Paper prepared for the 14th Global Economic Analysis Conference, *'Governing Global Challenges: Climate Change, Trade, Finance and Development'*,
Ca' Foscari University of Venice, Italy, June 2011

An Inventory of Datasets for the Compilation of Regional Social Accounting Matrices for the EU

Marc Mueller, Center for Development Research, University of Bonn, Germany
Emanuele Ferrari, European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS), Spain.

WORK IN PROGRESS: PRELIMINARY DRAFT.
PLEASE DO NOT QUOTE WITHOUT PRIOR AGREEMENT WITH THE AUTHORS.

Abstract:
Due to the ever-increasing demand for model-based analyses of regional development policies in a multi-sector context, the Institute for Prospective Technological Studies (IPTS) launched in 2009 a project on the compilation of regional Social Accounting Matrices (SAM) for the NUTS2 regions of the EU (IOTNUTS2). The SAMs cover the time span between 2000 and 2005. This database shall permit general equilibrium analyses of policies like reforestation programmes, the promotion of investment in agro-tourism or environmental services, and the support for the production of renewable energy by farming enterprises, and more generally to evaluate the rural development pillar of the European Common Agricultural Policy. Such measures target primarily the agricultural sector, but are likely to have an impact on other economic sectors and aggregate regional income, depending on the regional economic structure and the dominance of agriculture. Addressing regional heterogeneity requires multi-sector data on sub-national scale. Such datasets are usually not sufficiently detailed, if available at all, which gave raise to numerous non-survey methods to generate regional Input-Output tables based on combinations of available regional indicators and national datasets (e.g. Location Quotients, GRIT methods). One particular challenge encountered during the IOTNUTS2 project was the high level of sectoral aggregation in regional branch accounts provided by ESTAT, where e.g. agriculture, forestry, and fisheries are merged. Given the interest in spill over effects of dominantly agricultural policies, more detailed information was required. Therefore, statistical organisations of the 27 EU Member States were contacted and the results of previous projects on regional databases were screened. This paper gives an overview on the compiled inventory on regional datasets for EU27, starting with the target structure of the database and available national and regional datasets from ESTAT. Based on this, we discuss the datasets obtained from national statistical departments (NSO) and from previous projects with comparable aims. In general, we achieved a relevant informational gain over the exclusive use of ESTAT datasets for several member states while for some of them (i.e. Bulgaria) not as large as initially expected. Furthermore, we used the obtained NSO data to test the reliability of non-survey methods for the combination of national and regional datasets. It appeared that particularly forestry, mining/quarrying, and fuel industries displayed substantial deviations between derived indicators and those obtained from NSO, namely intermediate demand and gross output. For other branches, information could either be obtained (e.g. agriculture) or derived indicators proved to be close to the NSO values (most service sectors). In general, we conclude that for the majority of considered economic sectors, non-survey methods can generate reliable substitutes for otherwise collected indicators, but not for some critical branches which are usually concentrated in some regions and may

dominate the regional economic structure (forestry, mining, fisheries). This result can be helpful for future projects with comparable aims as we suggest that instead of attempting to sample economy-wide datasets, a focus on the mentioned critical sectors would provide higher marginal informational gains. The data collected from all the different sources are firstly utilized to populate national Input-Output tables for the EU 27 Member States. These matrixes are then balances following standard cross-entropy methods. These tables, with the suitable level of disaggregation, could be utilized as starting point to update the EU Input-Output tables that IPTS provided to the GTAP Consortium.

1. Background and Structure of the Target Database[1]

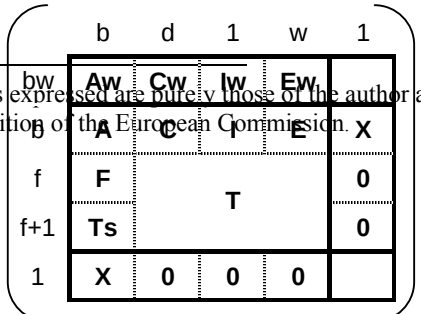
The reforms of the Common Agricultural Policy (CAP) will create an increasing interest of policy makers and analysts in the back- and forward linkages between regional agricultural and non-agricultural sectors, related labour markets, and regional development policies. The development of a European-wide database of Social Accounting Matrices (SAM) on regional scale according to the "Nomenclature of Territorial Units for Statistics (NUTS)" is a prerequisite for the usage of regionalised General Equilibrium Models (CGE) for quantitative evaluation of regional policies of the EU and their Member States on a regional scale. The resulting informational demand will have to be met by complete socio-economic datasets on regional scale. However, such a database does not yet exist with the needed level of details, mainly because of the scarcity of relevant data on regional scale. Consequently, any project aiming at the compilation of a full set of regional SAMs will have to anticipate the absence of survey data or recorded data from statistical departments and provide state-of-the-art solutions for the incorporation of any type of regional information that becomes available.

The goal of this paper is to provide an overview on the availability of relevant datasets on NUTS2 level from statistical organisations of the EU27 Member States. The existence and availability of regional datasets from ESTAT and statistical departments and research institutes of the Member States was investigated first by researching the respective internet sites. After having obtained an overview on the publicly available data, the respective contact centres or – if possible – individual desk officers were approached. The existence of non-published data and the conditions for release were investigated.

Apart from providing an overview on the existence and availability of regional datasets, this inventory prepares the ground for the subsequent stages. The inventory allows identifying those datasets, which are certainly available for all regions and those, which will have to be estimated. The latter information will be useful for the development of non-survey methods built on the certainly available datasets.

The targeted database consists of numerous sub-tables, classified into core-SAM accounts and satellite or auxiliary accounts. The core datasets appear in the final SAMs, such as intermediate demand distinguished by origin and economic branch or compensation of employees by branch. Satellite accounts serve as control-totals for the core datasets, i.e. gross value-added does not appear in the final SAM, but is the sum of "compensation of employees", "net taxes on production", and "operating surplus".

A crucial feature of the database is the distinction of some items by their origin, i.e. if they were produced in the same region in which they were consumed, or imported from a different region in the same country, or imported from abroad. The core SAMs and the respective sub-matrices are summarized in a simplified manner below:



[1] The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

S ≡

Where:

b: Index for economic branches

d: Index for regional institutions like private households and governments

w: Index for trade partners or origins of the items in the SAM (regional, domestic, foreign)

f: Index for primary factors of production (labour, land, capital)

S: Regional Social Accounting Matrix

Aw: Intermediate demand, distinguished by origin

A: Total intermediate demand ($A_{b,b'} = \sum_{w} Aw_{b,w,b'}$)

Cw: Final demand, distinguished by origin

C: Total final demand ($C_{b,d} = \sum_{w} Cw_{b,w,d}$)

Iw: Investment demand, distinguished by origin

I: Total investment demand ($I_b = \sum_{w} Iw_{b,w}$)

Ew: Exports, distinguished by origin and destination (e.g. exports of imported goods to other regions of the same country)

E: Total exports by destination

F: Payments to fixed factors ($E_{b,w} = \sum_{w'} Ew_{b,w',w}$)

Ts: Taxes on production and factors

T: Transactions between institutions (distribution of regional income)

X: Total regional supply (has to equal total regional demand):

$$X_{b'} = \sum_{b} A_{b,b'} + \sum_{f} \left[F_{f,b'} + Ts_{f,b'} \right] + Ts_{*,b'}$$

$$= \sum_{b} A_{b',b} + \sum_{d} C_{b',d} + I_{b'} + \sum_{w} E_{b',w}$$

The branch classification (b) for the target database follows the "Nomenclature for Economic Activities" (NACE) within the European System of National Accounts (ESA95) at first digit level (16 branches), with the exceptions of Agriculture, Hunting and Forestry (NACE: A) and Manufacturing branches (NACE: D), which will be further disaggregated into 2 and 3 sub-branches, respectively (Table 1). Therefore, dimension b will contain 19 single branches as elements, but data will also be collected for the aggregates A and D. Primary factors (f) contains the elements "Labour", "Land", and "Physical capital". Trade partners (w) for each region are either the intra-national markets for trade between regions or the external markets for inter-national trade between countries. A distinction between intra- and extra-European trade is not foreseen. Furthermore, the database does not include bi-lateral flows between regions of a country, or even between regions of the European Union with 27 Member States. Regional consumers (d) are regional governments and households.

A full set of the sub-matrices discussed above will not be available for most of the NUTS2 regions. Intermediate demand matrices (**A** and **Aw**) will be generally available on national scale. Primary factor payments (**F**) may be derived from a) value-added by sectors and b) employment by sectors. Data on regional trade flows (**Aw,Cw,Iw,Ew,E**) are likely to be unavailable for the major part of regions. Final consumption (**C**) and investment demand (**I**) will in the best of cases be available as an aggregate. The initially mentioned satellite accounts will therefore contain the type of information that is not directly part of the SAMs, but will serve to derive and control the required SAM entries. The structure of the satellite accounts and the correspondences with the SAM accounts **S** are illustrated below:

Y ≡

	b	d	1	w	1
w	**M**				
1	**At**	**Ct**	**It**	**Et**	
1	**V**				**Vt**
1	**L**		**P**	**G**	

Where:

Y: Regional Satellite Accounts

M: Total imports by branch or institution and by origin

At: Total intermediate demand in the region ($At_{b'} = \sum_{b} A_{b,b'}$)

Ct: Total final demand by institution (households, governments) ($Ct_d = \sum_{b} C_{b,d}$)

It: Total regional investment ($It = \sum_{b} I_b$)

Et: Total regional exports by destination

V: Gross value-added by branches ($V_b = \sum_{f} \left[F_{f,b} + Ts_{f,b} \right] + Ts_{*,b}$)

Vt: Total regional gross value-added (Gross Regional Product) ($Vt = \sum_{b} V_b$)

L: Employed persons by branch

P: Regional population

G: Net-migration to or from abroad or other regions of the same country

To ensure the consistency of the database to be developed with the ESA95 standards and to facilitate the communication with the Member States' statistical departments, the classification scheme used here follows in general the ESA95 nomenclature (see Tables 1 and 2).

Table 1 Target Branches of the IOTNUTS2 Database

Block	Code	Description
Target sectors	AA01	"Agriculture, hunting and related services"
	AA02	"Forestry, logging and related services"
	B000	"Fishing"
	C000	"Mining and quarrying"
	DA00	"Food products, beverages, and tobacco"
	DF00	"Coke, refined petroleum products, and nuclear fuels"
	DZ00	"Other manufacturing"
	E000	"Electricity, gas and water supply"
	F000	"Construction"
	G000	"Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods"
	H000	"Hotels and restaurants"
	I000	"Transport, storage and communication"
	J000	"Financial intermediation"
	K000	"Real estate, renting and business activities"
	L000	"Public administration and defence; compulsory social security"
	M000	"Education"
	N000	"Health and social work"
	O000	"Other community, social, personal service activities"
	P000	"Activities of households"
Sector aggregates (NACE16 and NACE06)	A000	"Agriculture, hunting and forestry"
	D000	"Manufacturing"
	A2B	"Agriculture, hunting, forestry and fishing"
	C2E	"Total industry (excluding construction)"
	F00	"Construction"
	G2I	"Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication"
	J2K	"Financial intermediation; real estate, renting and business activities"
	L2P	"Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies"

Note: Red fonts indicate that the respective elements are not part of the ESA95 classification scheme

Table 2 Selected branch and regional indicators in the IOTNUTS2 Database

Block	Code	Description
Institutional and factor accounts	d21md31	"Taxes less subsidies on products"
	p2pp	"Total intermediate consumption/Final use at purchasers' prices"
	d1	"Compensation of employees"
	d11	"Wages and salaries"
	d12	"Employers' social contributions"
	d29md39	"Other net taxes on production"
	d29	"Other taxes on production"
	d39	"Other subsidies on production"
	k1	"Consumption of fixed capital"
	b2npb3n	"Operating surplus, net"
	b2gpb3g	"Gross operating surplus and gross mixed income"
	b1g	"Value added at basic prices"

Block	Code	Description
	b1gm	"Gross domestic product at market prices"
	p1	"Output at basic prices"
Labour market	emp	"Total employment"
	sal	"Employees"
	self	"Self-employed"

2. Regional and National Data from ESTAT

The first step in the compilation of this inventory was the review of datasets available from the European Commission, namely national and regional branch accounts from ESTAT.

2.1. Supply, use and input-output tables (NAIO)

Supply and use tables provide information on intermediate and finals uses of products by industries and consumers as well as output of products by industries. A symmetric input-output does distinguish between products and industries; instead, it is arranged as product-by-product or industry-by-industry matrix. It rearranges both supply and use in a single table with identical classification of products (or industries respectively) applied for both rows and columns. The following tables are available from the Eurosat NAIO domain:

- 1500 "Supply Table"
- 1600 "Use Table"
- 1700 "Symmetric Input-Output Table"
- 1800 "Input-output Table for Domestic Output"
- 1900 "Input-Output Table for Imports"

Regarding products, ESTAT applies the CPA P60 classification that delineates 59 products. Industries are classified based on NACE rev.1.1 A60 as reference which distinguishes 59 industries.

Tables 1800 and 1900 permit the distinction of usage by their origin (domestic and foreign). Their availability is show in Table 3. For 2005, these tables are available for 21 Member States, whereas for the year 2000 missing tables can be filled by the tables created by Rueda-Cantuche et al (2009). It has to be noted that Table 3 indicates only that the IOT contain values, but not that they are complete. In some cases, particularly for smaller Member States, some industry accounts are filled with placeholders, indicating that the release of numerical values is subject to data protection regulations.

Table 3 Availability of Member States' National IOT No. 1800 and 1900

	2000	2001	2002	2003	2004	2005	Most recent IOT	Source for most recent IOT
BE000000	X					X	2005	NAIO
BG000000	**X**						**2000**	**RC09**
CZ000000	X					X	2005	NAIO
DK000000	X	X	X	X	X	X	2005	NAIO
DE000000	X	X	X	X	X	X	2005	NAIO
EE000000	X					X	2005	NAIO
IE000000	X					X	2005	NAIO
GR000000	X					X	2005	NAIO

ES000000	X					X	2005	NAIO
FR000000	X	X	X	X	X	X	2005	NAIO
IT000000	X					X	2005	NAIO
CY000000	**X**						**2000**	**RC09**
LV000000	**X**						**2000**	**RC09**
LT000000	X					X	2005	NAIO
LU000000	**X**						**2000**	**RC09**
HU000000	X					X	2005	NAIO
MT000000	**X**						**2000**	**RC09**
NL000000	X	X	X		X	X	2005	NAIO
AT000000	X					X	2005	NAIO
PL000000	X					X	2005	NAIO
PT000000	X					X	2005	NAIO
RO000000	X			X	X	X	2005	NAIO
SI000000	X	X				X	2005	NAIO
SK000000	X					X	2005	NAIO
FI000000	X	X	X	X	X	X	2005	NAIO
SE000000	X					X	2005	NAIO
UK000000	**X**						**2000**	**NAIO**
Total	**27**	**6**	**5**	**6**	**6**	**21**		

Note: Tables developed by Rueda-Cantuche et al (2009) are flagged as RC09 and appear in red fonts. Blue fonts indicate that Tables 1800 and 1900 are available from ESTAT but only for 2000.

2.2. National Accounts - Annual National Accounts (NAMA)

In addition to the national IOT from the NAIO domain, information from the Annual National Accounts (NAMA) was obtained. These datasets provide longer time-series for some indicators like "b1g: gross value-added at basic prices" or "p1: output at basic prices" and can be used to adjust the IOT available only for 2000 for the targeted base-year 2005. The coverage of indicators in the used NAMA datasets is listed in Table 4.

Table 4 Selected Datasets from the NAMA Domain

Code	Description	Covered items
nama_nace06_c	National Accounts by 6 branches - aggregates at current prices	b1g, d1, d11
nama_nace06_e	National Accounts by 6 branches - employment data	emp, sal, self
nama_nace31_c	National Accounts by 31 branches - aggregates at current prices	k1, p1, p2, p5, b1g, d1, d11, b2n_b3n, d29_m_d39
nama_nace31_e	National Accounts by 31 branches - employment data	emp, sal, self
nama_nace60_c	National Accounts by 60 branches - aggregates at current prices	k1, p1, p2, b1g, d1, d11, b2n_b3n, d29_m_d39
nama_nace60_e	National Accounts by 60 branches - employment data	emp, sal, self
nama_aux_pem	Population and employment - Annual data	pop, emp_nc, emp_dc, sal, self
nama_fcs_c	Final consumption aggregates - Current prices	p3, p31_s14_s15, p3_s13, p33, p34
nama_gdp_c	GDP and main components - Current prices	b1gm, p3, p31_s14_s15, p3_s13, p5, p6, p7, b1g, d21_m_d31, d1, b2g_b3g,

		d2_m_d3
Where:		
k1	Consumption of fixed capital	
p1	Output at basic prices	
p2	Total intermediate consumption/Final use at purchasers' prices	
b1g	Value added at basic prices	
d1	Compensation of employees	
d11	Wages and salaries	
b2n_b3n	Operating surplus, net	
d29_m_d39	Other net taxes on production	
emp	"Total employment"	
sal	"Employees"	
self	"Self-employed"	
pop	"Total population"	
p5	"Gross capital formation"	
p3_s14_s15	"Final consumption/expenditure by households and non-profit organisations serving households (NPISH)"	
p3_s13	"Final consumption/expenditure by general government"	

The datasets nama_nace60_e and nama_nace60_c have the highest level of detail for the economic branches and are the only ones that provide information for agriculture and forestry separately as needed for the targeted database. From 27 Member States' datasets, only 14 cover all 19 branches (Table 5). However, closer examination reveals that although figures for "a01: agriculture" and "a02: forestry" may not be available in all cases, the aggregate "a: agriculture and forestry" is available. As these sectors should be distinguished in the final database, additional sources of information will have to be used, e.g. the national IOTs form the NAIO domain.

Table 5 Number of Branches in nama_nace60_c for Gross Value-Added (b1g)

	2000	2001	2002	2003	2004	2005	Branches missing in 2005
BE000000	19	19	19	19	19	19	
BG000000	**16**	**16**	**16**	**16**	**16**	**16**	**a01, a02 (but a), p**
CZ000000	19	19	19	19	19	19	
DK000000	19	19	19	19	19	19	
DE000000	19	19	19	19	19	19	
EE000000	19	19	19	19	19	19	
IE000000	16	16	19	19	19	19	
GR000000	19	19	19	19	19	19	
ES000000	19	19	19	19	19	19	
FR000000	19	19	19	19	19	19	
IT000000	19	19	19	19	19	19	
CY000000	**19**	**19**	**19**	**19**	**19**	**18**	**df (but available in other years)**
LV000000	**18**	**17**	**18**	**18**	**18**	**18**	**p**
LT000000	**18**	**18**	**18**	**18**	**18**	**18**	**df**
LU000000	**17**	**17**	**17**	**17**	**17**	**17**	**b, df**
HU000000	**18**	**18**	**18**	**18**	**18**	**18**	**p**
MT000000	**16**	**16**	**16**	**16**	**18**	**17**	**a02 (but a), df (but avail. in other years)**
NL000000	19	19	19	19	19	19	

AT000000	19	19	19	19	19	19	
PL000000	**16**	**16**	**13**	**17**	**17**	**19**	
PT000000	16	16	16	16	16	19	
RO000000	**18**	**18**	**18**	**18**	**18**	**18**	**p**
SI000000	**19**	**19**	**19**	**19**	**19**	**19**	
SK000000	**18**	**18**	**18**	**18**	**18**	**18**	**p**
FI000000	19	19	19	19	19	19	
SE000000	**19**	**19**	**19**	**19**	**19**	**19**	**da15 (but da)**
UK000000	**17**	**17**	**17**	**17**	**17**	**17**	**a01, a02 (but a)**

Note: Bold fonts indicate missing branches. Red and blue fonts indicate missing IOTs for 2005 (see Table 3)

When comparing Tables 3 and 5, it appears that Bulgaria, Cyprus, Luxemburg, Malta, and UK are the most critical Member States in terms of data availability on national scale. Missing data for "a01: agriculture" in Bulgaria and UK may be filled by using the agriculture and forestry data from the available IOTs to split the aggregate "a". For Luxemburg, it has to be clarified whether "b: Fisheries" and "df: fuel industries" do not exist as branches, or if they are just not reported. In general, it appears that a full set of branch indicators (p1, p2, d1, b1g, emp) can be generated by combining NAIO and NAMA data on a case-by-case basis. Thus, it could be possible to update the missing national IOT for the year 2005.

2.3. National Accounts – Annual Sector Accounts (NASA)

The annual sector accounts (NASA) contain data on flows between sectors, domestic institutions, and the 'rest or the world'. In this respect, NASA is the only source for e.g. factor incomes from abroad, transfers received by households and direct taxes paid by enterprises and households. All information is available in one file (nasa_simplif), which contains a simplified version of the full sequence of monetary flows, which is still sufficient for the national SAMs.

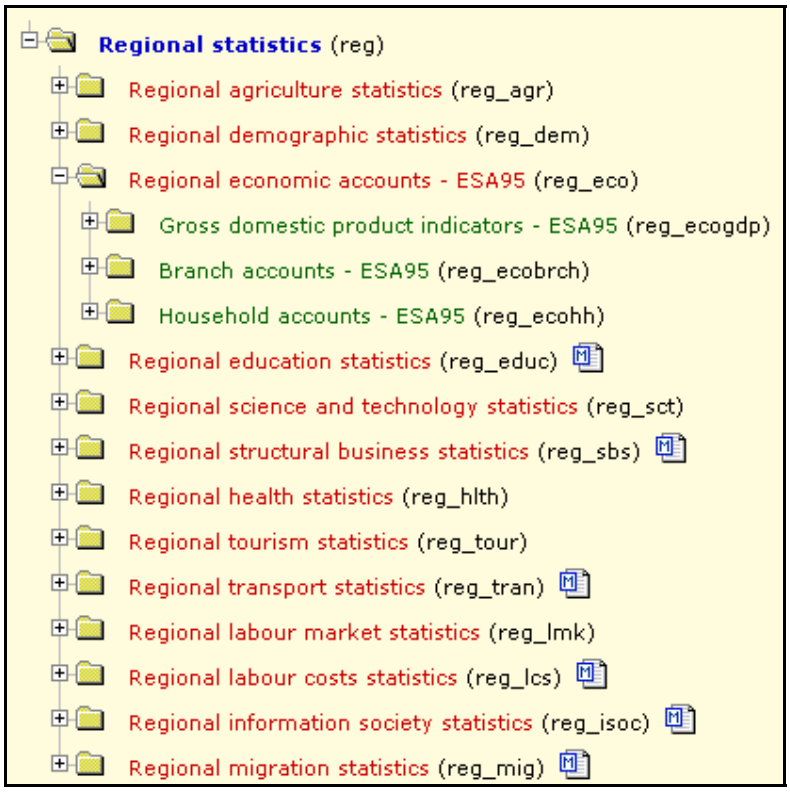
Following Müller et al 2009, the most relevant elements of the NASA datasets for the compilation of national SAMs are the items:

- d4 "Property income"
- d5 "Current taxes on income, wealth, etc"
- d6 "Social contributions and benefits"
- d7 "Other current transfers"
- d8 "Adjustment for the change in net equity of households in pension funds reserves"
- d9 "Capital transfers"
- d1 "Compensation of employees"
- b2g_b3g "Gross operating surplus and gross mixed income"

The taxes and transfer accounts (d4 to d9) between governments, households, and the "Rest of the World" are crucial to determine direct tax rates and governmental expenditures. For 2005, 24 Member States are covered, Cyprus, Luxemburg, and Malta being the exceptions for d5, d6, and d7. For the distribution of factor incomes (d1, b2g_b3g) to either domestic or foreign institutions, all 27 Member States are covered.

2.4. Regional Datasets from ESTAT

The "Regional statistics (reg)" section of ESTAT covers a wide range of indicators (see screenshot below). Particular attention was devoted to those sub-sections which provide information for the structure of economic branches in the NUTS2 regions.



Source: http://epp.ESTAT.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/data/database

2.4.1. Branch Accounts - ESA95 (reg_ecobrch)

The regional branch accounts provide information comparable to the name_nace06 datasets on national scale. The following 6 branches are covered:

- A_B "Agriculture, hunting, forestry and fishing"
- C_E "Total industry (excluding construction)"
- F "Construction"
- G_H_I "Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication"
- J_K "Financial intermediation; real estate, renting and business activities"
- L_TO_P "Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies"

The indicators available for these branches are listed in Table 6. The most important property of the datasets reg_e3vabp95 and reg_e3empl95 is the full coverage of NUTS2 regions in 2005 for the indicators "value-added at basic prices (b1g)" and "total employment (emp)", while Cyprus and Luxemburg are missing in the case of reg_e2rem – which is not relevant as the NUTS2 region of these Member States is equal to the national level. The obvious shortcoming of these datasets is the coarse representation of economic branches.

Table 6 Datasets from the "Branch accounts - ESA95" Domain

Code	Description	Regional coverage in 2005 (NUTS2)	Covered items
reg_e2gfcf	Gross fixed capital formation at NUTS level 2		p5
reg_e2rem	Compensation of employees at NUTS level 2	CY and LU missing in 2005, but other year are available	d1
reg_e2empl95_hw	Employment (in hours worked) at NUTS level 2		emp, sal
reg_e3vabp95	Gross value added at basic prices at NUTS level 3	**100%**	**b1g**
reg_e3empl95	Employment (in persons) at NUTS level 3	**100%**	**emp, sal**
Where:			
b1g	Value added at basic prices		
d1	Compensation of employees		
p5	"Gross capital formation"		
emp	"Total employment"		
sal	"Employees"		

2.4.2. Regional Structural Business Statistics (reg_sbs)

The structural business indicators cover the NACE Rev 1.1 sections C to K, with a breakdown of branches at the 2-digit level in the case of the dataset "Regional data (NUTS 06) (sbs_r_nuts03)". This dataset provides figures on employment (emp) and wages and salaries (d11), among others. The degree of completeness varies across the considered branches. When comparing the aggregate regional employment data by branches with the national total as derived from other ESTAT datasets (namely NAMA), it appears that e.g. employment in "mining and quarrying" (C000) is for 9 out of 27 Member States close to the national figures ("close" means here: within a +/- 10% range). Dull coverage of NUTS2 regions for manufacturing (D000) is reached in 12 Member States with more than one NUTS2 region. Coverage below 100% may indicate that regional markets are dominated by few enterprises, which would cause data protection regulations to apply. Given the severe amount of missing data and the lack of data on "gross value-added" by branches, these datasets have only a limited use, but will be exploited in case national statistic do not provide additional information.

Table 7 Coverage of employment data for selected branches from Regional Structural Business Statistics for 2005 (in percent of national employment as indicated by NAMA for the same year)

	C000	D000	E000	G000	H000	I000

BE000000	**108**	**102**	**101**	**101**	**106**	**94**
BG000000	**95**	**97**	**92**	**99**	**93**	83
CZ000000	**101**	**98**	**99**	**97**	87	**97**
DK000000						
DE000000	88	77	**91**	75	67	**90**
EE000000						
IE000000	79	80	24	**102**	119	75
GR000000	**100**	77	**100**	**99**	**96**	**99**
ES000000	87	84	75	113	**93**	**92**
FR000000	**97**	**99**	**103**	**92**	**91**	**99**
IT000000	85	**92**	88	**95**	**96**	**100**
CY000000	63	117	4	89	88	**92**
LV000000						
LT000000						
LU000000	112	**103**	64	302	301	301
HU000000	39	**91**	88	**99**	83	**93**
MT000000						
NL000000	**91**	**91**	**95**	127	132	126
AT000000	83	**99**	57	**103**	125	71
PL000000	80	86	78	**109**	**92**	84
PT000000	88	**93**	73	**100**	**90**	**96**
RO000000	**98**	83	75	**94**	**93**	81
SI000000						
SK000000	**100**	78	**97**	49	38	74
FI000000	74	**93**	**95**	79	73	87
SE000000	**109**	111	**94**	**108**	**99**	111
UK000000	65	**97**	65	112	153	54

Note: Bold fonts indicate a +/- 10% deviation from national employment data by branch

2.4.3. Regional Agriculture Statistics (reg_agr)

As “a01: agriculture” is neither covered in the branch accounts nor in the structural business statistics, the Economic Accounts for Agriculture (EAA), namely the dataset “Agricultural accounts according to EAA 97 Rev.1.1 (agr_r_accts)” was also evaluated in some detail. The EAA are a satellite account of ESA95, providing complementary information and concepts adapted to the particular nature of the agricultural industry. Despite the definitional differences to ESA95, the availability of “compensation of employees” (d1 in ESA95, 23000 in EAA) and “gross value-added” (b1g in ESA95, 20000 in EAA) was investigated on NUTS2 level, as shown in Table 8. Full coverage is given for 19 Member States in 2005, coverage above zero and below 100% can be observed for the Czech Republic, Italy has full coverage for b1g, but zero coverage for d1. The remaining countries have no entries.

Table 8 Availability of data on agricultural gross value-added and compensation of employees (ESA95: b1g, d1; EAA: 20000, 23000) (in percent of total NUTS2 regions, data for compensation of employees in brackets)

	2000	2001	2002	2003	2004	2005
BE000000						

BG000000					100 (100)	100 (100)
CZ000000	**88 (88)**	**88 (88)**	**88 (88)**	**88 (88)**	**88 (88)**	**88 (88)**
DK000000						
DE000000	95 (95)	95 (95)	95 (95)	95 (95)	100 (100)	100 (100)
EE000000	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**
IE000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
GR000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
ES000000	**89 (0)**					
FR000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
IT000000	**100 (0)**	**100 (0)**	**100 (0)**	**100 (0)**	**100 (0)**	**100 (0)**
CY000000	**100 (0)**	**100 (0)**	**100 (0)**	**100 (100)**	**100 (100)**	**100 (100)**
LV000000	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**
LT000000	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**
LU000000	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**
HU000000		100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
MT000000	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**	**100 (100)**
NL000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
AT000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
PL000000						
PT000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
RO000000						
SI000000						
SK000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
FI000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
SE000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
UK000000	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)

Note: Bold red fonts indicate that either b1g or d1 is missing in some or all NUTS2 regions in 2005. Data for d1 (compensation of employees) in brackets.

2.5. Summary of ESTAT Datasets

Concerning national IOTs in the desired format (i.e. NAIO Tables 1800 and 1900), 21 Member States are covered for the year 2005. As a full set of IOTs is available for the year 2000 (either from ESTAT or Rueda-Cantuche et al 2009) the first step will be to update the IOTs for the missing 6 Member States. This can be done by using a completed NAMA dataset for the relevant branch indicators. The distribution of the national income across the domestic institutions and the "Rest of the World" can implemented in the national SAMs by combining updated IOTs and NASA datasets. Critical Members States are Luxemburg, Cyprus, and Malta, as neither NAIO, nor NAMA, nor NASA datasets are fully available. With regard to the regional database, the problem is limited as these three Member States consist of only one NUTS2 region each.

The regional branch accounts have, despite their coarse representation of economic branches, full coverage for all NUTS2 regions and are therefore a most valuable asset for the subsequent compilation steps. A breakdown to 19 branches could be achieved by using the NAMA and NAIO datasets on national scale for the indicators d1, b1g, and emp. Structural Business Statistics and EAA data will then be used as supplements if available.

2.6. Derivation of Core Regional Datasets based on ESTAT

The ESTAT datasets represent the fundamental database for all subsequent steps as they provide full coverage of some regional branch accounts indicators, even if with a very coarse representation of branches. Statistical information from Member States will be used as

exhaustively as possible, but in case no additional information becomes available, ESTAT datasets will be used as fall-back option. Based on the evaluation of the national and regional datasets, the following steps are proposed:

1. Complete the NAMA series (namely nama_nace60_c, nama_nace60_e) for the 19 target sectors for the period between 2000 and 2005. This will be done by using the shares of the A60 aggregation level with practically 59 branches in the A16 aggregation level with 16 branches. In case national IOTs are not available, than the shares from the year 2000 will be used. The completion procedure could be thought of as follows:

a. Use the NAMA indicators if available to create a time series on aggregate intermediate demand, value-added components (b1g, d1), and employment (emp, sal):

$$AVLN^{MS}_{i,b19,t} = \sum_{A60}\left[G^{60\_19}_{A60,b19} \cdot NAMA^{MS}_{i,A60,t} \right] \forall\, NAMA^{MS}_{i,A60,t} \neq 0$$

Where:

MS:	Index for Member States
i:	Index for Indicators as available from the NAMA datasets (b1g, d1, p1, p2, emp, sal, …)
b19:	Target branches for the IOTNUTS2 database
t:	Time (years 2000 to 2005)
A60:	NACE with 59 branches
AVLN:	Indicators at national level (A: intermediate demand, V: value-added components, L: labour indicators)
G^{60_19}:	Aggregator matrix between 59 NACE and 19 target branches[2]

b. In case the NAMA does not provide information on the A60 but on the A16 aggregation level (which is the case for all Member States, see Table 5), then the respective NAIO tables 1800 and 1900 will be used to create this breakdown by calculating the share of the A60 branches in their respective aggregates at A16 level.

$$AVLN^{MS}_{i,b19,t} = \sum_{A60}\left[G^{60\_19}_{A60,b19} \cdot \sum_{A16}\left[G^{60\_16}_{A60,A16} \cdot NAMA^{MS}_{i,A16,t} \right] \cdot \frac{NAIO^{MS}_{i,A60,t}}{\sum_{A16}\left(G^{60\_16}_{A60,A16} \cdot \sum_{A60}\left[G^{60\_16}_{A60,A16} \cdot NAIO^{MS}_{i,A60,t} \right] \right)} \right]$$

$$\forall\, NAMA^{MS}_{i,A60,t} = 0 \wedge NAIO^{MS}_{i,A60,t} \neq 0 \wedge NAMA^{MS}_{i,A16,t} \neq 0$$

Where:

G60_16:	Aggregator matrix between 16 and 60 branches[3]
NAIO:	Indicators form national IOTs

c. Finally, if neither NAMA at A60 level, nor NAIO data are available, the IOTs from Rueda-Cantuche et al (2009) will be used, which provide a full coverage of the Member States in 2000:

[2] For the use of aggregator matrices, please see also section 4

[3] For the use of aggregator matrices, please see also section 4

$$AVLN^{MS}_{i,b19,t} = \sum_{A60}\left[G^{60\_19}_{A60,b19} \cdot \sum_{A16}\left[G^{60\_16}_{A60,A16} \cdot NAMA^{MS}_{i,A16,t} \right] \cdot \frac{IORC^{MS}_{i,A60}}{\sum_{A16}\left(G^{60\_16}_{A60,A16} \cdot \sum_{A60}\left[G^{60\_16}_{A60,A16} \cdot IORC^{MS}_{i,A60} \right] \right)} \right]$$

$$\forall\, NAMA^{MS}_{i,A60,t} = 0 \wedge NAIO^{MS}_{i,A60,t} = 0 \wedge NAMA^{MS}_{i,A16,t} \neq 0$$

Where:

IORC: IOTs from Rueda-Cantuche et al (2009)

2. Update the national IOTs to a common base-year 2005. This step can already be a test-case for the compilation procedures applied at the regional level: A limited set of yearly available branch indicators (gross value-added, compensation of employees, employment) will be combined with SAM-coefficients from another year or regional level. In the simplest case, an iterative procedure like RAS could be used to update the national IOTs to the new row- and column-totals for the branch accounts (p1). As such a procedure would not take advantage of other available information (b1g, d1, emp, etc.), a more refined compilation and balancing procedure will have to be developed.

3. Create a set of branch indicators at regional scale. This step could build on the fact that the regional branch accounts cover 100% of the NUTS2 regions, although with a coarse representation of six branches (A6, e.g. reg_e3vabp95, see Table 5). Similar to the usage of shares of A60 branches in their A16 aggregates on national scale, the shares of the 19 target branches (b19) in the A6 aggregates on national scale could be combined with the regional A6 data:

$$AVLR^{MS,R}_{i,b19,t} = \sum_{A6}\left[G^{19\_6}_{b19,A6} \cdot REBR^{MS,R}_{i,A6,t} \right] \cdot \frac{AVLN^{MS}_{i,b19,t}}{\sum_{A6}\left(G^{19\_6}_{b19,A6} \cdot \sum_{b19}\left[G^{19\_6}_{b19,A6} \cdot AVLN^{MS}_{i,b19,t} \right] \right)}$$

Where:

R: Index for NUTS2 regions

AVLR: Indicators at regional level (A: intermediate demand, V: value-added components, L: labour indicators)

REBR: Regional branch accounts with six branches

4. Compile a set of regional IOTs for 2005 based on national IOTs and regional branch indicators. Similar to updating the national IOTs, a simple RAS procedure could be thought of as a starting point to compile a prior set of regional IOTs. Again, this would not exploit the full range of available information and a more elaborate procedure will have to be developed

The main conclusion from the evaluation of the ESTAT datasets is that it is in general possible to derive regional IOTs. Also, the regional branch indicators AVLR as discussed under step 3 may serve as a benchmark for the evaluation of informational gain when considering the purchase of regional data from the Member States' statistical institutions

3. National Statistical Organisations

After screening the datasets available from ESTAT, the second step was to browse the homepages of the national statistical institutes of the Member States. On average, replies

came within two working days, but provided in many cases no additional information as could be found on the homepage itself. In general, the national classification schemes follow closely – but not fully – the ESA95 system.

3.1. Regional branch accounts from EU Member States' Statistical Organisations

To evaluate and compare the gain of information from contacting the Member States statistical institutions, the number of additional data points across all branches was summarized in an indicator. This "informational gain" indicator was constructed based on the following considerations:

- Number of branches in IOTNUTS2: 19
- Number of branch aggregates in ESTAT's regional branch accounts: 6
- Number of IOTNUTS2 branches in A6:

NACE A6 Name	Code	No. of IOTSNUTS2 branches in A6 (na6)
"Agriculture, hunting, forestry and fishing"	A2B	3
"Total industry (excluding construction)"	C2E	5
"Construction"	F00	1
"Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication"	G2I	3
"Financial intermediation; real estate, renting and business activities"	J2K	2
"Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies"	L2P	5

Under the assumption that the regional datasets are consistent with the ESTAT branch accounts, one would need (na6-1) additional data points to construct the full branch accounts (e.g. if data on a01 and a02 are available, then the remaining entry for b could be obtained residually, provided that a01+a02 < A2B).
The "informational gain" indicator over all branches (TIG) was then constructed as follows:

$$TIG_i^{MS,R} = \sum_{A6}\left(\sum_{b19}\left[G_{b19,A6}^{19\_6}\cdot\delta\left(\sum_{AR}G_{i,b19,AR}^{19\_AR}\right)\right]-1\right)\Bigg/\sum_{A6}\left(\sum_{b19}\left[G_{b19,A6}^{19\_6}\right]-1\right)$$

Where:

TIG: Informational gain indicator for all branches
AR: Branches of the regional datasets
G^{19_AR}: Aggregator from regional to IOTNUTS2 branches for the respective datasets i
G^{19_6}: Aggregator from 19 to 6 branches

$$\delta\left(\sum_{AR}G_{i,b19,AR}^{19\_AR}\right)=\begin{cases}1 & if\ \sum_{AR}G_{i,b19,AR}^{19\_AR}>0\\ 0 & if\ \sum_{AR}G_{i,b19,AR}^{19\_AR}=0\end{cases}$$

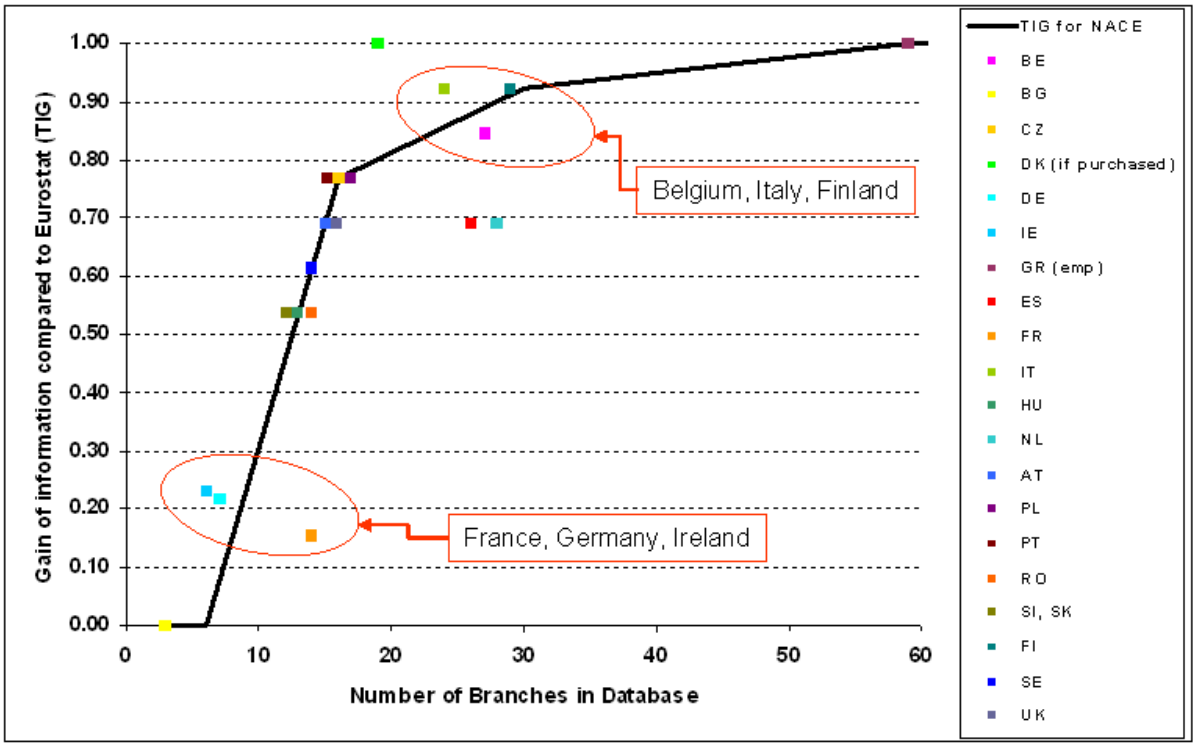
The aggregator matrix G^{19_AR} is constructed by assigning ones to branches that can be mapped in a many-to-one way to the 19 branches, and zeroes otherwise. The indicator can range between 0 (no informational gain compared to ESTAT data) and 1, which indicates a full coverage of branches for the respective indicator. It has to be noted that TIG measures only potential gain of information as it counts only the number of usable branch classifications in the national datasets.

A second purpose of screening the data supply of the national statistical organisations was to retrieve information on the availability of regional IOTs. In this respect, the situation proved to be modestly satisfying as survey-based regional IOTs could only be obtained for 12 Comunidades Autonomas (NUTS2) of Spain and the NUTS1 regions Scotland and Baden-Württemberg. The IOTs obtained for Finland and those potentially available for Austria, Italy, and Poland, are mainly based on non-survey methods.
Regarding the availability of regional branch accounts that provide additional information to the datasets obtained from ESTAT, screening the national statistical departments' supply resulted in a rather mixed picture. For freely available datasets, the informational gain as measured by the TIG-Indicator introduced in the beginning of section 3 never reached a level above 0.92 (see Figure 1).

Table 9 Table 10 – **Informational gain compared to ESTAT data – only Member States with more than 1 NUTS2 region, for b1g, d1 and emp.**

	N. of branches	b1g	d1	emp
BE000000	27	0.85	0.85	0.85
BG000000	3	0.00		
CZ000000	16	0.77		0.77
DK000000	19	1.00	1.00	1.00
DE000000	7	0.00 (0.77)	0.00 (0.77)	0.00 (0.77)
IE000000	6			
GR000000	59			1.00
ES000000	26	0.69	0.69	0.69
FR000000	14	0.15	0.69	0.69
IT000000	24	0.92	0.92	0.92
HU000000	13	0.54		
NL000000	28	0.69	0.69	0.69
AT000000	15	0.69	0.69	0.69
PL000000	16	0.77	0.77	0.54
PT000000	15	0.77	0.77	0.77
RO000000	13	0.54		0.69
SI000000	12	0.54		0.54
SK000000	12	0.54	0.54	0.54
FI000000	29	0.92	0.92	0.92
SE000000	14	0.62	0.62	0.62
UK000000	15	0.69	0.69	

Figure 1 Number of branches in regional accounts for b1g, and informational gain compared to ESTAT data – only Member States with more than 1 NUTS2 region



Notes:

- The black line indicates the gain of information for the different levels of NACE (A6, A16, A30, A60)
- Scale of informational gain indicator: 0: Same information as obtained from ESTAT; 1: Full coverage of all IOTNUTS2 branches

Source: Own presentation

The least informational gain was observed for Bulgaria (no gain) and France, Germany, and Ireland, for which the gain ranged below 0.3 points. One reason is the aggregate representation of agriculture, forestry, and fishery in the regional branch accounts. This is the case for 12 Member States (see Table 9). As a disaggregation of these branches is crucial, the availability of other indicators that could permit a split has been evaluated and summarized in Table 9.

3.2. Additional datasets

The focus of the survey of the EU Member States' statistical departments was on the availability of regional branch account indicators like gross value-added and employment as these indicators represent the corner points of the database to be developed. Datasets on population and migration are important indicators for the mobility of labour within and across the EU Member States. Population figures are available for all NUTS2 regions from both, ESTAT and national statistical organisation (NSO), while migration figures are not as easily obtainable. Table 32 lists the availability of migration figures (either net migration or

immigration/emigration and either by sex, or age group, or total). In all cases, the NSO provide more information than ESTAT, with the exceptions of Denmark, Greece, France, Malta, and Portugal.

Table 10 Availability of migration data on NUTS2 level

	ESTAT	NSO
BE000000		X
BG000000	X	X
CZ000000	X	X
DK000000		
DE000000		X
EE000000		**X**
IE000000		X
GR000000		
ES000000	X	X
FR000000		
IT000000	X	X
CY000000		**X**
LV000000		**X**
LT000000		**X**
LU000000		**X**
HU000000	X	X
MT000000		
NL000000	X	X
AT000000	X	X
PL000000	X	X
PT000000		
RO000000	X	X
SI000000	X	X
SK000000	X	X
FI000000	X	X
SE000000	X	
UK000000		X

Note: Bold black fonts indicate that a Member State consists of one NUTS2 region. Red bold fonts indicate that neither ESTAT nor the Member State's statistical organisation (NSO) provide migration data

Aggregate figures on agriculture and forestry (A) can be split based on agricultural data from either EAA (for d1 and b1g) or CAPRI data (for b1g). If the datasets are broadly consistent, forestry can be calculated as a non-negative residual. Procedures for negative residual values will have to be developed on a case-by-case basis. Splitting aggregate values for agriculture, forestry, and fishery (A_B) requires at least one indicator for either forestry or fishery in addition to the EAA/CAPRI figures. Table 34 shows the availability of forestry area data from ESTAT (black) or national statistical organisations (blue). Using these figures to derive average national employment and gross value-added coefficients per hectare of forestry area would supply the needed information, but is likely to yield negative values for fishery if calculated residually. Again, the procedures will have to be developed on a case-by-case

basis. The most critical Member State in terms of data availability is Bulgaria, for which no additional datasets could be found.

Table 11 Data availability for agriculture and forestry in 2005

	Regional branch accounts						
	A_B	A	A01, A02	EAA/ CAPRI	Forest area	A01 or (A+EAA) or (A_B+EAA +Forest Area)	Comment
BE000000	X			X	X	X	
BG000000	X			X		0	No additional information available
CZ000000		X		X	X	X	Forestry area available for 2004
DK000000			X	X		X	If purchased
DE000000	X			X	X	X	Forestry area available from NSO
EE000000			**X**	**X**		X	
IE000000		X	(X)	X		X	Additional data obtained from NSO
GR000000	X			X		0	Employment data available for A01, A02
ES000000	X			X		0	13 Regional IOTs
FR000000	X			X	X	X	
IT000000		X		X	X	X	
CY000000			**X**	**X**		X	
LV000000			**X**	**X**	**X**	X	
LT000000			**X**	**X**	**X**	X	
LU000000			**X**	**X**	**X**	X	
HU000000	X			X	X	X	
MT000000			**X**	**X**		X	
NL000000	X			X	X	X	
AT000000	X			X	X	X	
PL000000		X		X	X	X	
PT000000		X		X	X	X	
RO000000		X		X	X	X	
SI000000	X			X		0	Potentially available from ongoing projects
SK000000	X			X	X	X	
FI000000			X	X	X	X	
SE000000	X			X	X	X	Forestry area available from NSO
UK000000		X		X		X	
Total	**12**	**7**	**8**	**27**	**16**	**23**	

Notes:

Green: EAA data not available but gross value-added from CAPRI (GVAB)
Blue: Forestry area available from national statistics
Pink: Additional data available for previous year or from other source
Red: Splitting of A_B not possible based on regional data

3.3. Previous SAM Projects

Numerous research projects have developed regional IOT-based models on a regional scale for single EU Member States. The following sections provide an overview of projects on multi-country regional CGEs or other IOT-based models. For this study, the relevant publications of these projects were screened with focus on the potential use of the respective

databases for the IOTNUTS2 project. Particular attention was paid to the classification of economic branches as this was identified as a potential major constraint for the usability.

3.3.1. IASON

Within the FP5 project IASON (Integrated Appraisal of Spatial economic and Network effects of transport investments and policies, project site: http://www.iasonproject.eu/default.htm), a very comprehensive database for altogether 1342 NUTS3 regions in EU Member States and other European Countries was developed for the base year 1997 (Bröcker et al 2002). The database was built on national IOTs and miscellaneous regional datasets, featuring the 6 NACE Rev. 1.1 branch aggregates as used in the ESTAT regional branch accounts discussed in section 2.3.1. Although the project documentation appears to be a valuable source for data collection and compilation methods, the database as such is outdated and too coarse for the purposes of the IOTNUTS2 project: The IOTs developed in the course of the IASON project cannot be used as control totals (outdated and estimated), nor as benchmark for estimation procedures as they are themselves estimated, nor to construct a-priori information for branches missing in the regional accounts (too coarse).

3.3.2. DREAM

The dream model as discussed in Jean and Laborde (2004) is a regional CGE model operating on NUTS1 scale for EU25 with 1997 as base year. It is mainly built on national IOTs from the GTAP database, the regional breakdown was carried out by using the REGIO database from ESTAT and supplementary information. The breakdown of branches is in general not compatible with the requirements of the IOTNUTS2 project (Table 12). The branch breakdown and the fact that it is an estimated database limit severely the potential use for the IOTNUTS2 project.

Table 12 Branch classifications of DREAM

GTAP Code and Description	DREAM Code	IOTNUTS2 Code	Corresponding IOTNUTS2 branch: Description
b_t - Beverages and tobacco products, cmt - Meat:cattle,sheep,goats,horse, mil - Dairy products, ofd - Food productsnec, omt - Meat products nec, pcr - Processed rice, sgr - Sugar, vol -Vegetable oils and fats	Agri_Ind	AA01	"Agriculture, hunting and related services"
		DA15	"Food products and beverages"
ctl - Cattle,sheep,goats,horses, oap - Animal products nec, rmk - Rawmilk, wol - Wool, silk-worm cocoons	Animals	AA01	"Agriculture, hunting and related services"
gro - Cereal grains nec, pdr - Paddy rice, wht - Wheat	Cereals	AA01	"Agriculture, hunting and related services"
c_b - Sugar cane, sugar beet, for - Forestry, ocr - Crops nec, osd - Oilseeds, pfb - Plant-based fibers, v_f - Vegetables, fruit, nuts	OthVeg	AA01	"Agriculture, hunting and related services"
		AA02	"Forestry, logging and related services"
fsh - Fishing	Fishing	B000	"Fishing"
col - Coal, gas - Gas, oil - Oil, omn - Minerals nec	Extraction	C000	"Mining and quarrying"
crp - Chemical,rubber,plastic prods, p_c - Petroleum, coal products	Chim_Ind	DF23	"Coke, refined petroleum products and nuclear fuels"
		DZ00	"Other manufacturing"

isr - Insurance, ofi - Financial services nec	Finance	J000	"Financial intermediation"
ele - Electronic equipment, ome - Machinery and equipment nec	Mach_Ind	DZ00	"Other manufacturing"
fmp - Metal products, i_s - Ferrous metals, nfm - Metals nec, nmm -Mineral products nec	Metal_Ind	DZ00	"Other manufacturing"
omf - Manufactures nec	OthInd	DZ00	"Other manufacturing"
ppp - Paper products, publishing	Paper_Ind	DZ00	"Other manufacturing"
lea - Leather products, tex - Textiles, wap - Wearing apparel	Tex_Ind	DZ00	"Other manufacturing"
mvh - Motor vehicles and parts, otn - Transport equipment nec	Tran_Ind	DZ00	"Other manufacturing"
lum - Wood products	Wood_Ind	DZ00	"Other manufacturing"
cns - Construction, dwe - Dwellings, ely - Electricity, gdt - Gasmanufacture, distribution, obs - Business services nec, osg -PubAdmin/Defence/Health/Educat, ros - Recreation and other services, wtr - Water	**Services**	**E000**	**"Electricity, gas and water supply"**
		F000	**"Construction"**
		H000	**"Hotels and restaurants"**
		K000	**"Real estate, renting and business activities"**
		L000	**"Public administration and defence; compulsory social security"**
		M000	**"Education"**
		N000	**"Health and social work"**
		O000	**"Other community, social, personal service activities"**
		P000	**"Activities of households"**
trd - Trade	Trade	G000	"Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods"
atp - Air transport, cmn - Communication, otp - Transport nec, wtp -Sea transport	TransCom	I000	"Transport, storage and communication"
			Overall gain of information (TIG)
			0.15

Note: Bold red fonts indicate that a many-to-one mapping from national classification to IOTNUTS2 branches is not possible.

3.3.3. REAP&REEIO

The "Sustainable Consumption and Production Network" (SCPnet, http://www.scpnet.org.uk/index.html) is a partnership network dedicated to promoting the philosophy of sustainable consumption and production (SCP) at a sub-national level. The SCPnet is a reference point for Regional Development Agencies, Regional Assemblies, Government Offices for the Regions, Regional Observatories and the regional offices of the Environment Agency. It maintains two IOT-based models on their homepage:

The Regional Economy Environment Input Output Model (REEIO, http://www.scpnet.org.uk/reeio.html) is a decision support tool used to assess the environmental implications of production within a region. REEIO links economic activity in 42 industrial branches (most relevant for IOTNUTS2: Agriculture etc: a01,a02, b05; Food, Drink & Tob.: da15, da16; Manuf. Fuels: df23) to environmental components.

The Resource and Energy Analysis Program (REAP, http://www.scpnet.org.uk/reap.html) is designed for analyses of environmental pressures associated with consumption within a region. It can operate at regional and national levels. The production side of the economy is broken down according to the "UK Standard Industrial Classification of Economic Activities" into 123 sectors in the UK, which are compatible with the Standard Industrial Classification

(SIC92) and NACE Rev. 1. REAP was developed by the Stockholm Environment Institute (http://sei-international.org/), REEIO by Cambridge Econometrics (http://www.camecon.com/Home.aspx). The two models are designed as complementary as REAP models the impact of regional consumption and REEIO models the impacts of economic production.

3.3.4. REAPBALK

The European Commission's DG Research funded from October 2001 onwards a project named "Rural Employment and Agricultural Perspective in the Balkan Applicant Countries" (REAPBALK, Project ID: QLRT-2000-01608). The purpose of this project was to analyse the medium-term agricultural and employment perspectives in selected rural areas of four Balkan countries - Bulgaria, Croatia, Romania, and Slovenia. Greece was also included in the study to provide grounds for suggesting the likely path of employment development post-accession. I-O tables have been constructed for each region based on existing regional tables or, where these did not exist, on estimated regional I-O tables derived from the national ones. (http://ec.europa.eu/research/agriculture/projects/qlrt_2000_01608_en.htm)

3.3.5. ESPON

The ESPON 2013 Programme (European Observation Network for Territorial Development and Cohesion) offers on its homepage access to a database that provides regional statistics on various scales (http://www.espon.eu/main/Menu_ScientificTools/Menu_ESPON2013Database/). This database is compiled of datasets from ESTAT and ESPON projects, covering the entire European Union plus Switzerland, Norway, Iceland and Liechtenstein. Free access to the ESPON 2013 Database is granted upon acceptance of the pre-defined terms and conditions of use, which are not indicated further on the homepage. The database is announced to be available form March 2010, but by 17.03.2010 no possibilities to access the database was provided. An inventory (downloaded from an older version of the homepage) indicated that data relevant for the IOTNUTS2 project do not exceed the coverage of the ESTAT datasets. Once access to the database is published, a more detailed investigation will be carried out.

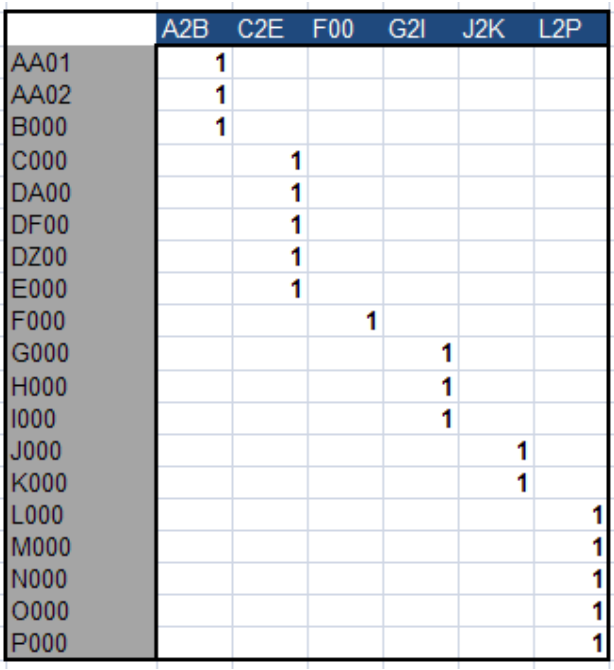
4. Performance of Non-survey Methods

4.1. Combination and completion of datasets

After screening the availability of datasets from various sources, the decision was made to use ESTAT branch accounts as benchmark, mainly because of the completeness of the regional branch accounts ($RAMA^{ESTAT}$). The next question is how the available data can be combined in the most efficient manner. In general, there are three types of data available: First, the mentioned ESTAT regional branch accounts ($RAMA^{ESTAT}$), the national branch account tables ($NAMA^{ESTAT}$) and the branch accounts from the national statistical organisations ($RAMA^{NSO}$). It appears that the NSO does not provide full information for the B19 branch classification but rather for an A16 scheme where Agriculture and Forestry (A000) and Manufacturing (D000) are combined (red shaded areas in Figure 2: $RAMA^{NSO}$ and $NAMA^{ESTAT}$).

So, in addition to three types of data, the information for three branch classifications (A6, A16, and B19) needs to be combined in the most exhaustive manner. This is basically done by defining aggregator matrices (G) between the branch classifications and the usage of shares of b19 and A16 entries in the respective A6 (and A16) entries. The following example illustrates the procedure for the simplified case without the A16 classification. As the described procedures rely on the definition of appropriate aggregator matrices, an example is given in figure 2:

Figure 2 Aggregator Matrix between b19 and A6 Classifications (G19_6)



	A2B	C2E	F00	G2I	J2K	L2P
AA01	1					
AA02	1					
B000	1					
C000		1				
DA00		1				
DF00		1				
DZ00		1				
E000		1				
F000			1			
G000				1		
H000				1		
I000				1		
J000					1	
K000					1	
L000						1
M000						1
N000						1
O000						1
P000						1

Using this aggregator matrix, the national shares of b19 entries in A6 are calculated for the national datasets (NASH):

$$NASH_{b19,i,t}^{ESTAT,MS} = \frac{NAMA_{b19,i,t}^{MS}}{\sum_{A6}\left(G_{b19,A6}^{19\_6} \cdot \sum_{b19}\left[G_{b19,A6}^{19\_6} \cdot NAMA_{b19,i,t}^{MS}\right]\right)}$$

Similarly, the shares of the regional datasets in A6 are calculated (RASH):

$$RASH_{b19,i,t}^{NSO,R} = \frac{RAMA_{b19,i,t}^{NSO,R}}{\sum_{A6}\left(G_{b19,A6}^{19\_6} \cdot \sum_{b19}\left[G_{b19,A6}^{19\_6} \cdot RAMA_{b19,i,t}^{NSO,R}\right]\right)}$$

Combination into a core-account dataset Y is then achieved by a simple expansion of the $RAMA^{ESTAT}$ datasets and multiplication with the shares, depending on the availability of datasets from NSO:

$$Y_{b19,i,t}^{MS,R} = \begin{cases} \sum_{A6}\left[G_{b19,A6}^{19\_6} \cdot RAMA_{A6,i,t}^{ESTAT,R}\right] \cdot RASH_{b19,i,t}^{NSO,R} & \forall RAMA_{b19,i,t}^{NSO,R} \neq 0 \\ \sum_{A6}\left[G_{b19,A6}^{19\_6} \cdot RAMA_{A6,i,t}^{ESTAT,R}\right] \cdot NASH_{b19,i,t}^{ESTAT} & \forall RAMA_{b19,i,t}^{NSO,R} = 0 \end{cases}$$

Where:

Y: Indicators at regional level (A: intermediate demand, V: value-added, L: labour indicators)

RAMA: Regional branch accounts (with six branches)

NAMA: National branch accounts (with 19 branches)

G^{19-6}: Aggregator matrix between A6 NACE and B19 target branches

i: Index for Indicators as available from the NAMA datasets (b1g, d1, p1, p2, emp, …)

b19: Target branches for the IOTNUTS2 database

A6: NACE with 6 branch aggregates

R: Index for NUTS2 regions

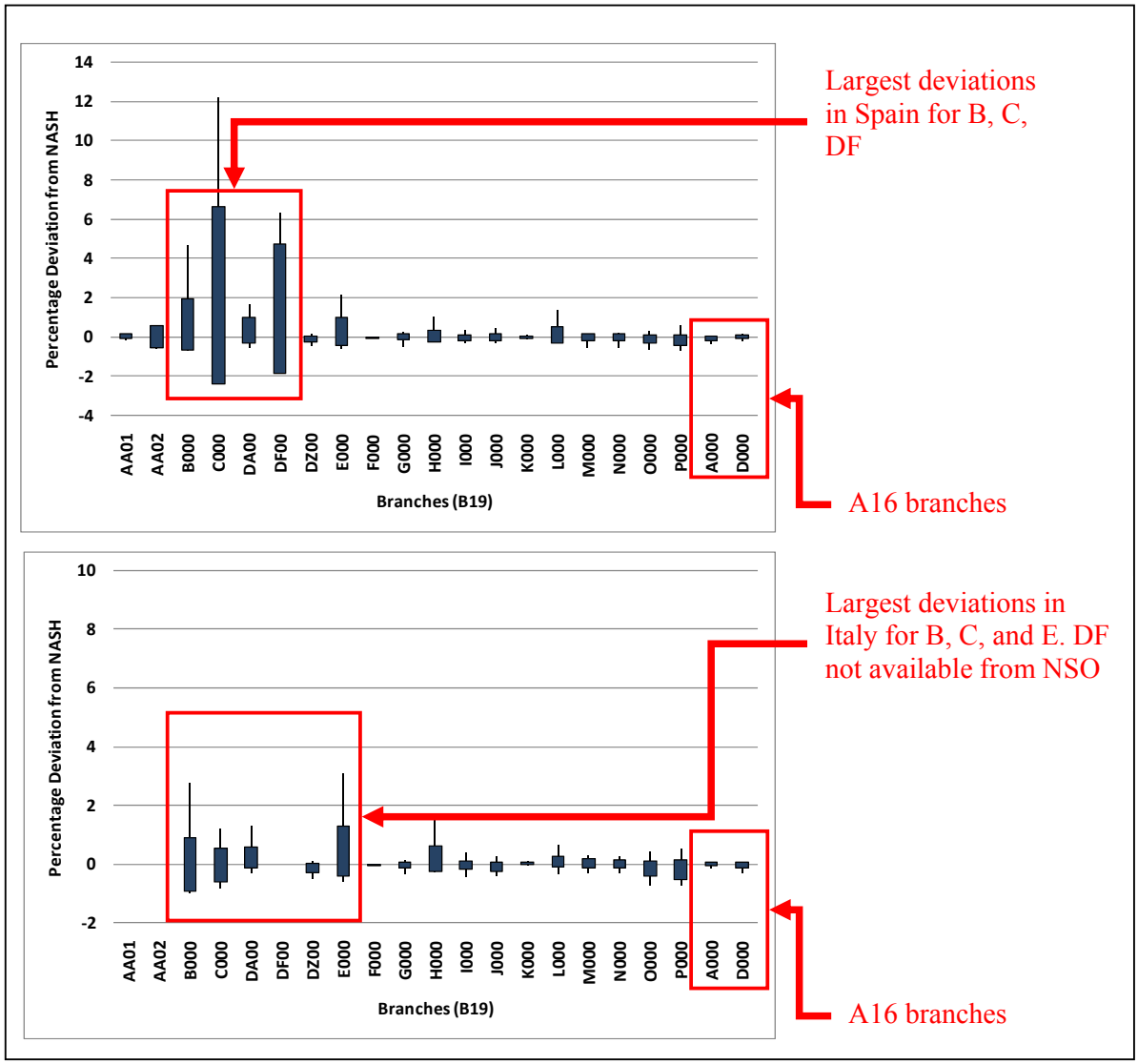
MS: Index for Members States

t: Time (years 2000 to 2005)

In case A16 information is used, the procedure involves more if-then conditions and calculation steps, but is in principle similar.

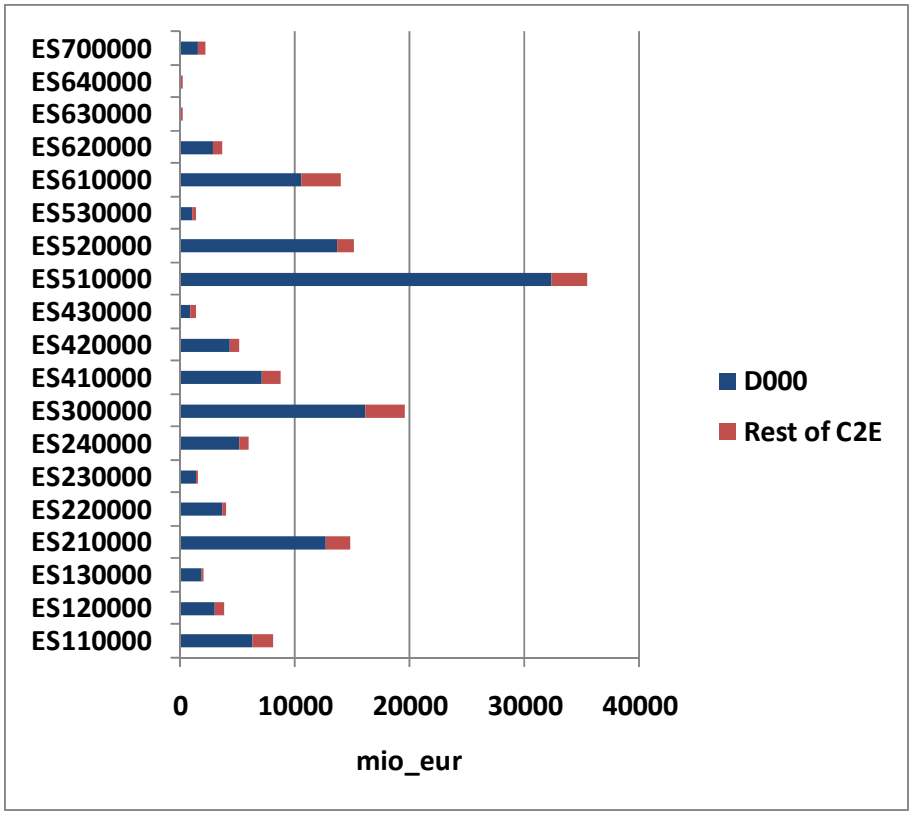
When comparing the deviation between NASH and RASH figures for Italy and Spain (Figure 3), it appears that in most cases the composition of the regional branch structure within the A6 aggregates does not deviate extensively from the national shares.
Notable exceptions for Spain are the sub-sectors of A2B (particularly B: Fisheries) and C2E (Particularly C: Mining and Quarrying as well as DF: Fuel industry). For Italy, the largest deviation can be observed for B: Fisheries, C: Mining and Quarrying and E: Electricity production. Data for DF: Fuel industry is not available from NSO. These observation is understandably as primary production like Fisheries and Mining and energy sectors like Fuel industries and Electricity tend to be concentrated in some regions (close to the sea, natural resources), while they may not appear in others. As the main objective of this project is the compilation of a database that characterises the regional economies appropriately and pays attention to regional particularities, this observation causes some concern.
In general, it can be stated that the shares of D (Manufacturing) and A (Agriculture and Forestry) do not deviate to the same extent form the national shares, so it would be possible to derive at least at the A16 level a more reliable set of core accounts. In the case of the share of D in C2E, it can also be stated that D (at least in Spain) has a particularly high value (see Figure 4), such that biased estimates for C and E will not have a huge importance for the overall picture of the regional economies.

Figure 3 Deviation between National and Regional Shares in Spain and Italy



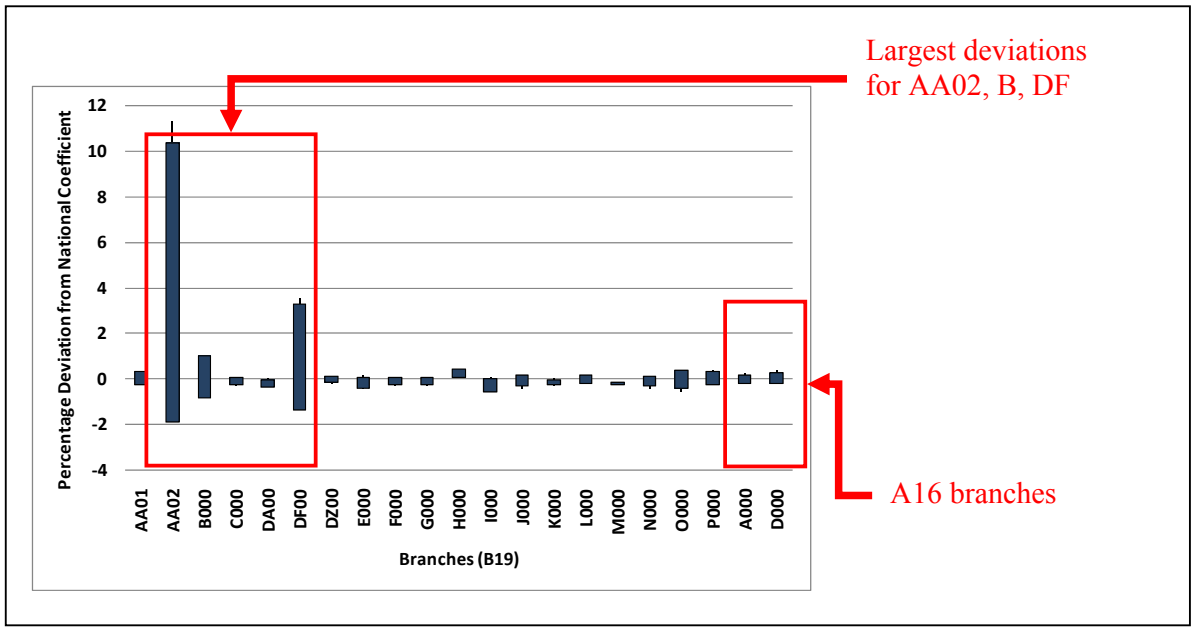
In summary, it may be concluded that the proposed derivation of regional core accounts for employment and gross value-added based on ESTAT regional branch accounts and national shares of b19 in A6 branch aggregates will produce acceptable results when considering the general characteristics of the regional economies (dominated by primary production or manufacturing). Nevertheless, detailed information on the share of fisheries and mining and quarrying industries will significantly improve the picture of the regional economies.

Figure 4 Composition of C2E in Spanish Regions



So far, only the regional account indicators gross value-added, employment, and compensation of employees has been discussed as these indicators are available for all NUTS2 regions from ESTAT. To create a complete picture of the regional core accounts and to derive ultimately regional IOTs, it will be necessary to derive a number of additional core indicators, such as gross output by branch, total intermediate demand by branch, or taxes on activities paid by branch. These figures are available in some cases, but it will be necessary to formulate a default strategy in case the NSO do not provide the needed information. Based on the fact that a full set of gross value-added and employment data can be compiled (as described in the previous section), a logical procedure would be to use again national coefficients for the completion of the core accounts. The assumption is that per unit of gross value-added generated by economic branch, a similar amount of intermediate input would be used and a similar amount of gross output would be produced as on national scale. As before, the outcomes of these computations are compared with the recorded figures from the NSO Spain in figures 4. With respect to intermediate demand (p2), figure 5 reveals that the largest deviations occur for forestry, fisheries and most notably fuel industries. This result can also be observed for other indicators like gross output (p1) consumption of fixed capital (k1).

Figure 5 Deviation between National and Regional Coefficients for Intermediate Inputs (p2)



These observations make clear that forestry, fisheries, and fuel industry have a different cost structure on regional scale compared to the national averages. One reason for this could be that policy measures like subsidies for some inputs are implemented on a regional scale. As the compared figures refer to output and intermediate demand at basic prices and not to physical units, the distortions caused by regional policies may be severe. In the absence of additional information, the only way to address this issue is to assume regionally different tax- and subsidy rates and account for this in the final balancing steps. At the stage of compilation of prior information, only the available data can be used.

4.2. Critical Regional Economic Branches

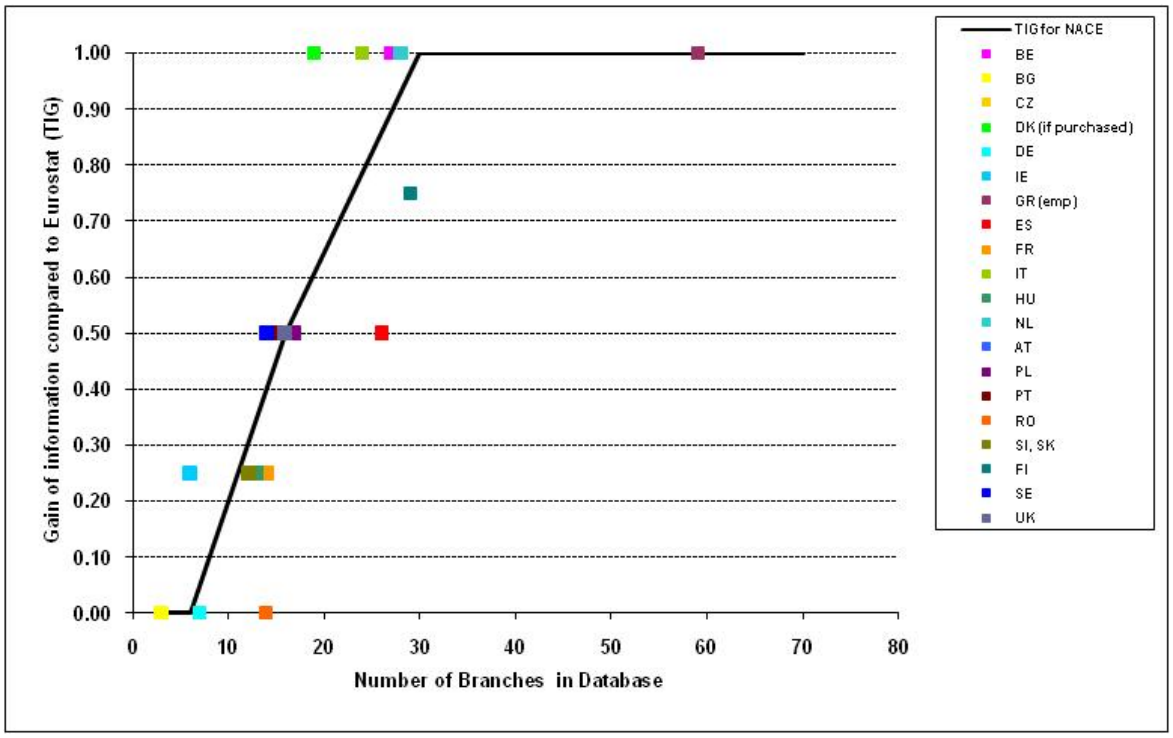
The previous section has given an outline on the concepts used to combine the available datasets and to complete the core accounts necessary for the subsequent SAM compilation. Basically two steps can be distinguished:

1. Expansion of available regional branch account data in A6 format to the target b19 branch classification based on national shares if datasets from NSAO were not available (e.g. Figure 3)
2. Derivation of missing branch account data based on national coefficients. (e.g. Figure 5)

When comparing the results obtained from these steps with observed branch account data from NSO, we observed substantial deviations in the regional composition of the A2B and C2E branch aggregates from the national averages, particularly for forestry (a02), fisheries (b05), mining and quarrying (C), and fuel industries (df). This observation is not surprising as these branches tend to be concentrated in certain regions and are actually branches which constitute regional economic heterogeneity. It became clear that for those branches additional information has to be collected. For industrial sectors, Structural Business Statistics proved to be a valuable if limited source (see section 2.4.2). Data on fisheries and forestry could not be obtained in a comparable manner, but the general availability of CAPRI data for agriculture (a01) and the fact that the share of the branch aggregate “agriculture and forestry” (A) within

the A2B aggregate was in many cases uniformly distributed across regions facilitated the derivation of consistent branch account data for forestry as residual. Based on this, figures for fisheries could also be derived residually.
For the industrial sectors (C2E), it appeared that the majority of NSO provided regional, dis-aggregated branch account data, although with varying levels of detail. The informational gain is depicted in Figure 6, which shows that 15 Member States with more than one NUTS2 region permit an improvement of at least 25% compared to the sole use of ESTAT branch accounts and national shares, so that the split into the targeted 5 sub-sectors (C,da,df,dz,E) can be performed with greater reliability.

Figure 6 Number of branches in regional accounts for b1g within the C2E aggregate, and informational gain compared to ESTAT data – only Member States with more than 1 NUTS2 region



5. Conclusions and Theses for Further Regional SAM Projects

This paper outlined several findings from an ongoing project on the compilation of regional SAMs for the European Union at NUTS2 level. The focus was on the creation of an inventory of existing regional branch account datasets which can serve as starting point for the derivation of regional SAMs. The inventory covered datasets from ESTAT, EU Member States' statistical organisations and previous projects on region SAM compilation. One major finding was that the datasets provided by NSOs greatly improved the core database for the IOTNUTS2 project but did not result in a full coverage of needed information. Consequently, we tested procedures to generate complete sets of core account data based exclusively on ESTAT data and compared the results with obtained NSO statistics. We could conclude that combinations of ESTATs regional branch accounts with national shares and coefficients resulted in satisfying estimates for missing values in many branches. Notable exceptions were

such branches that are usually concentrated in certain regions like mining and quarrying, fisheries, forestry, or fuel industries. For these branches, the default option to rely on ESTAT data was not applicable and we tested explicitly for alternative sources of data. In many cases, information from NSO and structural business statistics improved could be used for the derivation of plausible estimates.

After the derivation of regional core accounts, the regional SAMs should be derived. The related literature established a widely accepted sequence of procedural steps, titled the "Generation of Regional Input-Output Tables" procedure, or GRIT. GRIT draws mainly on the application of location quotients and gravity models. Location Quotients (LQs) appeared to play a crucial role on several stages of the compilation sequence. The objective of LQs is to derive the regional intermediate demand-from-regional-origin by adjusting the national coefficients according to the weight a particular branch has in the regional economy, usually measured by employment-shares of the specific branch in the region. It is assumed that a high weight of a certain branch in a region causes the development of other regional branches that may provide the needed intermediate inputs. This is plausible in the case of intermediates with high transportation cost, thus making regional branches competitive compared to suppliers from other regions. (see e.g. Flegg and Webber, 1995, or Bonfiglio and Chelli, 2008). Of particular interest are the two Flegg location quotients (Flegg LQ and Augmented Flegg LQ) as they do not only depend on sectoral national and regional employment data, but also on the choice of the parameter δ ($0 \leq \delta < 1$) that introduces an "element of flexibility" (Flegg et al 1995). The choice of δ depends on empirical considerations (Flegg at al 1995), statistical properties were investigated by Bonfiglio and Chelli (2008), finding higher values for δ to yield better results based on a Monte Carlo Analysis.

The performance of the alternative location quotients was tested for the available Spanish IOTs. The root mean squared error (RMSE) was calculated for each LQ and divided by the RMSE of the default setting LQ=1. In all cases, an improvement over the simple usage of national un-adjusted coefficients could be achieved (all values smaller than 1). However, the best performance can be observed by applying an AFLQ with a δ of 0.4, which is well in line of the findings of Bonfiglio and Chelli (2008), who observed best performance of AFLQ with a δ between 0.3 and 0.5.

Moreover, the database should include sub-matrices for intermediates from domestic and imported origin. The resulting need to estimate inter-national and inter-regional trade flows due to the lack of recorded data is a widely recognized challenge for multi-regional modelling. Gravity models are usually employed to solve this problem. In our case, a gravity model may help to derive total intermediate demands from domestic or imported origin, but not to derive the full sub-matrices. To derive the full set of needed sub-matrices, we will derive intermediate domestic and imported demand using national total intermediate demand coefficients. A first attempt on Spanish data shows that the goodness of fit of the thus derived a-priori values for the two sub-matrices are not entirely satisfying, nevertheless the correlation is still high. Having calculated the demand for regionally produced intermediates based on location quotients and total intermediate use as well as imported intermediate use as illustrated above, the sub-matrix of domestic imported intermediate use may now be computed residually.

Having generated a-priori entries for the production accounts of the target IOT, the value-added has to be distributed across the receiving institutions: government and households. For simplicity, we assume that private households are the primary recipients of wages and salaries (d11) and gross operating surplus (b2g_b3g), while the local government receives the indirect taxes paid by the production accounts (d29md39) and employers' social contributions (d12). The thus obtained aggregate private income (d11 + b2g_b3g) is then used to determine direct

taxes and transfers based on national tax and transfer rates as obtained from the NASA dataset. Due to the lack of information on regional consumption expenditures, we have to resort again to national consumption rates to derive aggregate private regional consumption The next item to be derived is aggregate regional investment. The ESTAT dataset "reg_e2gfcf" provides information for 175 NUTS2 regions, in some cases overlapping with data from NSO RAMA datasets. In case no information is available, we resort to the usage of investment shares in national income.

The prior database developed throughout the previous steps does not necessarily fulfil the requirement that regional expenditures equal regional revenues, nor that the regional IO-Tables add up to the national ones. Based on the experience gained on former projects (Mueller et al, 2009), we resort to the application of a Highest Posterior Density (Heckelei, et al., 2008 and Witzke and Britz, 2005) rather than a Generalised Cross Entropy (GCE, Golan et al 1994, Robinson et al 2001) procedure for balancing the IOTNUTS2 database.

The inventory of the database we developed within this paper and the described methodology to derive national and regional SAM for EU 27 member states might be important for the further development of the GTAP consortium database. The experience of the former AgroSAMs project (Mueller et al., 2009) confirmed the importance of building reliable a priori Social Accounting Matrices in order to produce a high quality database. A first step to produce a new set of GTAP format national I-O tables for EU27 should start from the creation of a complete set of macroeconomic indicators, which serve as control totals for the compilation of the final SAMs. National SAMs for 2005 derived from this work, although coarser than the required GTAP disaggregation, can be used, in combination with the CAPRI database to further disaggregate the agricultural sector, as control totals to update exiting tables. The IOTNUTS2 database will have additional advantages to compile new EU27 GTAP format tables. First of all the new baseyear will be 2005 compared with the actual one which is 2000. Secondly, the IOTNUTS2 database is derived from Input-Output tables, contrary to the AgroSAMs project where the starting point was a full set of Supply and Use tables. This will avoid a series of steps that were needed to convert AgroSAMs into GTAP format I-O tables as: revaluation of the total Use matrices to account for trade and transport margins, the split into domestic and imported Use matrices, the creation of a tax matrices for commodity taxes on purchases of domestic and imported commodities and the generation of a commodity by commodity input-output tables using the industry technology assumption required by GTAP (Lakatos et al.).

6. References

Bonfiglio, A. and F. Chelli (2008): Assessing the Behaviour of Non-Survey Methods for Constructing Regional Input-Output Tables through a Monte Carlo Simulation. In: Economic Systems Research, 20(3): 243-258

Bröcker, J., U. Kretschmer, C. Schürmann, D. Stelder, K. Spiekermann, M. Wegener (2002): The IASON Common Spatial Database. IASON (Integrated Appraisal of Spatial economic and Network effects of transport investments and policies) Deliverable 3. Funded by 5th Framework RTD Programme. Kiel/Dortmund: Christian-Albrechts-Universität Kiel/Institut für Raumplanung, Universität Dortmund.

ESTAT (2008): ESTAT Manual of Supply, Use and Input-Output Tables. URL: http://epp.ESTAT.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-07-013

Flegg A. T., C. D. Webber, and M. V. Elliott (1995): On the appropriate use of location quotients in generating regional input-output tables. In: Regional Studies 29: 547- 561

Golan, A., G. Judge, and S. Robinson (1994): Recovering Information from Incomplete or Partial Multisectoral Economic Data. In: Review of Economics and Statistics, 76(3): 541–49.

Heckelei, T., R. Mittelhammer, and T. Jansson (2008): A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models. Agricultural and Resource Economics, Discussion Paper 2008:2, Bonn, Germany.

Lakatos, C., McDonald S. and Mueller M. (2008), Data Sources and Construction Procedures, Europe in Badri Narayanan G. and Terrie L. Walmsley, Editors (2008). Global Trade, Assistance, and Production: The GTAP 7 Data Base, Center for Global Trade Analysis, Purdue University.

Jean S. and D. Laborde (2004): The Impact of Multilateral Liberalisation on European Regions: a CGE Assessment. CEPII Working Paper 2004-20. URL: http://www.cepii.fr/anglaisgraph/workpap/summaries/2004/wp04-20.htm

Müller, M., I. Pérez Domínguez, and S. H. Gay (2009): Construction of Social Accounting Matrices for EU27 with a Disaggregated Agricultural Sector (AgroSAM). Luxembourg: Office for Official Publications of the European Communities. EUR – Scientific and Technical Research series ISBN 978-92-79-13625-2

Robinson, S., A. Cattaneo, and M. El-Said (2001): Updating and estimating a Social Accounting Matrix using cross entropy methods. In: Economic Systems Research, 13(1): 47-64.

Rueda-Cantuche, J. M., J. Beutel, F. Neuwahl, I. Mongelli, and A. Loeschel (2009): A Symmetric Input-Output Table for EU27: Latest Progress. In: Economic Systems Research 21(1):59-79

Törmä, H. and K. Zawalinska (2007): Technical description of the CGE RegFin/RegPol models. University of Helsinki, RURALIA Institute. URL: http://www.helsinki.fi/ruralia/research/manuals.htm

Witzke, H.-P. and W. Britz (2005): Consolidating trade flows and market balances globally using a Highest Posteriori Density estimator. Presented at the 8th Annual Conference on Global Economic Analysis, Lübeck, Germany