commodity-by-commodity format for 2003. It represents 22 "single-product" producing sectors[3]; data is measured in thousands of Russian rubles.

The symmetric input-output table is accompanied by non-symmetric supply and use tables, tables of domestic and imported product use, tables of transport and trade mark-ups, and a tax table. All these tables include 24 producing sectors[4] and commodity groups aggregated according to Russian national industrial classification on a commodity-by-industry basis.

The make table and the use table are in consumer prices for year 2006[5]. Both tables consist of 15 industrial groups in the new Russian classification OKVED. Input-output tables for 2006 are in a highly aggregated (1-letter level)[6] format.

3.1.1 System of National Accounts 2006

The Rosstat publication "System of National Accounts 2001-2008" is the source of detailed information in UN SNA 93 format. There are data on different economic agents' accounts as well as integrated tables and use tables for several years. All information is presented in the OKVED format. The integrated table of the System of National Accounts for 2006 is a valuable source of information on income flows between economic agents. This information is essential for balancing the social accounting matrix.

[1] OKONH classification (http://www.standard.ru/classif/okonh/okonh.phtml) was the official industrial classification in the Soviet Union and Russia until recently (1976-2004). In 2004 Russia adopted a new classification OKVED based on "Statistical Classification of Economic Activities in the European Community" (NACE Rev. 2).
[2] All Rosstat publications are available online. Latest edition of Russian input-output tables covers year 2003; it was published in 2006. «Система таблиц «Затраты-выпуск» России за 2003 год», Росстат: Москва, 2006 г. (http://www.gks.ru/doc_2006/Zatrat06.zip).
[3] Description of IO tables methodology is published in Rosstat (1998), Chapter 5 "Input-Output Tables".
[4] Some differences in methodology should be noted; for instance, Rosstat does not calculate imputed rent for owner-occupied dwellings. "The value of housing services is treated as a sum of current expenditure of dwelling and consumption of fixed capital" (Masakova, 1998).
[5] «Национальные счета России 2001-2008 гг.», Росстат: Москва, 2009г. National Accounts of Russia, Tables 5.1 (make matrix) and 5.2 (use matrix in consumer prices) (http://www.gks.ru/doc_2009/nac_sh.zip).
[6] OKVED and NACE are very similar. The 2006 input-output tables consist of 15 industrial groups: A, B, C, D, E, F, G, H, I, K, L, M, N, O (http://www.fifoost.org/database/nace/nace-en_2002AB.php).

3.1.2 Interregional trade database

One of the main components of the raw data is the Russian interregional trade database. This database consists of information on exports of goods of Russian regions since 1999 to 2006. There is a list of 245 commodity groups that are monitored in the data. Total number of regions in the data is 89, which corresponds to the number of the subjects of the Russian Federation in 2006. Export destinations that are recorded in the database for each region consist of other Russian regions as well as CIS countries and the rest of the world. There is information on domestic regional consumption of each commodity group as well. Commodity groups in interregional trade database and their correspondence to NACE groups are presented in *Table 2*.

Table 2. Numbers of commodity groups in the Russian interregional trade database corresponding to selected NACE (rev 1) classification industrial codes

NACE classification	Number of commodity groups
A	8
CA_col	1
CA_gas	1
CA_oil	1
CB	5
DA	42
DB	22
DC	9
DD	8
DE	7
DF	9
DG	21
DH	6
DI	9
DJ	32
DK	28
DL	15
DM	14
DN	7
Grand Total	**245**

There is a detailed description of the preparatory data management that was involved before the actual use of the Russian regional database in the construction of the SUST-RUS database in the Appendix B.

We should notice that there is no information on trade in services between Russian regions. The problem of treating interregional trade in services between Russian regions would be discussed later on, at the stage of estimating regional input-output matrices.

3.1.3 Social taxes in the structure of the value added in 2006

Rosstat publication "Labour in Russia, 2008" contains information for estimation of the social taxes that are integrated in the value added.

3.1.4 International trade 2006

Data on international trade of the Russian regions was obtained from CEFIR's international trade database derived from original Federal Customs database.

3.1.5 Regional output statistics

Data on Russian regional output statistics is rather scarce. Unfortunately, at the date of SUST-RUS database construction there is no statistical information on Russian regional production across all categories of the 2-digit NACE codes. The origin of the regional output data is the Rosstat online database (http://www.gks.ru/dbscripts/Cbsd/DBInet.cgi).

Available information on the value of output consists of regional production of industries that belong to C, D and E categories. There is no data on value of regional output for categories A, B, F, G, H, I, K, L, M, N, and O, only some information on the volume of production in natural terms. The lack of reliable data introduces data risks in the SUST-RUS database. A number of assumptions on the structure of the value of regional output in 2006 were made.

4. Construction of the Russian regional social accounting matrices

4.1 Choice of the industrial classification in the SUST-RUS database

The choice of the industrial classification turns out to be quite a challenging task. There is an obvious trade-off between data constraints and a desire to be as detailed as possible in terms of the industries modeled. The detailed industrial structure is a very valuable feature of the model given environmental and climate policy applications anticipated as a major use of the SUST-RUS model.

Among the most binding constraints we can name the following:

Availability of regional output data. There is a lack of data on regional output.

High level of aggregation of the latest (2004-2006) Russian use and make matrices. Published versions consist only of 15 NACE categories corresponding to one-letter NACE code. For example, all manufacturing industries are aggregated into single code "D".

Differences in the statistical classifications between recent (2004-2006) (aggregated Russian IO tables) and older more disaggregated (1995-2003) Russian input-output tables. There was a change in the Russian statistical classification in 2004. An old OKONH classification does not match the new one, based on the standard international NACE classification. Industry matching is possible only on a highly disaggregated level of industrial structure. There is no one-to-one mapping on the aggregated level from OKONH to standard international classification like ISIC or NACE. We had to disaggregate the source table in order to match NACE sectoral classification. At this stage we need a more detailed version of input-output tables, like 1995 Russian symmetric input-output table with 110 industries. This level of details permitted us to build a one-to-one mapping to NACE.

The core of the SUST-RUS database is the symmetric input-output matrix. Thus, in order to create a meaningfully disaggregated database, we have to use input-output estimation techniques, entropy minimization technique and the RAS method. These techniques are based on the past information (priors) on matrix structure. The estimations methods impose structure of the prior data on the estimated matrix. Given the aggregated nature of the recent data the only prior that we could use in order to disaggregate the input-output matrix comes from the 1995 Russian input-output data. But this technique bears an unavoidable risk of imposing a 15-years old industrial structure on a transition economy with rapidly changing economic environment. Thus, our ability to use the prior was constrained by expert opinions of the level of possible disaggregation.

As the result of the described trade-off the following industrial structure for the database was chosen.

Table 3. The industrial structure of the SUST-RUS database

#	Code in SUST-RUS	NACE classification	Description
1	A	Section A	Agriculture, hunting and forestry
2	B	Section B	Fishing
3	CA_col	CA.10	Mining of coal and lignite; extraction of peat
4	CA_gas	CA.11.10.2-3	Extraction of natural gas; service activities incidental to gas extraction, excluding surveying
5	CA_oil	CA.11.10.1	Extraction of crude petroleum; service activities incidental to oil extraction, excluding surveying
6	CB	Subsection CB	Mining and quarrying, except of energy producing materials
7	DA	Subsection DA	Manufacture of food products, beverages and tobacco
8	DB	Subsection DB	Manufacture of textiles and textile products
9	DC	Subsection DC	Manufacture of leather and leather products
10	DD	Subsection DD	Manufacture of wood and wood products
11	DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing
12	DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel
13	DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres
14	DH	Subsection DH	Manufacture of rubber and plastic products
15	DI	Subsection DI	Manufacture of other non-metallic mineral products
16	DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products
17	DK	Subsection DK	Manufacture of machinery and equipment n.e.c.

18	DL	Subsection DL	Manufacture of electrical and optical equipment
19	DM	Subsection DM	Manufacture of transport equipment
20	DN	Subsection DN	Manufacturing n.e.c.
21	E_distr	41 + 40.2 + 40.3	Collection, purification and distribution of water; Manufacture of gas; distribution of gaseous fuels through mains; Steam and hot water supply
22	E_ely	40.1	Production and distribution of electricity
23	F	Section F	Construction
24	G	Section G	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
25	H	Section H	Hotels and restaurants
26	I_cmn	64	Post and telecommunications
27	I_trn	60 + 61 + 62 + 63	Land transport; transport via pipelines; Water transport; Air transport; Supporting and auxiliary transport activities; activities of travel agencies
28	J	Section J	Financial intermediation
29	LO	Section L and O	Public administration and defence; compulsory social security; Other community, social and personal service activities
30	K	Section K	Real estate, renting and business activities
31	M	Section M	Education
32	N	Section N	Health and social work

There is a lack of information on Subsections P and Q, as well as on Subsection CA. 12: Mining of uranium and thorium ores.

4.1.1 Choice of the base year for the SUST-RUS database

The choice of the base year is closely linked to the choice of the industrial structure. The tradeoff here lies on availability of data vs. the choice of the most recent year to base the SUST-RUS dataset on.

Arguments for and against different base years are listed below:

The 2003 year was a good candidate as a base year for the SUST-RUS dataset. There is a fairly detailed system of input-output matrices, regional output statistics as well as national accounts data. But all these statistics are in the old Soviet industrial classification which does not match current Russian or international statistics. Given the policy-oriented nature of the SUST-RUS project, the obsolete industrial classification would hinder presentation of our modeling results to the intended audience. Thus we make a decision that SUST-RUS should be based on the current classification system, but in the estimation process we would use all possible information, including information on 2003 input-output matrices.

New classification (OKVED – a NACE-based system) was introduced in 2004. There are no detailed input-output matrices in OKVED for Russia. But there is a number of aggregated (15 sectors, 1-letter NACE classification) use and make matrices for years 2004-2006. Thus, we could use any of these years as a base year for the SUST-RUS database.

At the time of decision making, the year richest in terms of the available regional data was 2006. The Rosstat on-line database included almost all manufacturing data for Russian regions for year 2006. Thus, the decision was made to use this year as the base year for the SUST-RUS database.

4.1.2 Overview of the proposed method

There are several major steps in the construction of the Russian regional social accounting matrices for the SUST-RUS project:

Estimation of the Russian country symmetric input-output table (SIOT) for 2003 in NACE format.

As we noted earlier, there is a difference in statistical classifications between input-output tables published in years 1995-2003 and the present statistical format. These input-output tables have a good representation of the manufacturing sector – 14 of total 23 industries/commodity groups are of manufacturing origin. The latest available input-output table with disaggregated manufacturing sector is IO table for 2003. Thus, to construct the social accounting matrices for the SUST-RUS database, we have to estimate the symmetric input-output table in OKVED (NACE) format.

The estimation of the 2003 symmetric input-output table is possible due to the existence of the 1995 symmetric input-output table with 110 industries. The estimation technique used for this purpose is quite close to the entropy minimization method, where we use the 1995 table as a prior.

Estimation of the Russian SIOT for 2006 in disaggregated NACE classification.

The use of the 1995 year table as a prior imposes the old structure on the 2003 data. In order to minimize adverse effects of an old prior, and in order to use the most recent data, we calculate a disaggregated symmetric input-output table for 2006. The estimation procedure is close to the cross-entropy minimization as well, but we use the 2003 table as a prior this time.

Estimation of the Russian regional symmetric input-output matrices in disaggregated NACE classification.

Due to the lack of regional input-output tables we used a number of assumptions in order to estimate regionally disaggregated input-output tables. We assumed that technological coefficients in each region are the same and are identical to the country coefficients.

We assumed that final consumption on the regional level has the same structure as on the country level. We also assumed that ratio of the intermediate consumption to final consumption in each region is the same and corresponds to the ratio in the country table.

Due to the lack of the interregional data on trade in services, we assumed that there was no trade in services among regions, only international trade for each region.

We assumed that there was no re-export on the regional level in the base year. This assumption led to changes in the structure of interregional trade. Thus, in order to meet this constraint we estimated interregional trade with cross-entropy method, using actual data from interregional trade as a prior.

Estimation of the Russian regional social accounting matrices in disaggregated NACE classification.

Data in the format of the input-output table is almost the same as in the social accounting matrix format. The only difference is that the social accounting matrix incorporates information on income flows between economic agents. Thus, in order to go from input-output table to the SAM, we need national accounts data.

Each step of the proposed methodology includes several tasks as displayed in the diagrams below.

Figure 1. Estimation of the Russian SIOT for 2003 in NACE format

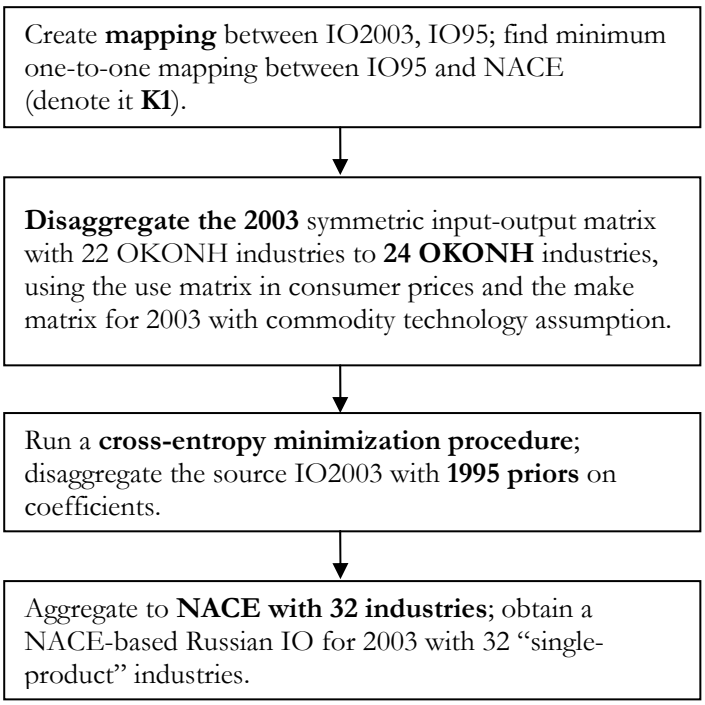


Figure 2. Estimation of the Russian SIOT for 2006 in disaggregated NACE classification

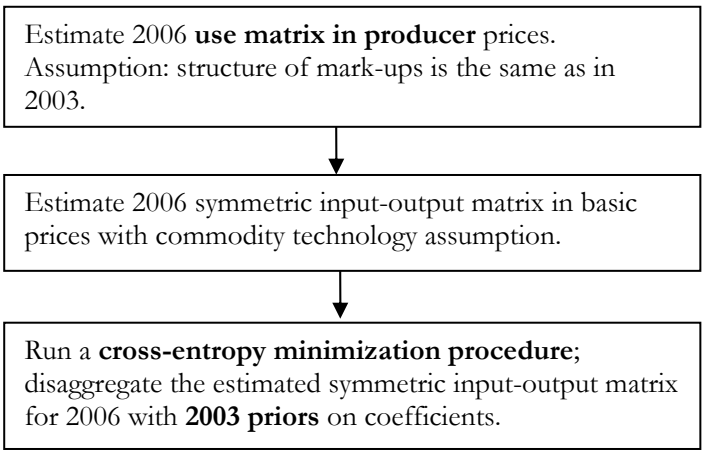
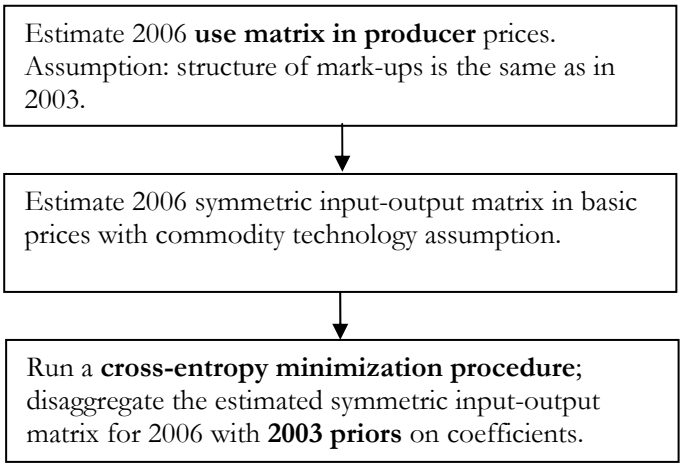


Figure 3. Estimation of the Russian regional symmetric input-output matrices in disaggregated NACE classification



4.1.3 Cross-entropy minimization technique

We used an entropy minimization technique, similar to Robinson, Cattaneo and El-Said (2001) for the disaggregation of the Russian IO for 2003. The detailed description of this numerical exercise is in the appendix C.

First we used cross-entropy minimization method in order to disaggregate 23 sectors of the symmetric input-output table to 59 sectors[7]. This operation was possible since 2003 table is in OKONH classification as well as the earlier more disaggregated IO table which we used as a prior. After obtaining a balanced symmetric input-output table with 59 sectors for 2003, we aggregated it to NACE format. This is possible since on the disaggregated level (2-digit NACE) there is a good concordance between OKONH and OKVED.

The same method of estimating Russian input-output table was used in Tourdyeva and Shkrebela (2006) for estimation of the IO table in the GTAP format.

5. Disaggregation of trade flow data and social accounting matrices

5.1 Introduction

Regions are linked among themselves through trade in goods and services, factor and income flows. Regions import and export goods and services from other regions in the same country and from other countries. In contrast to international trade, usually, trade of goods and services between regions is not recorded statistically. As a result, no good statistical data exists for inter-regional trade, which could be readily used for modelling purposes.

[7] Russian I-O tables report data in OKONH classification. In order to find a correspondence between I-O data and GTAP sectors we build a mapping from I-O sectors to OKONH, then from OKONH to ISIC. We base our classification on a mapping between OKONH and OKVED classifications published by the Ministry of Economy of the RF and Rosstat in 2002 (http://okpd.org/product/okonh_okved.zip). The minimum common classification contains 59 sectors.

Such data, however, is necessary for empirical implementation of multi regional general equilibrium models, such as SUST-RUS. The objective of this note is to discuss data and methodological issues of modelling the inter-regional trade flows and, based on the gained insights, to develop a strategy for implementing inter-regional trade in SUST-RUS.

In general equilibrium models of trade, the trade in goods and services is determined endogenously as an optimization outcome of the optimal strategy of firms, households and other economic 'actors'. Hence, solving the model for the equilibrium set of prices and quantities yields also inter-regional trade flows, which arise from the underlying conceptual framework and the estimated parameters. For the model's base year, the predicted trade flows need to satisfy also exogenous data constraints. This can be done via the cross-entropy minimization method. This method allows obtaining the inter-regional trade flows that, on the one hand, satisfy the imposed data constraints and, on the other hand, minimize the distance to the initial (unconstrained) estimates of trade flows.

In the context of our study the minimum entropy approach has several advantages. First, the cross-entropy approach is flexible in the sense that it takes into account different data sources for inter-regional trade flows and restrictions on the data, e.g. the production and consumption accounts in regions. Second, the minimum entropy approach utilizes nonlinear programming technique and, hence, can handle non-linear problems. Third, it can be implemented in GAMS, which is the language used in our model.

6. Creating and optimizing a regional database

When a database at regional level is constructed, data from several sources has to be combined and balanced. In general, to construct a regional general equilibrium model, at least the following data are necessary.

- A balanced social accounting matrix at national level
- Input-output coefficients at national level, based on a balanced input-output table
- International imports and exports at regional level
- Regional production data (outputs)
- Regional consumption data
- Information on regional input-output coefficients (in case these are unavailable, national estimates will be projected at regional level)
- Data on interregional trade flows and transport costs between regions

The preferred method in regional economic modelling is top-down disaggregation of a national social accounting matrix at regional level, based on regional output data. However, to account for heterogeneity among regions, it requires a consistent and robust method to rebalance data, based on the constraints of the system. A scheme of the procedure is presented in Figure 4.

To create a mature database at the regional level, several constraints must be met. We sum these up mathematically, using the notations from Table 4 and Table 5.

Table 4: Notation used

Description	Subscript
Sectors/products (each sector produces only one product)	*i*
Intermediate inputs (products ii, sectors i)	*ii,i*
Regions (Federal regions of Russia)	*r*
Rest of the world region	*RoW*
Flows of goods, labour and capital (from region r to region rr)	*r,rr*
Index used in social accounting matrix	*b*
Superscript 0 is used to indicate the initial (previous period) level of variable	*0*

Table 5: Variables in optimization procedure

Description	Notation
Social accounting matrix on national level	$SAM_{b,bb}$
Social accounting matrix on regional level	$SAMR_{b,bb,r}$
Input output (use) table	$IO_{i,ii,r}$
Production	$XD_{i,r}$
Sales	$X_{i,r}$
Final consumption	$XF_{i,r}$
Value added	$VA_{i,r}$
Taxes on production	$TAX_{i,r}$
Imports	$M_{i,r}$
Exports	$E_{i,r}$
Interregional trade flows (by good)	$XDDE_{i,r,rr}$
Interregional margin on trade (by good)	$trm_{i,r,rr}$

The first, most obvious constraint is that the social accounting matrices at regional levels have to sum up to the national one. While this may seem trivial, this is the basis of the top-down disaggregation procedure. Posing this constraint is actually much more restrictive than might initially be expected. Recall that a social accounting matrix contains all data on production, intermediate consumption, final consumption, exports & imports, production factors, etc. Applying this constraint, we assume that the social accounting matrix at national level is correct. As such, the data we have at regional levels have to be balanced (calibrated) at the national level social accounting matrix.

$\sum_r SAM_{b,bb,r} = SAM_{b,bb}$	(1)

Thus, by setting this first constraint, we have already assumed that all our regional values will exactly sum up to the national level data.

$\sum_r XD_{i,r}$ = Production on national level	(2)
$\sum_r X_{i,r}$ = Sales on national level	(3)
$\sum_r IO_{i,r}$ = Intermediate consumption on national level	(4)
$\sum_r VA_{i,r}$ = Value added on national level	(5)
$\sum_r M_{i,r}$ = Imports on national level	(6)
$\sum_r E_{i,r}$ = Exports on national level	(7)

Another important constraint is the one put at the interregional trade flows. This may also seem trivial, but is essential in estimating the regional database. The equation below actually states that the sum of the interregional trade balances, at country level have to sum up to zero. The parameter of trade margins (trm) is equal to the margins on trade, which is the mark up on the value of the traded good. In the model, this parameter is determined exogenously, based on data on transport and trade costs.

$\sum_{i,r} \sum_{rr} XDDE_{i,r,rr} \cdot (1+\mathrm{trm}_{i,r,rr}) - \sum_{rr} XDDE_{i,rr,r} \cdot (1+\mathrm{trm}_{i,rr,r}) = 0$	(8)

Let us reconsider the steps to create a regional social accounting matrix. We assume that we were able to create a balanced national social accounting matrix. The first steps are to disaggregate the national matrix at regional level. The key data here are the output and consumption data from regional account statistics. In general, when the data at regional level is limited, the regional matrix will have several features in common with the national one. If only output and consumption data at regional level are available, one will assume that the value-added shares, input-output coefficients, taxation, etc. in the regional matrix will be proportional to the national matrix. This makes the balancing of the regional social accounting matrix simpler.

The procedure above ensures internal output balance of the regional matrix is correct. This signifies that the production is accounted for by the following equation. The value of production has to be equal to the sum of the value-added, intermediate inputs in production and the net tax on production.

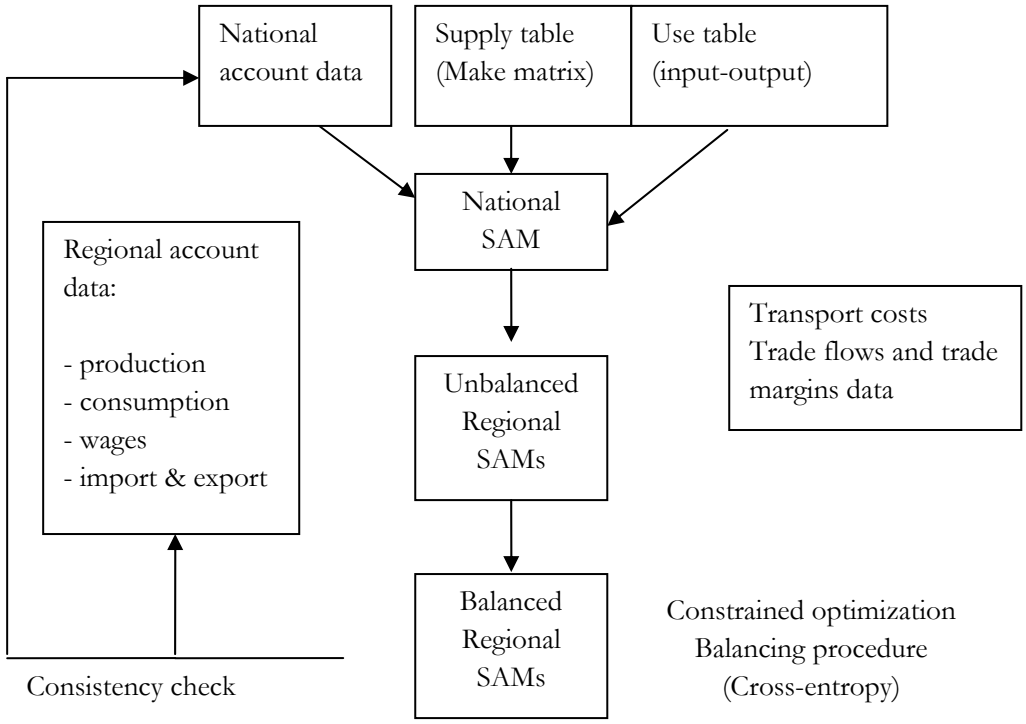
$$XD_{i,r} = VA_{i,r} + \sum_{ii} IO_{i,ii,r} + TAX_{i,r} \quad (9)$$

The next restriction is based on the external output balance. This signifies that all production has a destination, either on the domestic market or on the international market.

$$XD_{i,r} = \sum_{rr} XDDE_{i,r,rr} + E_{i,r} \quad (10)$$

Equations 9 and 10 tell us that production there cannot be more than the value of the inputs of production and that exports cannot be higher than production.

Figure 4. Top-down disaggregation of a balanced national SAM



The next elements are related to the use of goods in each region. The internal balance of sales is defined as the sum of final consumption (including taxes) (XF) and intermediate consumption (IO). Note that the input-output data is now summed up over the rows (sectors) unlike the equation 9, where the inputs are summed up over columns (goods).

$$X_{i,r} = XF_{i,r} + \sum_{ii} IO_{ii,i,r} \quad (11)$$

The external balance of sales is essential in the calibration of the regional dataset. Regional sales are equal to the sum of interregional imports (augmented with trade margins) and international imports.

$$X_{i,r} = \sum_{rr} XDDE_{i,rr,r} \cdot (1 + trm_{i,rr,r}) + M_{i,r} \quad (12)$$

Equation 11 and 12 tell us that there are no more sales than consumption and that goods supplied to a region should be equal to the products sold in that region.

While up till now, these equations may not look particularly special, creating a database which satisfies to equations 1-12 is a difficult and complex task. The main reason for this is related to features of many regional and national datasets. The most troublesome ones are related to the data on interregional and international trade. Often, these data contain re-export and re-import data, meaning that equation 10 and 12 (or the external balances) do not hold. Further complications can arise as international trade flows can be double-counted as interregional ones or get assigned to the wrong regions. To have a consistent database the share of imports to sales and exports to production should thus strictly be less or equal to 1.

$$shareE_{i,r} = \frac{E_{i,r}}{XD_{i,r}} \leq 1 \quad \text{and} \quad shareM_{i,r} = \frac{M_{i,r}}{X_{i,r}} \leq 1 \quad (13)$$

7. Balancing by cross-entropy

The balancing procedures based on Robinson, Cattaneo, and El-Said (2001) and applied to create the SUST-RUS 2003 and 2006 input-output tables, can be modified at regional level. This technique is described in Canning P. and Wang Z. (2005) as a flexible method for creating interregional input-output accounts. Again, the main element is the minimization of entropy (cfr. Introduction). This method has proved powerful in creating balanced regional datasets, based on the main idea that the original proportions in the initial dataset should be (as close as possible) maintained.

The 'workhorse' function of this model is the minimization of entropy, implied by minimizing the Kullback-Liebler Divergence (cfr. Appendix C).

In the model code, the following expression is used to evaluate the distance between the 'new' trade flows and the 'old' trade flows (initial) data, where E is the value of the entropy.

$$\mathrm{E} = \sum_{i} \sum_{r,rr} XDDE'_{i,r,rr} \left(\ln XDDE_{i,r,rr} - \ln XDDE^{0}{}_{i,r,rr} \right) \quad (14)$$

The model could however not be restricted to only optimize the trade flows to the constraints (equation 1-12) we pose on the model. Therefore we added terms to the entropy function, related to the import and exports from each region.

$$\mathrm{E} = \sum_{i} \sum_{r,rr} XDDE'_{i,r,rr} \left(\ln XDDE_{i,r,rr} - \ln XDDE^{0}{}_{i,r,rr} \right) + \sum_{i} \sum_{r} E_{i,r} \left(\ln E_{i,r} - \ln E^{0}{}_{i,r} \right)$$
$$+ \sum_{i} \sum_{r} M_{i,r} \left(\ln M_{i,r} - \ln M^{0}{}_{i,r} \right) \quad (15)$$

This entropy function was minimized, while posing the constraints discussed in the previous paragraph. For this equations 2-12 were integrated in a GAMS calibration model.

Using cross-entropy minimization however, poses some problems. The main difficulty is avoiding negatives in the entropy function during the optimization process. This inevitably leads to a crash of the GAMS program (a negative log is not-defined). To avoid these problems we add a very small, but

positive value to the logarithm (0.000000000001). Also, by definition, trade flows and export/imports should be strictly larger or equal to zero.

An additional problem with cross-entropy is that values which are initially zero, are not handled well by the function. In general it is better to fix these (by definition) to zero and not take these up in the entropy function.

8. An additional complication

As is stated in Deliverable D2.1, data on interregional trade in services was not at our disposal. Therefore we assumed that services are non-tradable between regions. This signifies that equation 10 and 12 should be modified in the following way. The production of services can only be destined to the domestic region or to the international market. The demand for services can only be satisfied by a regional (domestic) production and imports:

$$XD_{i,r} = XDDE_{i,r,r} \Big|_{regionalFlow} + E_{i,r} \qquad (10')$$

$$X_{i,r} = XDDE_{i,r,r} \Big|_{regionalFlow} \cdot (1 + trm_{i,r}) + M_{i,r} \qquad (12')$$

Posing these restrictions also requires a change in the set-up of the calibration procedure. Note that equations 10' and 12' impose only a limited flexibility to adjust the data. One possibility is to loosen the strains on the import and export structure at national level. This would mean that we 'comment out' equations 6 and 7 and allow that the regions import and export more services to balance their accounts.

While this may seem an attractive feature at first glance, it leads to distortions in the international trade equilibrium. Also it seriously inflates trade in services with the Rest of the World, up to values which are even higher than trade in goods.

If we do not balance through the external markets, we need to achieve an internal balance. We did not want to deviate from the location of production (outputs) as given in the regional database. Its means that the only possibility to achieve the balance is through consumption and value-added adjustments.

The 'entropy' function of the previous section is different for services. Note that imports and exports are also parts of the balancing procedure, but under the restriction of equation 6 and 7.

$$\begin{aligned} E = & \sum_{i=services} \sum_{r} XF_{i,r} \left(\ln XF_{i,r} - \ln XF^{0}{}_{i,r} \right) + \sum_{r} \sum_{i,ii} IO'_{i,r,rr} \left(\ln IO_{i,r,rr} - \ln IO^{0}{}_{i,r,rr} \right) \Big|_{=services} \\ & + \sum_{i=services} \sum_{r} VA_{i,r} \left(\ln VA_{i,r} - \ln VA^{0}{}_{i,r} \right) + \sum_{i=services} \sum_{r} M_{i,r} \left(\ln M_{i,r} - \ln M^{0}{}_{i,r} \right) \\ & + \sum_{i=services} \sum_{r} E_{i,r} \left(\ln E_{i,r} - \ln E^{0}{}_{i,r} \right) \end{aligned} \qquad (16)$$

9. Conclusion

To disaggregate the national SAM at regional level for the SUST-RUS project we modified the cross-entropy approach to allow taking up a mix of interregional and international trade statistics. Under relatively general constraints and assumptions this model has a remarkable capability to balance the regional dataset in a consistent way.

Although the procedure was successfully implemented and was manually checked for errors and inconsistencies we have some caveats. The capability of this procedure to derive an optimal solution is constrained by the logarithmic function, which does not allow the model to have negatives and zero values. Also, the method has a 'black-box' nature, offering solutions which can be hard to explain without a good consistency check of the initial data. If researchers implement a similar procedure we advise to check the limits for each variable and keep good track on the input data. Different configurations of the entropy function should be checked to make sure the results of the procedure are robust.

10. Reference

Canning P., Wang Z., (2005), A flexible mathematical programming model to estimate interregional input-output accounts, Journal of Regional Science, Vol. 45, No.3, 2005, pp539-563.

Francois J.,(2001), Flexible estimation and inference within CGE models, Tinbergen institute.

Huff, K., McDougall, R. and Walmsley, T. (2000). "Contributing Input-Output Tables to the GTAP Data Base," GTAP Technical Paper No. 1, Release 4.2, January.

Kiselev, Sergey, and Romashkin, Roman. (2006). 11.O Russian Federation. Chapter in "*Global Trade, Assistance, and Production: The GTAP 6 Data Base"* Dimaranan, edited by Betina V., (2006), Center for Global Trade Analysis, Purdue University.

Lofgren, Hans, Rebecca Lee Harris, Sherman Robinson. (2002) A standard computable general equilibrium (CGE) model in GAMS / with assistance from Marcelle Thomas and Moataz El-Said, IFPRI, Microcomputares in Policy Research.

Maidment, Terry and Owen Gabbitas. (2006). Chapter 11.B. Regional Input-Output Tables: Australia. in Dimaranan, Betina V., Editor Global Trade, *Assistance, and Production: The GTAP 6 Data Base*, Center for Global Trade Analysis, Purdue University.

Masakova Irina. (1998). Practice of 1993 SNA: Implementation in Russia. Report on Joint OECD/ESCAP meeting on national accounts, 1998. (www.oecd.org/dataoecd/16/62/2665332.pdf).

Robinson, Sherman, Andrea Cattaneo, and Moataz El-Said. (2001). "Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods." Economic Systems Research, Vol. 13, No. 1, pp. 47-64. (http://www.ifpri.org/divs/tmd/dp/tmdp33.htm).

Rosstat. (1998). *Statistical methodology* . Part 2. Moscow. (in Russian) [Росстат. Методологические положения по статистике. Выпуск второй. Москва 1998 год. (http://www.gks.ru/documents/metod/met98_2.arj)].

Rosstat. (2006). *System of Input-Output tables of Russia for 2003*. Moscow. (in Russian) [Росстат. Система таблиц "Затраты-Выпуск" России за 2003 год. Статистический сборник. Москва 2006 год. (http://www.gks.ru/doc_2006/Zatrat06.zip)].

Rosstat. (2008). *National Accounts in Russia, 2000-2007*. Moscow. (in Russian) [Росстат. Национальные счета России в 2000 - 2007 годах. Статистический сборник. Москва 2008 год. (http://www.gks.ru/doc_2008/nac_sh.zip)].

United Nations. (1993). System of National Accounts, United Nations publication, Sales No.E.94. XVII.4.

United Nations. (1999). "Handbook of National Accounting – Handbook of Input-Output Table Compilation and Analysis", Studies In Methods Series F, No.74, Statistics Division, New York (http://unstats.un.org/unsd/publication/SeriesF/SeriesF_74E.pdf).

Appendix A: Aggregated SAM of the SUST-RUS project

Table A1. The Aggregated Social Accounting Matrix of the SUST-RUS project.

	Commo-dities	Activities	Factors: Labour (Wage bill)	Factors: Capital (Profit and mixed income)	Other taxes on production	Production subsidies	Commo-dity and Income Taxes	Households	Government	Investment	ROW account - Exports	Totals
Commodities	-	22 574 710.47	-	-	-	-	-	11 477 043.79	4 729 287.67	5 514 918.41	7 205 326.16	51 501 286.5(
Activities	46 338 692.8(	-	-	-	-	-	-	-	-	-	-	46 338 692.8(
Factors: Labour (Wage bill)	-	9 627 029.75	-	-	-	-	-	-	-	-	-	9 627 029.75
Factors: Capital (Profit and mixed income)	-	12 888 771.77	-	-	-	-	-	-	-	-	-	12 888 771.77
Other taxes on production	-	436 601.20	-	-	-	-	-	-	-	-	-	436 601.20
Production subsidies	-	- 17 366.29	-	-	-	-	-	-	-	-	-	-17 366.29
Commodity and Income Taxes	-	828 945.88	-	-	-	--	-	1 179 455.50	1 615.89	213 974.99	1 715 865.33	3 939 857.60
Households	-	-	9 627 029.75	12 888 771.77	-	-	-	-	2 272 868.00	-	-	24 788 669.53
Government	-	-	-	-	436 601.20	- 17 366.29	3 939 857.60	6 418 984.05	-	-	-	10 778 076.56
Savings	-	-	-	-	-	-	-	5 713 186.19	3 774 305.00	-	-	9 487 491.19
ROW account - Imports	5 162 593.70	-	-	-	-	-	-	--	-	3 758 597.79	-	8 921 191.49
Totals	51 501 286.5(	46 338 692.80	9 627 029.75	12 888 771.77	436 601.20	-17 366.29	3 939 857.60	24 788 669.53	10 778 076.56	9 487 491.19	8 921 191.49	

Appendix B: Preparatory data management with Russian regional trade statistics

State of the original data

Original data on Russian interregional trade for years 2001-2006 are stored in approximately 2500 excel files. Each file contains data for a certain year, direction of trade flows, and the commodity title. Within each file data are further split up to approximately 15-20 tables. Tables have a certain pattern: rows refer to destination regions, columns refer to regions of departure, destination regions are listed by names (sometimes with different spelling), regions of departure listed by their IDs (with some possible mistakes) and names. Each excel file name begins with the letters "BBF", followed by the code of flow direction (1 or 2) and later by the commodity code, for example, "BBF1_221.XLS". Original data are transferred to STATA software using a Perl program (listed below).

The Perl program, used for raw data processing, parses each file (converting excel file to two-dimensional array), runs through array recognizing distinct tables, processes inconsistencies between regions' names and their IDs and puts each observation to output file, which, in turn, is transferred to STATA format. The program code is available upon request.

List of variables in STATA database:

Tablenumber – the number of the table this observation was taken from,
To – destination region,
From – region of departure,
Value – value of trade,
Filename – the name of the file this observation was taken from,
Year – year,
Code – commodity code,
Unit – 1 stands for quantity, 2 stands for nominal value.

Further data adjustments

Russian interregional trade data does not fully coincide with regional output data, i.e. total export of the region to itself, to other regions and abroad is not equal to the total regional output. And since reliability of regional output data is much higher, we decided to revalue interregional trade flows. Further, we assume that interregional data imply correct proportions of the export of a certain region to other regions and abroad.

Let $T_{r_1 r_2}$ be a trade from the region r_1 to the region (or abroad) r_2 (according to interregional trade data) and O_r - the total output of the region r. Then estimated value of interregional trade is $T^E_{r_1 r_2} = \frac{O_r T_{r_1 r_2}}{\sum_r T_{r_1 r}}$.

Table B-1. Exports of the Central Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from CFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
301			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry	297029852	42247914	11239817	5285808.6	886559.92	1167344.6	3562147.9
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	106256173	368823.92	0	0	0	0	0
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	43253792	221351.84	79512.332	5631618.8	3476531.2	5873954.4	256.2844
DA	Subsection DA	Manufacture of food products, beverages and tobacco	415050447	163361587	6970831.3	20662217	1150940.9	30650183	9468456.4
DB	Subsection DB	Manufacture of textiles and textile products	64682841	1248612.2	1755127.9	2222227.8	195022.67	385048.91	247506.98
DC	Subsection DC	Manufacture of leather and leather products	7228991.4	608416.05	779241.88	580588.55	382781.86	192639.54	8664.5662
DD	Subsection DD	Manufacture of wood and wood products	28846299	2628177.2	1751896.9	2967725.7	680891.01	585722.17	59641.214
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	136203228	8800292.4	3605866.8	7727542.8	784960.05	5622613.4	47792.824
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	98727686	30375954	8020493.1	10985039	1617407.5	2766511.5	1479882.2
DH	Subsection DH	Manufacture of rubber and plastic products	58400081	6889843.6	1800883	7483926.1	1867796.4	821237.11	113495.35
DI	Subsection DI	Manufacture of other non-metallic mineral products	203300731	6407972.5	5210544.9	6046790.8	2216668.3	2103208.6	70787.88
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	77535815	8116085.2	19486096	53935521	12713602	5935539.1	7754066.2
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	138811314	18494514	5060077.5	18532679	11046019	7779378.9	17571547
DM	Subsection DM	Manufacture of transport equipment	148044703	25311094	3857419.3	9833112.1	5014800.6	2004437	1377849
DN	Subsection DN	Manufacturing n.e.c.	75352624	5446210.9	3507418.9	8564375.1	2289722.4	4703123.6	1806327.1

Table B-2. Exports of the Northwestern Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from NWFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
311			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry							
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	24325360.53	90974628.78	152847.2307	1868041.991	17159774.73	0	0
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	5740151.819	46710388.71	1401.477346	162372.3171	1428870.584	426287.0084	0
DA	Subsection DA	Manufacture of food products, beverages and tobacco	5285843.26	233579513.3	19294314.67	1123628.131	3801955.876	7053207.612	0
DB	Subsection DB	Manufacture of textiles and textile products	6364416.629	4772266.41	0	586853.8589	0	0	0
DC	Subsection DC	Manufacture of leather and leather products	142066.1656	987436.7059	171.551603	17780.60815	0	302052.339	0
DD	Subsection DD	Manufacture of wood and wood products	4280347.38	17018023.69	584048.9258	1254297.905	92795.68357	173522.1826	5150.654765
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	25550037.63	20608302.36	2896278.143	5236674.941	801768.8711	2058158.942	225331.3103
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	7966250.517	9724029.777	2690846.94	3639822.34	638837.7729	541787.4991	272421.3738
DH	Subsection DH	Manufacture of rubber and plastic products	827242.0164	14826824.84	0	128590.3189	149627.2227	0	0
DI	Subsection DI	Manufacture of other non-metallic mineral products	12006122.19	47260188.82	465269.1625	1299612.752	653554.619	1000207.805	27467.0801
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	52730774.84	73191921.31	9537204.286	72916601.24	8609746.16	3403322.661	101097.0642
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	33479442.19	39611086.91	5127266.341	8622390.207	3364655.564	1763757.659	683274.9318
DM	Subsection DM	Manufacture of transport equipment	36861235.91	31492926.45	2957645.065	15194612.04	2421787.535	1008017.951	1835650.754
DN	Subsection DN	Manufacturing n.e.c.	6810032.093	11226551.66	1655583.412	4269049.391	3182658.78	3005049.547	1162620.637

Table B-3. Exports of the Southern Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from SFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
321			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry	62305423	4520051.2	264975655	2535635.4	167492.09	543397.76	0
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	28915999	0	31573322	0	0	0	0
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	548033.58	768839.18	3503880.6	52795.362	135.93457	0	0
DA	Subsection DA	Manufacture of food products, beverages and tobacco	58675977	22160075	87110412	10875500	2292004.3	13100251	2737638.8
DB	Subsection DB	Manufacture of textiles and textile products	8784246.5	106014.27	2477854.4	115187.52	319641.54	0	0
DC	Subsection DC	Manufacture of leather and leather products	251837.38	1927.0576	1515551.6	103836.72	184.78634	0	0
DD	Subsection DD	Manufacture of wood and wood products	171793.04	47540.415	3645561.2	8205.5505	9119.4242	10691.003	5157.9485
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	536295.89	128620.42	14029590	59797.954	40408.256	50901.873	0
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	12714080	1190653.6	10380295	3336060.4	1060734.3	1117798.2	141718.31
DH	Subsection DH	Manufacture of rubber and plastic products	3496859.7	531314.95	5540897.4	1472517.2	703300.35	539905.93	20312.798
DI	Subsection DI	Manufacture of other non-metallic mineral products	5099008.1	262415.77	44655032	1169402.2	237638.84	587248.1	38770.77
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	18660197	3964283.6	42910781	6069661.8	12633535	2827484.3	180084.1
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	10298185	1773883.7	4908268.2	3226433.4	192909.04	84728.333	283932.53
DM	Subsection DM	Manufacture of transport equipment	36168184	405606.82	10022415	1215113.2	881434.98	411706.97	0
DN	Subsection DN	Manufacturing n.e.c.	2680171.9	1229339.1	8049343	0	0	1260865.3	0

Table B-4. Exports of the Volga Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from VFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
331			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry	53834341	16287866	16953393	344843712	2487290.5	0	0
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	635804070	0	0	0
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	4332493.1	1556226.1	922313.99	23781722	2184872.1	124179.54	93.239787
DA	Subsection DA	Manufacture of food products, beverages and tobacco	76208732	41534015	7226786.7	103528673	5456046.9	20833764	8059805.3
DB	Subsection DB	Manufacture of textiles and textile products	8866960.8	342873.06	1933113.7	9206084.9	905048.67	516209.43	60139.93
DC	Subsection DC	Manufacture of leather and leather products	1354158.6	93420.646	121694.92	1399513.4	203654.13	5919.9402	0
DD	Subsection DD	Manufacture of wood and wood products	3397562.5	1133752	967510.44	8943560.3	257276.8	34422.928	25974.995
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	13248659	2808213	1871349.1	12105167	1954575.1	2269387.2	415863.47
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	36562598	14489885	12727409	47786993	7358101.3	7555551.8	806090.2
DH	Subsection DH	Manufacture of rubber and plastic products	17414454	2657427.2	7239696.5	39250532	4561267.1	5535902	264625.18
DI	Subsection DI	Manufacture of other non-metallic mineral products	14173791	1435769.7	1342009	62639893	7151929	1767548.7	38312.441
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	34871062	34073378	7552196.7	44089911	80440201	36102495	1063291.7
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	31797249	7237943.4	6127977.2	58582664	9707917.2	4341483.5	793008.62
DM	Subsection DM	Manufacture of transport equipment	167637077	31618791	14486716	125519510	42970872	32083102	7678914.9
DN	Subsection DN	Manufacturing n.e.c.	42756184	66656.401	1081538.9	3635557.5	242309.76	762762.26	420863.18

Table B-5. Exports of the Urals Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from UFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
341			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry	16827493	10567236	7552509.1	25649554	67403993	3191211.1	0
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	4650278.7	1.974E+09	0	0
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	2492836.5	221553.66	0	7077779.9	38445797	121884.38	0
DA	Subsection DA	Manufacture of food products, beverages and tobacco	2772822.8	32498773	105887.68	4484276	31057187	8917451	0
DB	Subsection DB	Manufacture of textiles and textile products	1557932.1	0	0	654976.65	1785003.7	59731.268	0
DC	Subsection DC	Manufacture of leather and leather products	81116.158	641743.05	0	7962.5025	790973.91	37363.742	0
DD	Subsection DD	Manufacture of wood and wood products	588662.57	235582.85	322788.23	1182536.9	5097411.9	155774.58	241485.37
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	1494663.3	155520.36	15710.287	863940.61	3974219.7	1661596.4	40073.674
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	4907575.9	1817822.7	1785608.7	4518070.6	3913492.1	3326326.7	918615.48
DH	Subsection DH	Manufacture of rubber and plastic products	2005183.7	181405.06	243452.18	720153.31	9530378.3	393782.59	0
DI	Subsection DI	Manufacture of other non-metallic mineral products	326904.14	661860.06	264990.72	9755691.1	54987120	1949497.3	116121.5
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	79696240	24831330	25005761	96259823	217745418	43479226	4658355.1
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	10116781	3750071.1	2856684.6	8231532.1	11538284	6832012.7	489047.2
DM	Subsection DM	Manufacture of transport equipment	21917064	3039264.2	609041.65	3693035.7	23486928	2582073.3	526897.52
DN	Subsection DN	Manufacturing n.e.c.	3538155.6	1755887.2	2860799.6	5029931.7	18965295	4954136.4	3636239.4

Table B-6. Exports of the Siberian Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from SFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
351			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry	9336429.4	5647186.1	72554582	2980860.3	1420425.7	151266802	6709711.3
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	27524799	8149531.8	1701149.1	5054817.2	33962601	111266715	2115441.4
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	48386.214	0	2639.8077	0	12837763	45469899	195194.96
DA	Subsection DA	Manufacture of food products, beverages and tobacco	5335764.3	42045472	0	5679261.4	12737605	87594715	16742700
DB	Subsection DB	Manufacture of textiles and textile products	1606828.3	0	29507.228	2391.598	147.99492	3543069.4	1331.9543
DC	Subsection DC	Manufacture of leather and leather products	0	0	0	0	0	1062821.1	0
DD	Subsection DD	Manufacture of wood and wood products	102797.93	54079.919	409094.96	62347.36	192411.34	11052801	460495.07
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	497477.61	2032052.2	130250.36	3882381.9	367653.58	3755321.6	184550.34
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	18889726	2845802.3	1672253.3	7279922.2	3311309.2	19090672	745851.47
DH	Subsection DH	Manufacture of rubber and plastic products	4260818.6	210944.7	1059758.4	2420962	2225858.8	6773482.9	471038.54
DI	Subsection DI	Manufacture of other non-metallic mineral products	1423800.9	206668.05	24088.473	511319.9	1134540.7	37808158	297857.51
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	81459737	12152924	20884520	48415551	40174483	126205136	8280342.2
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	13162978	1972765	2112587.1	7412263.9	5227825.2	13005567	1592246.5
DM	Subsection DM	Manufacture of transport equipment	37830329	4808951.7	250026.15	42387.807	307758.28	3593074.4	626177.66
DN	Subsection DN	Manufacturing n.e.c.	5549861.8	1796176.7	834321.85	1786061.8	1373321	2183554.7	893958.93

Table B-7. Exports of the Far Eastern Federal District to other Russian Federal Districts, according to the Russian regional Statistics database

from FEFD			Central Federal District	Northwestern Federal District	Southern Federal District	Volga Federal District	Urals Federal District	Siberian Federal District	Far Eastern Federal District
361			301	311	321	331	341	351	361
A	Section A	Agriculture, hunting and forestry	0	0	0	0	0	0	61477000
B	Section B	Fishing							
CA_col	Subsection CA	Mining and quarrying of energy producing materials	4255045	0	0	590643.02	5282244.4	498336.54	36669322
CA_gas	Subsection CA	Mining and quarrying of energy producing materials	0	0	0	0	0	0	0
CB	Subsection CB	Mining and quarrying, except of energy producing materials	0	0	0	0	0	0	145841654
DA	Subsection DA	Manufacture of food products, beverages and tobacco	0	0	0	0	0	0	51066846
DB	Subsection DB	Manufacture of textiles and textile products	0	0	0	0	0	0	0
DC	Subsection DC	Manufacture of leather and leather products	0	0	0	0	0	0	345883.9
DD	Subsection DD	Manufacture of wood and wood products	269.18746	0	442.29405	0	721.43601	5203.2252	633717.29
DE	Subsection DE	Manufacture of pulp, paper and paper products; publishing and printing	0	0	0	0	0	0	3605536.8
DF	Subsection DF	Manufacture of coke, refined petroleum products and nuclear fuel							
DG	Subsection DG	Manufacture of chemicals, chemical products and man-made fibres	0	0	0	0	0	0	2882243.5
DH	Subsection DH	Manufacture of rubber and plastic products	0	0	0	0	0	0	2372296.3
DI	Subsection DI	Manufacture of other non-metallic mineral products	77257.452	0	0	0	0	188661.7	11434193
DJ	Subsection DJ	Manufacture of basic metals and fabricated metal products	6592.5825	8402.7551	0	1618.5278	376078.61	3530277.4	6405555.4
DK	Subsection DK	Manufacture of machinery and equipment n.e.c.							
DL	Subsection DL	Manufacture of electrical and optical equipment	279065.13	61049.78	84309.015	1244986	172799.21	763259.34	2019559.7
DM	Subsection DM	Manufacture of transport equipment	0	0	0	0	0	0	0
DN	Subsection DN	Manufacturing n.e.c.	166223.76	54299.979	9534.3767	169615.45	9632.0755	389780.27	5427164.6

Appendix C: Cross-entropy minimization technique for the IO disaggregation

This Appendix is formalizing the problem of disaggregation of an input-output matrix.

Definition. Let's define a *classification* K as a non-intersecting sum of sets I, R, C, where set I consists of names of industries presented in an input-output table, set R consists of names of value-added categories of an input-output table, and set C consists of names of final consumption accounts in an input-output table.

Definition. An *input-output matrix* M in a classification K has the following block structure:

M_{11}	M_{12}	M_{13}
M_{21}	0	0
M_{31}	0	0

where M_{11} is an $I \times I$ matrix, M_{12} - $I \times C$ matrix, M_{21} - $R \times I$ matrix, M_{13} is a $I \times 1$ column, and M_{31} is a $1 \times I$ row. An *input-output matrix* M satisfies the following properties:

1. Row sum of M_{11} and M_{21} gives M_{31}

$$\sum_i (M_{11})_{ij} + \sum_i (M_{21})_{ij} = (M_{13})_j$$

2. Column sum of M_{11} and M_{12} gives M_{13}

$$\sum_j (M_{11})_{ij} + \sum_j (M_{12})_{ij} = (M_{31})_i$$

3. $M_{13}{}^T = M_{31}$

4. M_{11} contains only non-negative elements.

Let's define M_K as the family of all IO tables in classification K that satisfy properties (1)-(4).

Definition. We say that classification K_1 is more detailed compared to classification K_2 (in other words: K_1 can be aggregated to K_2), if there is a single-value surjective operator $g(\cdot)$ from K_1 to K_2 that preserves classification's subdivision on industries, VA rows and final consumption columns (preserving subdivision means that if s, for example, is in $I_1 \subset K_1$ then $g(s)$ is in $I_2 \subset K_2$). That is, each element of the more detailed classification K_1 is contained in some element of K_2, and each element of K_2 contains at least one of K_1 (here we implicitly assume that an element of some classification is a subset of the set of all types of economic activities).

Let's describe a case in which there is no aggregation relation between classifications. Let $K = \{1,2,3\}$ be a set of economic activities; there are two classifications $K_1 = \{(1,2),3\}$ and $K_2 = \{1,(2,3)\}$. It is obvious that neither K_1 is more detailed than K_2, nor K_2 is. But a more detailed classification K_3 always exists, in our case it is the classification K.

To point that, consider two classifications K_1 and K_2 such that K_2 is less detailed table $M_1 \in M_{K_1}$ (IO table in the classification K_1). Define aggregating operator $A: M_{K_1} \to M_{K_2}$ (an operator from space of tables in classification K_1 to space M_{K_2}) such that for all elements of classification K_2 $s,t \in K_2$ $(M_2)_{st} = A(M_1)_{ij} = \sum_{i:g(i)=s} \sum_{j:g(i)=t} M_{1ij}$, where $i, j \in K_1$. Let's note that aggregation is defined correctly, so the image of operator $A(\cdot)$ is a subset of M_2. In other words aggregating operator preserves all the properties (1)-(4). More over it is important to note that image of $A(\cdot)$ is unique.

Let's define inverse operator to $A(\cdot)$ – a *disaggregating operator* $A^{-1}: M_{k_2} \to M_{K_1}$. Let $M_2 \in M_{K_2}$ be an IO table in less detailed classification K_2, then the set $A^{-1}(M_2) \subset M_{K_1}$ is all tables M_1 in a more detailed classification K_1 such that $A(M_1) = M_2$. However, operator $A^{-1}: M_{k_2} \to M_{K_1}$ is not a single-value mapping in contrast to aggregating procedure. In other words, its image contains not a single IO table, but a set of tables.

First, we will show that disaggregating operator is defined correctly; that is, an image of $A^{-1}(\cdot)$ is nonempty and is contained in M_{K_1}. $A^{-1}(M_2)$ is *non-empty* if there exists M_1 from M_{K_1} such that $A(M_1) = M_2$. Indeed, let's construct an IO table in classification K_1 that is in image of disaggregating operator $A^{-1}(\cdot)$ for certain. Recall that operator aggregating classification K_1 to K_2 we denote as $g(\cdot)$. Define proportional disaggregating operator $A^{-1}{}_{pr}(\cdot)$ as mapping from set of tables in less detailed classification M_{K_2} that divide each element $(M_2)_{st}, s,t \in K_2$ into equal parts between all the elements $(M_1)_{ij}$, $i, j \in K_1$ such that $g(i) = s, g(j) = t$. It is obvious that this procedure is an aggregation. Indeed, the only thing we have to check is that it preserves IO properties (1)-(4). Consider some element s of classification K_2, and assume that it corresponds to elements $i_1, i_2, ..., i_{|s|}$ of classification K_1 (i.e. $g(i_s) = s$). If s is an industry ($s \in I_{K_1}$), it is sufficient to check that the sum of every column i_s equals the sum of row i_s. As soon as disaggregating operator domain is a set of IO tables, matrix M_2 satisfies all IO table conditions, and consequently, the sum of any row (column) i_s of table M_1 equals the sum of row (column) s of table M_2 divided by $\bar{s} = |g^{-1}(s)|$ (where $|g^{-1}(s)|$ is the power of set $g^{-1}(s)$). Thus, we see that condition (3) for IO table M_2 is identical with the condition for M_1. Conditions (1), (2) and (4) are also obviously satisfied.

So we checked that disaggregation procedure is well-defined, but disaggregating operator $A^{-1}(\cdot)$ is not a single-value mapping. Thus, all elements of set $A^{-1}(M_2)$ become equivalent without some extra information about the structure of the matrix M_1 in a more detailed classification. Further, we will consider a setting that solves difficulties arising with multiple solutions of the disaggregation problem.

Let M_1^{y1} be an IO table for some year y_1 in a classification K_1 (more detailed classification), and M_2^{y2} is an IO table for the benchmark year y_2 in classification K_2. We assume that tables for different years in one classification have close structures, thus, we are estimating IO table M_1^{y2} of the year y_2 in the classification K_1 extracting information about its structure from M_1^{y1} table. Thus, our problem can be reformulated as a problem of finding a closest table M_1^{y2} to the prior M_1^{y1}.

$$\rho(\overline{M}, M_1^{y1}) \longrightarrow \max$$

$$s.t. \quad \overline{M} \in A^{-1}(M_2^{y2}) \qquad (*)$$

Where $\rho(\cdot,\cdot)$ – is a measure of closeness in the space of IO tables. Note that $\rho(\cdot,\cdot)$ need not be a metric in a mathematical sense. For our problem we don't need the triangle inequality property.

There are some examples of a measure of closeness $\rho(\cdot,\cdot)$:

1. Euclidian metric.

$$\rho_1(M', M'') = \sum_i \sum_j (M'_{ij} - M''_{ij})^2$$

2. Weighted sum of squares.

$$\rho_2(M', M'') = \sum_i \sum_j \frac{(M'_{ij} - M''_{ij})^2}{M''_{ij}}$$

3. Kullback-Leibler divergence.

$$\rho_3(M', M'') = \sum_i \sum_j M'_{ij}(\ln M'_{ij} - \ln M''_{ij})$$

In fact, Kullback-Liebler divergence is defined only for distributions, not for arbitrary non-negative vectors. This is due to fact that the Gibbs' inequality ($D_{KL}(p,q) \geq 0, \forall p,q, \quad and = 0, iff \quad p = q$) holds for sure only for p,q such that $0 \leq p,q \leq 1, \sum_i p_i = \sum_i q_i = 1$. However, it is possible to prove an extension of the Gibbs' inequality.

Let $P_i, Q_i, i = 1...N$, be arbitrary positive numbers $P_i, Q_i \geq 0, \forall i$. Then consider two auxiliary vectors $p_i = \frac{P_i}{\overline{P}}$, where $\overline{P} = \sum_i P_i$, and $q_i = \frac{Q_i}{\overline{Q}}$, where $\overline{Q} = \sum_i Q_i$. It is easy to check that p and q are distributions. And thus, Gibbs' inequality holds: $D_{KL}(p,q) \geq 0$.

Extended inequality for vectors P,Q takes on the following form:

$$D_{KL}(P,Q) = \sum_{i=1}^{N} P_i(\ln P_i - \ln Q_i) = \sum_i p_i \overline{P}(\ln \overline{P} + \ln p_i - \ln \overline{Q} - \ln q_i) =$$

$$\overline{P}\sum_i \left[p_i(\ln \overline{P} - \ln \overline{Q}) + \overline{P}(p_i(\ln p_i - \ln q_i))\right] = \overline{P}(\ln \overline{P} - \ln \overline{Q}) + \overline{P} D_{KL}(p,q) \geq \overline{P}(\ln \overline{P} - \ln \overline{Q})$$

so that $D_{KL}(P,Q) = c_1 D_{KL}(p,q) + c_2$, where c_1, c_2 are constants. And since $D_{KL}(P,Q)$ is just a monotone transformation of the distance measure for normalized vectors p, q, the following identity holds:

$$\arg\max_{P \in F} \{D_{KL}(P,Q)\} = \arg\max_{p \in F} \{D_{KL}(p,q)\}.$$

That is, although $\rho_3(M', M'')$ can take on a negative value and it is not a true distance, the disaggregation problem is defined correctly.

While choosing the measure of closeness $\rho(\cdot,\cdot)$, one should take into account the following:

domain of $\rho(\cdot,\cdot)$;
suitability of numerical methods;
table structure preserving.

Domain of $\rho(\cdot,\cdot)$. Note that functions $\rho_2(\cdot,\cdot)$ and $\rho_3(\cdot,\cdot)$ are not defined for a table with non-positive elements, whereas any IO table contains a block of zero elements, and there could be negative elements in blocks A_{12}, A_{21}. Consequently, most of suggested "distance measures" fail to work in this case.

However, we can adjust metric to extend its domain; for example, $\rho_4(\cdot,\cdot)$ is an adjustment of WSS:

$\rho_4(M', M'') = \sum_{i,j} \frac{(M'_{ij} - M''_{ij})^2}{f(M''_{ij})}$, where

$$f(x) = \begin{cases} x, \text{ for } x \geq \varepsilon \\ \varepsilon, \text{ for } x \in [-\varepsilon, \varepsilon] \\ -x, \text{ for } x \leq -\varepsilon \end{cases}$$

In contrast to $\rho_3(\cdot,\cdot)$, metric $\rho_4(\cdot,\cdot)$ (adjusted weighted sum of squares) is defined on all matrices (even with negative elements). However, it is not continuous with respect to the second argument, which could cause difficulties with numerical solution described below. The domain of Kullback-Leibler divergence could be expanded similarly, but again with discontinuities. Thus, any metric adjustment complicates the disaggregation problem. This suggests that the object to adjust is an IO table, not the metric.

So, let M be some IO table with the set of rows $\overline{R} = I \cup R \cup SUMR$ and the set of columns $\overline{C} = I \cup C \cup SUMC$. We will expand the table M to M' so that the expanded sets of rows and columns are $\overline{R}^* = I \cup R \cup C \cup SUMR$, $\overline{C}^* = I \cup C \cup R \cup SUMC$; define M' as the follows:

M'_{11}	M'_{12}	M'_{13}	M'_{14}
M'_{21}	0	0	0
M'_{31}	0	0	0
M'_{41}	0	0	0

Where $M'_{11} = M_{11}$ - $I \times I$ block,

$M'_{21} = \max\{0, M_{21}\}$ - $C \times I$ block,

$M'_{31} = (-\min\{0, M_{12}\})$ - $R \times I$ block,

$M'_{41} = sum(M_{11}, M_{21}, M_{31})$ (rows summation) - $1 \times I$ block,

And similarly for M_{12}, M_{13}, M_{14} .

Thus, using one-to-one transformation of an arbitrary table M, we get the table M' that has no negative elements. Note that it is a one-to-one procedure, i.e. it preserves all the information about M . Zero elements can be excluded as well by adding small positive ε to every element of matrix M': $M'' = M' + \varepsilon \cdot E$.

Numerical methods

Since disaggregation of huge tables is a complex computational process, one needs to use numerical methods. Thus, any distance measure that worsens the convergence of the method due to its functional characteristic is not valid. Continuity and differentiability of value function and of the set of constraints are the characteristics required for numerical methods. Also note that the minimization problem (*) can give a solution that doesn't satisfy IO properties, especially for one – non-negativity of table elements. Thus, we have to add some constraints which ensure that the solution will be an IO table (satisfying all the IO properties). New constraints complicate the problem further, so we have to choose the value function with least possible rigid constraints.

Table structure preserving

Choosing the metric $\rho(\cdot,\cdot)$, we tend to get a solution that has the closest structure to the prior. That is, changes of the estimated table elements compared to the prior are almost the same around the table. We might consider two types of changes – absolute and proportional. Below we discuss advantages and disadvantages of these structure types.

Again, consider the disaggregation problem with metrics (1)-(3) mentioned above.

$$\rho(\overline{M}, M_1^{y1}) \longrightarrow \max$$
$$s.t. \quad \overline{M} \in A^{-1}\left(M_2^{y2}\right) \qquad (*)$$

To make some conclusions about the structure type of metric, write down the analytical solution of the problem (*)

Euclidian metric – $\rho_1(M, M') = \sum_{ij}\left(M_{ij} - M'_{ij}\right)^2$

$FOC.$ $2\left(M_{ij} - M'_{ij}\right) = \lambda_{st}$, for $\forall i, j : g(i) = s, g(j) = t$,

where λ_{st} are Lagrange multipliers of the corresponding aggregation constraint.

Weighted sum of squares – $\rho_2(M', M'') = \sum_{ij} \frac{\left(M'_{ij} - M''_{ij}\right)^2}{M''_{ij}}$.

$FOC.\quad 2\frac{\left(M_{ij}-M^{'}{}_{ij}\right)}{M^{'}{}_{ij}}=\lambda_{st}$, for $\forall i,j: g(i)=s, g(j)=t$.

Or equivalently, $\frac{M_{ij}}{M^{'}{}_{ij}}=\mu_{st}$, where $\mu_{st}=\frac{\lambda_{st}+2}{2}$.

Cross-entropy – $\rho_3\left(M^{'},M^{''}\right)=\sum_{ij}M^{'}_{ij}\left(\ln M^{'}_{ij}-\ln M^{''}_{ij}\right)$.

$FOC.\quad \left(\ln M_{ij}-\ln M^{'}{}_{ij}\right)+1=\lambda_{st}$, for $\forall i,j: g(i)=s, g(j)=t$,

or, $\frac{M_{ij}}{M^{'}{}_{ij}}=\upsilon_{st}$, where $\upsilon_{st}=e^{\lambda_{st}-1}$.

Note that in each case we face convex minimization problem, and consequently, first order conditions are sufficient for it being the minimum. Also, note that the Euclidian metric tends to equalize absolute changes in estimation compared to the prior, whereas WSS and KLD equalize proportional ones. And thus, there could arise negative element is the estimation $M_1^{y_1}$ even if both tables M_1^{y1} and M_2^{y2} were non-negative. The example is the following:

M_1^{y1}	a	b1	b2	c	sum
a	1	1	1	1	4
b1	1	100	1	100	202
b2	1	100	1	1	103
c	1	1	100	0	0
sum	4	202	103	0	0

M_2^{y2}	a	b	c	sum
a	1	2	1	4
b	2	1	1	4
c	1	1	0	0
sum	4	4	0	0

Hence, the **optimal criterion** is the one that tends to equalize proportions of the estimation to the prior.