Social Accounting Matrices and Satellite Accounts for EU27 on NUTS2 Level (SAMNUTS2)

Marc Mueller
Emanuele Ferrari

2013
Executive Summary

The reforms of the Common Agricultural Policy (CAP) are creating increasing interest amongst policy makers and analysts in the backward and forward linkages between regional agricultural and non-agricultural sectors, related labour markets, and regional development policies. There is a need to capture the effects of these policies on all branches of affected regional economies and thus ensure a better regional implementation of agricultural and regional development policies. Consequently, the demand for model-based analyses of regional development policies in a multi-sector context is increasing.

A relevant factor for the evaluation of such policies is the heterogeneity of the affected regions. Addressing regional heterogeneity requires multi-sector data on a sub-national scale. Existing datasets with these characteristics are usually not sufficiently detailed, which propelled the development of numerous non-sample methods to generate regional Social Accounting Matrices (SAM) based on combinations of regional indicators and national datasets. For some EU Member States, the construction of consistent regionalized tables has been pursued mainly by National Statistical Offices following survey-based methods. Also, several national research institutes applied non-survey-based methods and linked them to multi-sector regionalized national models. To the best of our knowledge, a complete set of SAMs for all the EU NUTS2 regions does not yet exist and this knowledge gap was addressed by the SAMNUTS2 project presented here, which was funded and supported by the European Commission's Joint Research Centre, Institute for Prospective Technological Studies (IPTS). The importance of the database of regional SAMs developed in the course of the SAMNUTS2 project is two-fold: first, it permits the implementation of regionalized Computable General Equilibrium (CGE) models for policy analyses. This allows the feedbacks of agricultural and regional development policies on non-agricultural sectors and on regional factor markets to be assessed. Secondly, it permits the construction of soft and hard linkages between existing regionalized CGE models and partial equilibrium (PE) models focusing on the agricultural sector.

This report describes the steps we took to build a database of Social Accounting Matrices for all of the 271 NUTS2 regions of the EU. This database is called SAMNUTS2. As a first conceptual step, we acknowledged that the relevant information available from international statistical data providers may not be as detailed and complete as data possibly available from national statistical organisations (NSO). Consequently, we contacted the NSO in the EU Member States and developed an inventory of relevant datasets for the compilation of the specified regional SAMs. In addition, reports from previous projects on the creation of regional SAMs were consulted. It turned out that a significant informational gain could be achieved by consulting the Member States NSOs and using their data.

To combine the obtained information with readily available datasets from international data providers, we applied and developed statistical methods to exploit and harmonize the available data as much as possible. To ensure consistency between regional and national levels, particularly in the context of sector employment, generation of value-added, consumption, and intra-national trade flows, we applied a Bayesian estimation approach. The SAMNUTS2 estimation procedure generates a database that can be used to put regional CGEs for all EU Member States at NUTS2 into operation. It should be noted again that the SAMNUTS2 database uses all available information in an efficient manner and preserves the recorded data structure as long as accounting and modelling constraints are not violated. This emphasizes the nature of SAMNUTS2 as a model and
not a statistical database. It also means that the actual entries in the SAMNUTS2 database are not static. Depending on the targeted models, the database may change due to new accounting or calibration constraints. Should the number of models increase, the parallel existence of different versions of the SAMNUTS2 database could follow. To what extent such a development would be desirable is questionable and should be taken into account in further stages. Finally, the quality of the SAMNUTS2 database can only be improved through actual usage as many implausible entries can only be detected through calibration of and simulation with a variety of computational models. We are looking forward to the future developments.

Disclaimer:
The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.
Table of Contents

Executive Summary............................................................................................................. 1
Table of Contents ............................................................................................................. 3
1. Introduction and Motivation ............................................................................................. 5
2. Contents and Structure of the SAMNUTS2 Database ....................................................... 7
   2.1 Nomenclature of Territorial Units for Statistics......................................................... 12
3. Review of datasets available from the European Commission ........................................ 14
   3.1 IPTS ........................................................................................................................ 14
   3.2 National Datasets from Eurostat ............................................................................. 14
      3.2.1 National Accounts – Supply, use and Input-output tables (NAIO) ................. 14
      3.2.2 National Accounts - Annual National Accounts (NAMA) ......................... 15
      3.2.3 National Accounts – Annual Sector Accounts (NASA) .............................. 17
      3.2.4 Summary National Datasets from Eurostat ...................................................... 18
   3.3 Regional Datasets from Eurostat ............................................................................ 18
      3.3.1 Branch Accounts - ESA95 (reg_ecobrch) ......................................................... 19
      3.3.2 Regional Structural Business Statistics (reg_sbs) .................................. 20
      3.3.3 Regional Agriculture Statistics (reg_agr) ......................................................... 21
      3.3.4 Summary Regional Datasets from Eurostat ...................................................... 22
3.4 Preliminary Conclusion ............................................................................................... 23
3.5 Multi-country IOT Projects ....................................................................................... 30
   3.5.1 IASON ............................................................................................................... 30
   3.5.2 DREAM ......................................................................................................... 30
   3.5.3 REAP&REEIO ................................................................................................. 30
   3.5.4 REAPBALK ..................................................................................................... 31
   3.5.5 ESPON .......................................................................................................... 32
4. Core Accounts ................................................................................................................ 33
   4.1 Expanding and Combining ESTAT and NSO Datasets ............................................ 34
   4.2 Derived Core Accounts .......................................................................................... 39
5. Derivation of IO-Tables .................................................................................................. 42
   5.1 Generation of Regional Input-Output Tables: GRIT ................................................ 42
      GRIT Phase 1: Adjustment of National IOT ......................................................... 42
      GRIT Phase 2: Adjustment for Regional Imports ................................................. 43
      GRIT Phase 3: Regional Branches ......................................................................... 43
      GRIT Phase 4: Derivation of Prototype Transactions Table .................................. 43
      GRIT Phase 5: Derivation of Final Transactions Table ........................................... 44
      GRIT Summary ..................................................................................................... 44
   5.2 Location Quotients ................................................................................................. 44
   5.3 Foreign or Domestic Origin of Used Commodities .................................................. 47
   5.4 Taxes, Transfers, and Institutions .......................................................................... 51
   5.5 Demand-side of the IO-Tables ................................................................................. 52
6. Balancing Procedure ....................................................................................................... 55
   6.1 General Methodological Considerations .................................................................. 55
   6.2 Balancing Approach for Core Accounts .................................................................. 57
   6.3 SAM Balancing Approach ...................................................................................... 60
      6.3.1 National SAM ............................................................................................... 60
      6.3.2 Total Transaction Matrices on NUTS2 Level ................................................. 61
      6.3.3 Detailed Transaction Matrices on NUTS2 Level .............................................. 63
7. Implementation ................................................................................................................ 64
8. Summary and Conclusion ............................................................................................... 67
References .......................................................................................................................... 69
Annex 1: Example for the Computation of TIG ................................................................. 72
Figure 1  Number of NUTS2 regions in the EU Member States ......................................... 13
Figure 2  Number of branches in regional accounts for b1g, and informational gain compared to Eurostat data – only Member States with more than 1 NUTS2 region .............. 28
Figure 3  Comparison of Regional Branch Account Data (RAMA) from NSO and ESTAT for Spain ..................................................................................................................................... 33
Figure 4  Three Datasets for Cornwall and Isles of Scilly (UKK3) ........................................ 35
Figure 5  Aggregator Matrix between b19 and A6 Classifications (G[19,6]) ............................... 36
Figure 6  Deviation between National and Regional Shares in Spain and Italy ................... 38
Figure 7  Composition of C2E in Spanish Regions ................................................................. 39
Figure 8  Deviation between National and Regional Coefficients for Gross Output (p1) ....... 40
Figure 9  Deviation between National and Regional Coefficients for Intermediate Inputs (p2) ........................................................................................................................................... 41
Figure 10 Performance of LQs in the Case of Spanish Regional SAMs ................................. 47
Figure 11 SAMNUTS2 Sub-matrices and Methods for Derivation ........................................ 48
Figure 12 Derived versus Recorded Values of Total Intermediate Demand in Spanish Regions .................................................................................................................................. 49
Figure 13 Derived versus Recorded Values of Imported Intermediates ................................. 50
Figure 14 Deviation between National and Regional Coefficients for Imported Intermediate Inputs (p2i) ........................................................................................................................................... 50
Figure 15 Deviation between prior and balanced regional core accounts for p2pp (S/S0) ....... 60
Figure 16 Deviation between balanced national and regional wages (wage/wageMS) ......... 60
Figure 17 Deviation between prior and balanced national SAM for UK 2005 (S/S0) ........... 61
Figure 18 Deviation between prior and balanced regional total transaction matrix for Andalucía ES61 (S/S0) .................................................................................................................................... 62
Figure 19 Deviation between prior and balanced regional detailed transaction matrix for Andalucía ES61 (S/S0) .................................................................................................................................... 63
Figure 20 Flow of Raw-Data Import in the SAMNUTS2 Procedure ..................................... 65
Figure 21 Flow of Compilation Steps in the SAMNUTS2 Procedure ..................................... 65
Figure 22 Directories of the SAMNUTS2 Main Folder ............................................................ 66
Figure 23 Number of NUTS2 regions and resource use of compilation steps (in CPU seconds on normal desktop PC) ......................................................................................... 66

Table 1  Target Accounts of the SAMNUTS2 Database .......................................................... 10
Table 2  Availability of Member States’ National IOT No. 1800 and 1900 ................................. 15
Table 3  Selected Datasets from the NAMA Domain ................................................................. 16
Table 4  Number of Branches in nama_nace60_c for Gross Value-Added (b1g) ......................... 17
Table 5  Datasets from the “Branch accounts - ESA95” Domain ............................................. 20
Table 6  Availability of employment data for da15 and DA (in brackets) from sbs_r_nuts03 (in percent of total NUTS2 regions) ............................................................................. 21
Table 7  Availability of data on agricultural gross value-added and compensation of employees ................................................................................................................................. 22
Table 8  Contact details of national statistical offices .............................................................. 25
Table 9  Data availability for agriculture and forestry in 2005 .................................................. 29
Table 10 Location Quotients ..................................................................................................... 46
Table 11 Core Accounts in First Balancing Step ...................................................................... 58
1. Introduction and Motivation

The reforms of the Common Agricultural Policy (CAP) is creating increasing interest amongst policy makers and analysts in the backward- and forward linkages between regional agricultural and non-agricultural sectors, related labour markets, and regional development policies. Policy makers need to capture the effects of these policies on all branches of the economy and to allow a better regional scaling of agricultural policies. Thus, demand for model-based analyses of regional development policies in a multi-sector context is increasing.

Mattas et al. (2011), in their analysis for 5 regions in the EU, highlight that an inspection of regional Input-Output Tables (IOTs) reveals deep differences in the structure of the regional economy. The size of same sectors and the distribution of multipliers diverge among the regions. According to Mattas et al. (2011) this is a clear indication that Pillar II programmes to be effective and boost the regional economy should be highly flexible. If this is true for a sample of 5 regions, the differences will be magnified when the all set of regions of the EU 27 will be taken into account. Addressing regional heterogeneity requires multi-sector data on a sub-national scale. Existing datasets with these characteristics are usually not sufficiently detailed, which give rise to numerous non-survey methods to generate regional IOTs based on combinations of regional indicators and national datasets. In the literature, there are many examples of regionalization of national tables for single or multiple regions. In addition, at national level some tentative to construct consistent regionalized tables have been pursued mainly by National Statistical Offices (NSO) following survey-based method (i.e. Finland, OFS) or national research institutes following non-survey-based methods (Fritz et al., 2003) and link them to multi-sectoral regionalized national models (Casini-Benvenuti and Paniccia, 2003). At the best of our knowledge, a complete set of SAMs for all the EU NUTS2 does not yet exist and this work fulfil this absence in the literature.

The goal of this report is to describe the steps to build a database of Social Accounting Matrices (SAM) for all the 271 NUTS2 regions of the EU. The database is called SAMNUTS2. In the first step, we create an inventory of datasets relevant for the compilation of the specified regional SAMs. The inventory include regional datasets for the EU: national and regional databases from Eurostat (ESTAT), national statistical departments as well as databases developed in comparable previous projects (e.g. DREAM, RAEM). The report illustrates the significant informational gain attained with the datasets coming from Member States NSOs. In the second step, we develop a statistical method to exploit the data. Following standard non-survey procedures, the data are combined to populate the regional SAMs. Survey-based regional tables coming from NSOs are used to test the reliability of the techniques adopted in this work to combine national and regional datasets. This test shows that for the majority of economic sectors, non-survey methods generate reliable substitutes for otherwise collected indicators. In the third step, the matrices produced are balanced following a modified Stone-Byron method.

The importance of this database is two-folded. A SAMs based dataset permits to exploit the characteristics of Computable General Equilibrium (CGE) analyses. Firstly, this allows assessing the feedbacks of agricultural policies on non-agricultural sectors and on the factor markets. 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 all the policies related to the so-called Pillar II of the CAP can be regionally modelled. Such measures primarily target the agricultural sector, but are likely to influence other economic sectors and aggregate regional income, depending on the regional economic structure and the dominance of agriculture. Secondly, the construction of such a regional database allows constructing soft and hard linkages between existing regionalized national Computable General Equilibrium (CGE) models (e.g. RegFin/RegPol - Torma and Zawalinska, 2007) and partial equilibrium models, which already covers all the EU NUTS2 regions, like the "Common Agricultural Policy Regionalized Impact analysis modelling system" (CAPRI – Britz and Witzke, 2008). According to Torma et al. 2010, the advantages of both types of modelling, the generality of the CGE and its capacity of taking into account all the aspects of an economy and the "depth" of the PE model and the abundance of details in the modelling of a single sector, can be exploited in three main ways. They might be integrated through their database (Mueller et al., 2009) or with a sequential implementation of scenarios, where results of one model serve as input for the other models (Nowicki et al., 2009 for the Scenar2020 Project). A third approach, which is allowed only with the development of a database like the SAMNUTS2, is the iterative calibration of structural model parameters, as developed within the SEAMLESS project (Jansson et al. 2009). This approach ensured the harmonized simulation behaviour of the models for matching endogenous variables.

The rest of the report is organized as follows. Section 2 presents the structure of the database produced. In section 3 the inventory of existing database and the information gains from NSO database is presented. Section 4 details the creation of the core accounts while Section 5 illustrates the procedure to produce the regional I-O tables. Section 6 describes the balancing procedure while Section 7 presents the technical implementation of the different steps. Section 8 concludes.
2. Contents and Structure of the SAMNUTS2 Database

The targeted SAMNUTS2 database consists of numerous sub-tables, which may be broadly classified into core-SAM accounts and satellite or auxiliary accounts. The core datasets are those which 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, for instance in the case of gross value-added, which 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 SAMNUTS2 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 of the same country or imported from the rest of the world. The core SAMs and the respective sub-matrices are summarized below:

\[
S = \begin{pmatrix}
  b & d & 1 & w & 1 \\
  Aw & Cw & lw & Ew & \\
  b & A & C & I & E & X \\
  f & F & T & 0 & \\
  f+1 & Ts & 0 & \\
  1 & X & 0 & 0 & 0
\end{pmatrix}
\]

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
  \[
  (C_{b,d} = \sum_w Cw_{b,w,d})
  \]
- \(lw\): Investment demand, distinguished by origin
- \(I\): Total investment demand
  \[
  (I_b = \sum_w lw_{b,w})
  \]
- \(Ew\): 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
\( E_{b,w} = \sum_w E_{b,w} \)

Ts: Taxes on production and factors

T: Transactions between institutions (distribution of regional income)

X: Total regional supply (has to equal total regional demand):
\[
X_{b'} = \sum_b A_{b,b'} + \sum_f \left[ F_{f,b'} + TS_{f,b'} \right] + TS_{s,b'}
\]
\[
= \sum_b A_{b,b'} + \sum_d C_{b,d} + I_{b'} + \sum_w E_{b',w}
\]

The branch dimension (b) follows the “Nomencature for Economic Activities” (NACE) within the European System of National Accounts (ESA95). It follows the NACE classification at first digit, with the exceptions of Agriculture, Hunting and Forestry (NACE: A) and Manufacturing branches (NACE: D). These branches have been further disaggregated into 2 and 3 sub-branches. Therefore, dimension b contains 19 branches, but data have been also been collected for the aggregates A and D.

Primary factors (f) contain the elements “Labour”, “Land”, and “Physical capital”. Trade partners (w) for each region are the intra-national markets for trade between regions and 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 EU with 27 Member States. Regional consumers (d) are the regional government and a single representative household.

A full set of the sub-matrices discussed above is not available for most of the NUTS2 regions. Intermediate demand matrices (A and Aw) were 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, lw, Ew, E) were not unavailable for the major part of regions. Final consumption (C) and investment demand (I) were in the best of cases available as an aggregate. The satellite accounts contain the type of information that is not directly part of the SAMs, but 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:

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 \( (V_b = \sum_{f} [F_{f,b} + Ts_{f,b}] + Ts_{s,b}) \)

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 with the ESA95 standards and to facilitate the communication with the Member States' statistical departments, the classification scheme used follows in general the ESA95 nomenclature (Table 1). Exceptions originate from the specific needs of the regionalised database, for which e.g. the category "p2i: Use of imported products, cif" is subdivided into "p2d: Use of imported products of domestic origin, cif" and "p2f: Use of imported products of foreign origin, cif". Other deviations from the ESA95 standard are the introduction of taxes and subsidies on the primary factor "land", as well as associated land revenues. The newly introduced codes are highlighted in red in Table 1.
Table 1  Target Accounts of the SAMNUTS2 Database

<table>
<thead>
<tr>
<th>Block Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA01</td>
<td>&quot;Agriculture, hunting and related services&quot;</td>
</tr>
<tr>
<td>AA02</td>
<td>&quot;Forestry, logging and related services&quot;</td>
</tr>
<tr>
<td>B000</td>
<td>&quot;Fishing&quot;</td>
</tr>
<tr>
<td>C000</td>
<td>&quot;Mining and quarring&quot;</td>
</tr>
<tr>
<td>DA00</td>
<td>&quot;Food products, beverages, and tobacco&quot;</td>
</tr>
<tr>
<td>DF00</td>
<td>&quot;Coke, refined petroleum products, and nuclear fuels&quot;</td>
</tr>
<tr>
<td>DZ00</td>
<td>&quot;Other manufacturing&quot;</td>
</tr>
<tr>
<td>E000</td>
<td>&quot;Electricity, gas and water supply&quot;</td>
</tr>
<tr>
<td>F000</td>
<td>&quot;Construction&quot;</td>
</tr>
<tr>
<td>G000</td>
<td>&quot;Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods&quot;</td>
</tr>
<tr>
<td>H000</td>
<td>&quot;Hotels and restaurants&quot;</td>
</tr>
<tr>
<td>I000</td>
<td>&quot;Transport, storage and communication&quot;</td>
</tr>
<tr>
<td>J000</td>
<td>&quot;Financial intermediation&quot;</td>
</tr>
<tr>
<td>K000</td>
<td>&quot;Real estate, renting and business activities&quot;</td>
</tr>
<tr>
<td>L000</td>
<td>&quot;Public administration and defence; compulsory social security&quot;</td>
</tr>
<tr>
<td>M000</td>
<td>&quot;Education&quot;</td>
</tr>
<tr>
<td>N000</td>
<td>&quot;Health and social work&quot;</td>
</tr>
<tr>
<td>O000</td>
<td>&quot;Other community, social, personal service activities&quot;</td>
</tr>
<tr>
<td>P000</td>
<td>&quot;Activities of households&quot;</td>
</tr>
</tbody>
</table>

Target sectors

<table>
<thead>
<tr>
<th>Block Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A000</td>
<td>&quot;Agriculture, hunting and forestry&quot;</td>
</tr>
<tr>
<td>D000</td>
<td>&quot;Manufacturing&quot;</td>
</tr>
<tr>
<td>A2B</td>
<td>&quot;Agriculture, hunting, forestry and fishing&quot;</td>
</tr>
<tr>
<td>C2E</td>
<td>&quot;Total industry (excluding construction)&quot;</td>
</tr>
<tr>
<td>F00</td>
<td>&quot;Construction&quot;</td>
</tr>
<tr>
<td>G21</td>
<td>&quot;Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication&quot;</td>
</tr>
<tr>
<td>J2K</td>
<td>&quot;Financial intermediation; real estate, renting and business activities&quot;</td>
</tr>
<tr>
<td>L2P</td>
<td>&quot;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&quot;</td>
</tr>
</tbody>
</table>

Sector aggregates (NACE16 and NACE06)

<table>
<thead>
<tr>
<th>Block Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2r</td>
<td>&quot;Use of domestic/regional products&quot;</td>
</tr>
<tr>
<td>p2i</td>
<td>&quot;Use of imported products, cif&quot;</td>
</tr>
<tr>
<td>p2d</td>
<td>&quot;Use of imported products of domestic origin, cif&quot;</td>
</tr>
<tr>
<td>p2f</td>
<td>&quot;Use of imported products of foreign origin, cif&quot;</td>
</tr>
<tr>
<td>d21</td>
<td>&quot;Taxes less subsidies on products&quot;</td>
</tr>
<tr>
<td>d21md31</td>
<td>&quot;Taxes less subsidies on products&quot;</td>
</tr>
<tr>
<td>d21</td>
<td>&quot;Taxes on products&quot;</td>
</tr>
<tr>
<td>d31</td>
<td>&quot;Subsidies on products&quot;</td>
</tr>
<tr>
<td>p2pp</td>
<td>&quot;Total intermediate consumption/Final use at purchasers' prices&quot;</td>
</tr>
<tr>
<td>d1</td>
<td>&quot;Compensation of employees&quot;</td>
</tr>
<tr>
<td>d11</td>
<td>&quot;Wages and salaries&quot;</td>
</tr>
<tr>
<td>d12</td>
<td>&quot;Employers' social contributions&quot;</td>
</tr>
<tr>
<td>d29md39</td>
<td>&quot;Other net taxes on production&quot;</td>
</tr>
<tr>
<td>d29</td>
<td>&quot;Other taxes on production&quot;</td>
</tr>
<tr>
<td>d29ln</td>
<td>&quot;Taxes on land (~d29l)&quot;</td>
</tr>
<tr>
<td>d29cap</td>
<td>&quot;Taxes on other physical capital&quot;</td>
</tr>
<tr>
<td>d39</td>
<td>&quot;Other subsidies on production&quot;</td>
</tr>
<tr>
<td>d39ln</td>
<td>&quot;Subsidies on land&quot;</td>
</tr>
<tr>
<td>d39cap</td>
<td>&quot;Subsidies on other physical capital&quot;</td>
</tr>
<tr>
<td>k1</td>
<td>&quot;Consumption of fixed capital&quot;</td>
</tr>
<tr>
<td>Block</td>
<td>Code</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>b2npb3n</td>
</tr>
<tr>
<td></td>
<td>b2gpb3g</td>
</tr>
<tr>
<td></td>
<td>b3g</td>
</tr>
<tr>
<td></td>
<td>b2g</td>
</tr>
<tr>
<td></td>
<td>b2glnd</td>
</tr>
<tr>
<td></td>
<td>b2gcap</td>
</tr>
<tr>
<td></td>
<td>b1g</td>
</tr>
<tr>
<td></td>
<td>b1gm</td>
</tr>
<tr>
<td></td>
<td>p34</td>
</tr>
<tr>
<td></td>
<td>p34s1</td>
</tr>
<tr>
<td></td>
<td>p34s2</td>
</tr>
<tr>
<td></td>
<td>b8g</td>
</tr>
<tr>
<td></td>
<td>b9</td>
</tr>
<tr>
<td></td>
<td>b12</td>
</tr>
<tr>
<td></td>
<td>d4</td>
</tr>
<tr>
<td></td>
<td>d5</td>
</tr>
<tr>
<td></td>
<td>d6</td>
</tr>
<tr>
<td></td>
<td>d7</td>
</tr>
<tr>
<td></td>
<td>d8</td>
</tr>
<tr>
<td></td>
<td>d9</td>
</tr>
<tr>
<td></td>
<td>p1</td>
</tr>
<tr>
<td></td>
<td>p119</td>
</tr>
<tr>
<td></td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>p3s14s15</td>
</tr>
<tr>
<td></td>
<td>p3s1311</td>
</tr>
<tr>
<td></td>
<td>p3s1313</td>
</tr>
<tr>
<td></td>
<td>p3</td>
</tr>
<tr>
<td></td>
<td>p5</td>
</tr>
<tr>
<td></td>
<td>p6s2</td>
</tr>
<tr>
<td></td>
<td>p6s1</td>
</tr>
<tr>
<td></td>
<td>p6</td>
</tr>
<tr>
<td></td>
<td>tfinu</td>
</tr>
<tr>
<td></td>
<td>tu</td>
</tr>
</tbody>
</table>

**Institutional accounts (columns)**

**Labor market (rows)**

<table>
<thead>
<tr>
<th>Block</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ulc</td>
<td>&quot;Unit labor cost&quot;</td>
</tr>
<tr>
<td></td>
<td>Pop</td>
<td>&quot;Total population&quot;</td>
</tr>
<tr>
<td></td>
<td>Act</td>
<td>&quot;Economically active population&quot;</td>
</tr>
<tr>
<td></td>
<td>Une</td>
<td>&quot;Unemployed persons&quot;</td>
</tr>
<tr>
<td></td>
<td>Emp</td>
<td>&quot;Total employment&quot;</td>
</tr>
<tr>
<td></td>
<td>Sal</td>
<td>&quot;Employees&quot;</td>
</tr>
<tr>
<td></td>
<td>Self</td>
<td>&quot;Self-employed&quot;</td>
</tr>
<tr>
<td></td>
<td>migs1</td>
<td>&quot;Net-migration, domestic&quot;</td>
</tr>
<tr>
<td></td>
<td>migs2</td>
<td>&quot;Net-migration, foreign&quot;</td>
</tr>
<tr>
<td></td>
<td>Mig</td>
<td>&quot;Net-migration&quot;</td>
</tr>
</tbody>
</table>

Note: Red fonts indicate that the respective elements are not part of the ESA95 classification scheme.
2.1 Nomenclature of Territorial Units for Statistics

The Nomenclature of Territorial Units for Statistics (NUTS) classification is a hierarchical system to divide the EU territory for the purpose of collection, development and harmonisation of EU regional statistics and socio-economic analyses of the regions. The regional classification follows this hierarchy: NUTS 1 (major socio-economic regions), NUTS 2 (basic regions for the application of regional policies), and NUTS 3 (small regions for specific diagnoses). The current NUTS classification lists 97 regions at NUTS 1, 271 regions at NUTS 2 and 1303 regions at NUTS 3. The internal administrative structure of the Member States is generally based on two of these three main regional levels. This existing national administrative structure may be, for example:

- Germany: at NUTS 1 and NUTS 3 levels (respectively the Länder and Kreise)
- United Kingdom: at NUTS 1 and NUTS 3 levels (standard regions and counties)
- France: at NUTS 2 and NUTS 3 (regions and départements)
- Spain: at NUTS 2 and NUTS 3 (Comunidades autónomas and provincias)
- Italy: at NUTS 2 and NUTS 3 (regioni and province)

Providing a complete breakdown, i.e. at all three NUTS, therefore means identifying a regional level for each Member State in addition to the two main levels mentioned above. This additional level thus corresponds to a regional structure that is less extensively used for administrative purposes - or which may indeed be instituted solely for this statistical purpose, without having any administrative function whatever. Depending on which levels already exist, the additional level may be created at any one of the three NUTS levels. Since France, for example, has functional administrative units at level 2 and 3, the additional level is introduced at NUTS level 1. This is also the case for Italy, Greece and Spain. By contrast, the additional "non-administrative" level is at NUTS level 2 for Germany and the United Kingdom and at NUTS level 3 for Belgium.

The NUTS regulation lays down the following minimum and maximum population thresholds for the average size of the NUTS regions:

- **LEVEL** Min./Max.
- NUTS 1 3 mlns /7 mlns
- NUTS 2 800 000 /3 mlns
- NUTS 3 150 000 /800 000

The number of NUTS2 regions per Member State varies remarkably (Figure 1). UK and Germany have the highest number of NUTS2 regions (37 and 39, respectively), followed by France, Italy and Spain. These five Member States cover more than half of the 271 NUTS2 regions, thus potentially making it possible to retrieve the bulk of needed data by contacting five statistical organisations. In contrast, the three Baltic States Estonia, Latvia, and Lithuania, the islands Cyprus and Malta and Luxemburg consist of only one NUTS2 region, so that data on national scale are sufficient for the purposes of this project.
Figure 1  Number of NUTS2 regions in the EU Member States

Note: Green bars indicate that the Member State consists of only one NUTS 2 region.
3. Review of datasets available from the European Commission

The first step in the compilation of the datasets inventory is the review of datasets available from the European Commission, namely Eurostat and the Institute for Prospective Technological Studies (IPTS) of the Joint Research Centre (JRC).

3.1 IPTS

Of particular relevance are the IPTS' national IOT developed by Rueda-Cantuche et al. (2009) as they cover all 27 EU Member States for the year 2000. The Eurostat homepage presents the different datasets on regional and national scale in numerous domains. The work by Müller et al (2009) is crucial to identify the linkages between Eurostat datasets from these domains.

3.2 National Datasets from Eurostat

3.2.1 National Accounts – Supply, use and Input-output tables (NAIO)

Supply and use tables are matrices by product and industry describing production processes and the transactions in products of the national economy with great detail. A symmetric input-output table is a 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 provided:

- 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, Eurostat applies the CPA P60 classification that delineates 59 products. For the classification of industries Eurostat uses NACE rev.1.1 A60 as reference which distinguishes 59 industries. As tables 1800 and 1900 permit the distinction of usage by their origin (domestic and foreign), they represent the most valuable input for the SAMNUTS2 database. Their availability is show in Table 2. For 2005, these tables are available for 20 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 2 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 2 Availability of Member States’ National IOT No. 1800 and 1900

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>Most recent IOT</th>
<th>Source for most recent IOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 NAIO</td>
<td></td>
</tr>
<tr>
<td>BG000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 IPTS</td>
<td></td>
</tr>
<tr>
<td>CZ000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
</tr>
<tr>
<td>DK000000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2005 NAIO</td>
<td></td>
</tr>
<tr>
<td>DE000000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2005 NAIO</td>
<td></td>
</tr>
<tr>
<td>EE000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IE000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR000000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2005 NAIO</td>
<td></td>
</tr>
<tr>
<td>IT000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CY000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 IPTS</td>
<td></td>
</tr>
<tr>
<td>LV000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 IPTS</td>
<td></td>
</tr>
<tr>
<td>LT000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LU000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 IPTS</td>
<td></td>
</tr>
<tr>
<td>HU000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 IPTS</td>
<td></td>
</tr>
<tr>
<td>NL000000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2005 NAIO</td>
<td></td>
</tr>
<tr>
<td>AT000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RO000000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2005 NAIO</td>
<td></td>
</tr>
<tr>
<td>SI000000</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
</tr>
<tr>
<td>SK000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FI000000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2005 NAIO</td>
<td></td>
</tr>
<tr>
<td>SE000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>2005</td>
<td>NAIO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK000000</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000 NAIO</td>
<td></td>
</tr>
</tbody>
</table>

Total 27 6 5 6 6 20

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

3.2.2 National Accounts - Annual National Accounts (NAMA)

In addition to the national IOT from the NAIO domain, information from the Annual National Accounts (NAMA) is 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 3.
Table 3  Selected Datasets from the NAMA Domain

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Covered items</th>
</tr>
</thead>
<tbody>
<tr>
<td>nama_nace06_c</td>
<td>National Accounts by 6 branches - aggregates at current prices</td>
<td>b1g, d1, d11</td>
</tr>
<tr>
<td>nama_nace06_e</td>
<td>National Accounts by 6 branches - employment data</td>
<td>emp, sal, self</td>
</tr>
<tr>
<td>nama_nace31_c</td>
<td>National Accounts by 31 branches - aggregates at current prices</td>
<td>k1, p1, p2, p5, b1g, d1, d11, b2n_b3n, d29_m_d39</td>
</tr>
<tr>
<td>nama_nace31_e</td>
<td>National Accounts by 31 branches - employment data</td>
<td>emp, sal, self</td>
</tr>
<tr>
<td>nama_nace60_c</td>
<td>National Accounts by 60 branches - aggregates at current prices</td>
<td>k1, p1, p2, b1g, d1, d11, b2n_b3n, d29_m_d39</td>
</tr>
<tr>
<td>nama_nace60_e</td>
<td>National Accounts by 60 branches - employment data</td>
<td>emp, sal, self</td>
</tr>
<tr>
<td>nama_aux_pem</td>
<td>Population and employment - Annual data</td>
<td>pop, emp_nc, emp_dc, sal, self</td>
</tr>
<tr>
<td>nama_fcs_c</td>
<td>Final consumption aggregates - Current prices</td>
<td>p3, p31_s14_s15, p3_s13, p33, p34</td>
</tr>
<tr>
<td>nama_gdp_c</td>
<td>GDP and main components - Current prices</td>
<td>b1gm, p3, p31_s14_s15, p3_s13, p5, p6, p7, b1g, d21_m_d31, d1, b2g_b3g, d2_m_d3</td>
</tr>
</tbody>
</table>

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 SAMNUTS2 database. From 27 Member States’ datasets, only 14 cover all 19 branches (Table 4). However, closer examination reveals that although figures for “da15: Food products and beverages” is not available, the aggregate “da: food products, beverages, and tobacco” exists. The same applies for “a01: agriculture” and “a02: forestry”, for which in some cases also the aggregate “a: agriculture and forestry” is available. As these sectors are distinguished in the final database, additional sources of information have to be used, e.g., the IOTs form the NAIO domain. Concerning “p: Activities of households”, it has to be clarified if missing entries indicate that such a branch is not considered in the respective national accounting schemes or merged with other branches.
### Table 4: Number of Branches in nama_nace60_c for Gross Value-Added (b1g)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BE0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>BG0000000</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>a01, a02 (but a), da15 (but da), p</td>
</tr>
<tr>
<td>CZ0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>DK0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>DE0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>EE0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>EL0000000</td>
<td>16</td>
<td>16</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>GR0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>ES0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>FR0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>IT0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>CY0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>df (but available in other years)</td>
</tr>
<tr>
<td>LV0000000</td>
<td>18</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>p</td>
</tr>
<tr>
<td>LT0000000</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>df</td>
</tr>
<tr>
<td>LU0000000</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>b, da15 (but da), df</td>
</tr>
<tr>
<td>HU0000000</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>p</td>
</tr>
<tr>
<td>MT0000000</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>a02 (but a), df (but avail. in other years)</td>
</tr>
<tr>
<td>NL0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>AT0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>PL0000000</td>
<td>15</td>
<td>15</td>
<td>13</td>
<td>16</td>
<td>16</td>
<td>18</td>
<td>da15 (but da)</td>
</tr>
<tr>
<td>PT0000000</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>RO0000000</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>p</td>
</tr>
<tr>
<td>SI0000000</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>da15 (but da)</td>
</tr>
<tr>
<td>SK0000000</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>p</td>
</tr>
<tr>
<td>FI0000000</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>SE0000000</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>da15 (but da)</td>
</tr>
<tr>
<td>UK0000000</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>a01, a02 (but a), da15 (but da)</td>
</tr>
</tbody>
</table>

Note: Bold fonts indicate missing branches. Red and blue fonts indicate missing IOTs for 2005 (Table 2)

Comparing Tables 2 and 4, 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”. Missing data for “da15” may be replaced by “da” if it could be decided to deviate from the original “da15” disaggregation. For Luxemburg, it has to be clarified whether “b: Fisheries” and “df: fuel industries” do not exist as branches, or if they are just missing. 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.

#### 3.2.3 National Accounts – Annual Sector Accounts (NASA)

The annual sector accounts (NASA) contain data on flows between sectors, domestic institutions, and the ‘rest of 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. 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.

3.2.4 Summary National Datasets from Eurostat

Concerning national IOTs in the desired format (i.e. NAIO Tables 1800 and 1900), 20 Member States are covered for the year 2005. As a full set of IOTs is available for the year 2000 (either from Eurostat or Rueda-Cantuche et al. 2009) the first step is to update the IOTs for the missing 7 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. The critical Member States are Luxemburg, Cyprus, and Malta, as neither NAIO, NAMA or 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.

3.3 Regional Datasets from Eurostat

The "Regional statistics (reg)" section of Eurostat covers a wide range of indicators (see screenshot below). Particular attention is devoted to those sub-sections which provide information for the structure of economic branches in the NUTS2 regions.
3.3.1 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 5. 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 main shortcoming of these datasets is the coarse representation of economic branches.

Table 5 Datasets from the “Branch accounts - ESA95” Domain

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Regional coverage in 2005 (NUTS2)</th>
<th>Covered items</th>
</tr>
</thead>
<tbody>
<tr>
<td>reg_e2gfcf</td>
<td>Gross fixed capital formation at NUTS level 2</td>
<td>p5</td>
<td></td>
</tr>
<tr>
<td>reg_e2rem</td>
<td>Compensation of employees at NUTS level 2</td>
<td>CY and LU missing in 2005, but other year are available</td>
<td>d1</td>
</tr>
<tr>
<td>reg_e2empl95_hw</td>
<td>Employment (in hours worked) at NUTS level 2</td>
<td>emp, sal</td>
<td></td>
</tr>
<tr>
<td>reg_e3cvabp95</td>
<td>Gross value added at basic prices at NUTS level 3</td>
<td>100% b1g</td>
<td></td>
</tr>
<tr>
<td>reg_e3empl95</td>
<td>Employment (in persons) at NUTS level 3</td>
<td>100% emp, sal</td>
<td></td>
</tr>
</tbody>
</table>

Where:
- b1g Value added at basic prices
- d1 Compensation of employees
- p5 "Gross capital formation"
- emp "Total employment"
- sal "Employees"

3.3.2 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 significantly across the considered branches. The availability of data for “da15” and “da” is shown in Table 6. Full coverage of NUTS 2 regions for “da” is reached in 13 Member States with more than one NUTS2 region. As “food industry” is most likely to be present in all NUTS2 regions, coverage below 100% may indicate that the regional markets are dominated by a few enterprises, which would cause data protection regulations to apply. If this is the case, then national or regional statistical institutions may neither be entitled to publish the respective information. 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 for the SAMNUTS2 project, but have been exploited in case national statistic do not provide additional information.
### Table 6: Availability of employment data for da15 and DA (in brackets) from sbs_r_nuts03 (in percent of total NUTS2 regions)

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE000000</td>
<td>55 (91)</td>
<td>64 (100)</td>
<td>100 (100)</td>
<td>55 (100)</td>
<td>27 (100)</td>
<td></td>
</tr>
<tr>
<td>BG000000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CZ000000</td>
<td>25 (100)</td>
<td>13 (100)</td>
<td>63 (100)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DK000000</td>
<td>33 (41)</td>
<td>31 (38)</td>
<td>56 (74)</td>
<td>56 (74)</td>
<td>46 (41)</td>
<td>59 (74)</td>
</tr>
<tr>
<td>DE000000</td>
<td>0 (100)</td>
<td>0 (100)</td>
<td>0 (100)</td>
<td>0 (100)</td>
<td>0 (100)</td>
<td>0 (100)</td>
</tr>
<tr>
<td>EE000000</td>
<td>62 (92)</td>
<td>69 (92)</td>
<td>54 (92)</td>
<td>69 (92)</td>
<td>69 (92)</td>
<td>69 (100)</td>
</tr>
<tr>
<td>FR000000</td>
<td>46 (92)</td>
<td>50 (100)</td>
<td>50 (100)</td>
<td>42 (100)</td>
<td>42 (100)</td>
<td>54 (100)</td>
</tr>
<tr>
<td>IT000000</td>
<td>43 (90)</td>
<td>29 (81)</td>
<td>38 (100)</td>
<td>67 (90)</td>
<td>95 (100)</td>
<td>90 (90)</td>
</tr>
<tr>
<td>CY000000</td>
<td>-</td>
<td>100 (100)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LV000000</td>
<td>57 (100)</td>
<td>43 (43)</td>
<td>100 (43)</td>
<td>29 (100)</td>
<td>71 (100)</td>
<td>-</td>
</tr>
<tr>
<td>LT000000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LU000000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HU000000</td>
<td>100 (42)</td>
<td>58 (100)</td>
<td>33 (83)</td>
<td>50 (100)</td>
<td>25 (100)</td>
<td>33 (100)</td>
</tr>
<tr>
<td>MT000000</td>
<td>100 (44)</td>
<td>44 (100)</td>
<td>44 (100)</td>
<td>44 (100)</td>
<td>44 (78)</td>
<td>33 (78)</td>
</tr>
<tr>
<td>PL000000</td>
<td>44 (81)</td>
<td>44 (88)</td>
<td>63 (94)</td>
<td>44 (100)</td>
<td>38 (100)</td>
<td>19 (100)</td>
</tr>
<tr>
<td>PT000000</td>
<td>29 (29)</td>
<td>29 (29)</td>
<td>43 (100)</td>
<td>43 (100)</td>
<td>43 (100)</td>
<td>43 (100)</td>
</tr>
<tr>
<td>RO000000</td>
<td>25 (100)</td>
<td>25 (100)</td>
<td>13 (100)</td>
<td>13 (100)</td>
<td>13 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>SI000000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SK000000</td>
<td>75 (100)</td>
<td>100 (75)</td>
<td>100 (75)</td>
<td>50 (50)</td>
<td>50 (100)</td>
<td>50 (100)</td>
</tr>
<tr>
<td>FI000000</td>
<td>20 (20)</td>
<td>40 (60)</td>
<td>40 (60)</td>
<td>20 (100)</td>
<td>60 (100)</td>
<td>-</td>
</tr>
<tr>
<td>SE000000</td>
<td>63 (100)</td>
<td>63 (100)</td>
<td>38 (100)</td>
<td>13 (100)</td>
<td>0 (100)</td>
<td>38 (100)</td>
</tr>
<tr>
<td>UK000000</td>
<td>51 (84)</td>
<td>38 (89)</td>
<td>49 (95)</td>
<td>59 (95)</td>
<td>51 (62)</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Bold black fonts indicate that a Member State consists of one NUTS2 region. Blue bold fonts indicate full regional coverage for da in 2005. Data availability for DA in brackets.

### 3.3.3 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)” is also evaluated. 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) is investigated on NUTS2 level (Table 7). 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 7  
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)

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CZ000000</td>
<td>88 (88)</td>
<td>88 (88)</td>
<td>88 (88)</td>
<td>88 (88)</td>
<td>88 (88)</td>
<td>88 (88)</td>
</tr>
<tr>
<td>DK000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE000000</td>
<td>95 (95)</td>
<td>95 (95)</td>
<td>95 (95)</td>
<td>95 (95)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>EE000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>ES000000</td>
<td>89 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>IT000000</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
</tr>
<tr>
<td>CY000000</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
</tr>
<tr>
<td>LV000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>LT000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>LU000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>HU000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>NL000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>AT000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>PL000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>RO000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>FI000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>SE000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>UK000000</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
</tbody>
</table>

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.

3.3.4  Summary Regional Datasets from Eurostat

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 have then be used as supplements if available.
3.4 Preliminary Conclusion

The Eurostat datasets represent the fundamental database for the SAMNUTS2 project. Statistical information from Member States is used as exhaustively as possible, but in case no additional information is available, Eurostat datasets has been 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 is done by using the shares of the A60 aggregation with 59 branches in the A16 aggregation with 16 branches. In case national IOTs are not available, then the shares from the year 2000 is 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):

   $AVLN_{,ib,t}^{MS} = \sum_{A60} \left[ G_{A60,b19}^{60,19} \cdot NAMA_{i,A60,t}^{MS} \right] \forall NAMA_{i,A60,t}^{MS} \neq 0$

   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 SAMNUTS2 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)
   - $G_{A60,19}^{60,19}$: Aggregator matrix between 59 NACE and 19 target branches

   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, Table 4), then the respective NAIO tables 1800 and 1900 were used to create this breakdown by calculating the share of the A60 branches in their respective aggregates at A16 level.

   $AVLN_{,ib,t}^{MS} = \sum_{A60} \left[ \sum_{A16} \left[ G_{A60,b19}^{60,19} \cdot NAMA_{i,A16,t}^{MS} \right] \right] \sum_{A60} \left[ G_{A60,16}^{60,16} \cdot NAIO_{i,A60,t}^{MS} \right] \forall NAMA_{i,A60,t}^{MS} = 0 \wedge NAIO_{i,A60,t}^{MS} \neq 0 \wedge NAMA_{i,A16,t}^{MS} \neq 0$

   Where:
   - $G_{A60,16}^{60,16}$: Aggregator matrix between 16 and 60 branches

---

1. For the use of aggregator matrices, please see also Annex 1
NAIO: Indicators form national IOTs

c. Finally, if neither NAMA at A60 level, or NAIO data are available, the IOTs from Rueda-Cantuche et al. (2009) have been used, which provide a full coverage of the Member States in 2000:

\[
AVLN_{i,b19,r}^{MS} = \sum_{A60} G_{i,A60,b19}^{60,19} \cdot \sum_{A16} G_{i,A60,A16}^{60,16} \cdot NAMA_{i,A16,r}^{MS} \\
\sum_{A60} G_{i,A60,A16}^{60,16} \cdot IORC_{i,A60}^{MS} \\
\forall NAMA_{i,A60,r}^{MS} = 0 \land NAIO_{i,A60,r}^{MS} = 0 \land NAMA_{i,A16,r}^{MS} \neq 0
\]

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) are 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 have been 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, 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:

\[
AVLR_{i,b19,r}^{MS,R} = \sum_{A6} G_{i,A6,b19}^{19,6} \cdot REBR_{i,A6,r}^{MS,R} \\
\cdot \sum_{A6} G_{i,A6,b19}^{19,6} \cdot AVLN_{i,b19,r}^{MS}
\]

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

\[\text{For the use of aggregator matrices, please see also Annex 1}\]
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 information and a more elaborate procedure have been developed.

The main conclusion from the evaluation of the Eurostat 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.

### Table 8 Contact details of national statistical offices

<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>Email</th>
<th>Internet</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE000000 Directorate-general Statistics Belgium</td>
<td>Rue de Louvaing 44 · 1000 Bruxelles</td>
<td><a href="mailto:info.stat@economie.fgov.be">info.stat@economie.fgov.be</a></td>
<td><a href="http://www.statbel.fgov.be/">http://www.statbel.fgov.be/</a></td>
</tr>
<tr>
<td>BG000000 National Statistical Institute</td>
<td>2, P. Volov Str.; 1038 Sofia</td>
<td><a href="mailto:info@nsl.bg">info@nsl.bg</a></td>
<td><a href="http://www.nsi.bg/index_e.htm">http://www.nsi.bg/index_e.htm</a></td>
</tr>
<tr>
<td>CZ000000 Czech Statistical Office</td>
<td>Na padesatem 81, 100 82 Praha 10</td>
<td><a href="mailto:infoservis@czso.cz">infoservis@czso.cz</a></td>
<td><a href="http://www.czso.cz/eng/redakce.nsf/i/home">http://www.czso.cz/eng/redakce.nsf/i/home</a></td>
</tr>
<tr>
<td>DK000000 Danmarks Statsistik</td>
<td>Sejrøgade 11, Postboks 2550 · 2100 KØBENHAVN Ø</td>
<td><a href="mailto:dst@dst.dk">dst@dst.dk</a></td>
<td><a href="http://www.dst.dk/">http://www.dst.dk/</a></td>
</tr>
<tr>
<td>DE000000 Statistisches Bundesamt</td>
<td>Gustav-Stresemann-Ring, 11 Postfach 5528 · 65189 WIESBADEN</td>
<td><a href="mailto:info@destatis.de">info@destatis.de</a></td>
<td><a href="http://www.destatis.de/">http://www.destatis.de/</a></td>
</tr>
<tr>
<td>EE000000 Statistics Estonia</td>
<td>15 Endla Street · 15174 Tallinn</td>
<td><a href="mailto:stat@stat.ee">stat@stat.ee</a></td>
<td><a href="http://www.stat.ee/">http://www.stat.ee/</a></td>
</tr>
<tr>
<td>IE000000 Central Statistics Office</td>
<td>Skehard Road – CORK</td>
<td><a href="mailto:dst@dst.dk">dst@dst.dk</a></td>
<td><a href="http://www.dst.dk/">http://www.dst.dk/</a></td>
</tr>
<tr>
<td>GR000000 National Statistical Service of Greece</td>
<td>46, PEIRAIOΣ str. and EPONITON · 185.47 PEIRAIAΣ</td>
<td><a href="mailto:info@statistics.gr">info@statistics.gr</a></td>
<td><a href="http://www.stat.gr/portal/page/portal/ESYE">http://www.stat.gr/portal/page/portal/ESYE</a></td>
</tr>
<tr>
<td>ES000000 Instituto Nacional de Estadística [<em>INE</em>]</td>
<td>Paseo de la Castellana, 183 · MADRID 28046</td>
<td><a href="mailto:info@ine.es">info@ine.es</a></td>
<td><a href="http://www.ine.es/en/welcome_en.htm">http://www.ine.es/en/welcome_en.htm</a></td>
</tr>
<tr>
<td>FR000000 INSEE, Direction générale</td>
<td>18, boulevard Adolphe Pinard · 75675 PARIS · Cedex 14</td>
<td><a href="mailto:dipdiff@istat.it">dipdiff@istat.it</a></td>
<td><a href="http://www.istat.it/">http://www.istat.it/</a></td>
</tr>
<tr>
<td>IT000000 Istituto Nazionale di Statistica</td>
<td>Via Cesare Balbo · 00184 Roma</td>
<td><a href="mailto:dipdiff@istat.it">dipdiff@istat.it</a></td>
<td><a href="http://www.istat.it/">http://www.istat.it/</a></td>
</tr>
<tr>
<td>LV000000 Central Statistical Bureau of Latvia</td>
<td>Lacplesa Street 1 · 1301 Riga</td>
<td><a href="mailto:csh@csh.lv">csh@csh.lv</a></td>
<td><a href="http://www.csh.lv/advisus.cfm">http://www.csh.lv/advisus.cfm</a></td>
</tr>
<tr>
<td>LT000000 Statistics Lithuania</td>
<td>29 Gedimino pr. · 2746 Vilnius</td>
<td><a href="mailto:statistika@stat.gov.lt">statistika@stat.gov.lt</a></td>
<td><a href="http://www.stat.gov.lt/en/">http://www.stat.gov.lt/en/</a></td>
</tr>
<tr>
<td>LU000000 Service Central de la Statistique et des Etudes Economiques</td>
<td>Centre Administratif Herr Werner 13, rue Érasme L · 1468 Luxembourg-Kirchberg</td>
<td>info@stateclu</td>
<td><a href="http://www.stateclu">http://www.stateclu</a></td>
</tr>
<tr>
<td>HU000000 Hungarian Central Statistical Office</td>
<td>Keleti Karoly u. 5-7 P.O. Box 51 · 1525 Budapest</td>
<td><a href="mailto:ksh@ksh.hu">ksh@ksh.hu</a></td>
<td><a href="http://portal.ksh.hu/">http://portal.ksh.hu/</a></td>
</tr>
</tbody>
</table>
To evaluate and compare the gain of information from contacting the Member States statistical institutions, the number of additional data points across all branches is summarized in an indicator. This “informational gain” indicator is constructed based on the following considerations:

- Number of branches in SAMNUTS2: 19
- Number of branch aggregates in Eurostat’s regional branch accounts: 6
- Number of SAMNUTS2 branches in A6:

<table>
<thead>
<tr>
<th>NACE A6 Name</th>
<th>Code</th>
<th>No. of IOTSNUTS2 branches in A6 (na6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Agriculture, hunting, forestry and fishing&quot;</td>
<td>A2B</td>
<td>3</td>
</tr>
<tr>
<td>&quot;Total industry (excluding construction)&quot;</td>
<td>C2E</td>
<td>5</td>
</tr>
<tr>
<td>&quot;Construction&quot;</td>
<td>F00</td>
<td>1</td>
</tr>
<tr>
<td>&quot;Wholesale and retail trade, repair of motor vehicles, motor cycles and personal and household goods; hotels and restaurants; transport, storage and communication&quot;</td>
<td>G2I</td>
<td>3</td>
</tr>
<tr>
<td>&quot;Financial intermediation; real estate, renting and business activities&quot;</td>
<td>J2K</td>
<td>2</td>
</tr>
<tr>
<td>&quot;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&quot;</td>
<td>L2P</td>
<td>5</td>
</tr>
</tbody>
</table>
Under the strong assumption that the regional datasets are consistent with the Eurostat branch accounts, one would need (na6-1) additional data points to construct the full SAMNUTS2 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) is then constructed as follows:

\[
TIG_{i}^{MS,R} = \frac{\sum_{A6} \left( \sum_{b19} \left( G^{19,6}_{b19,A6} \cdot \delta \left( \sum_{AR} G^{19,AR}_{1,b19,AR} \right) -1 \right) \right) - \sum_{AR} \left( \sum_{b19} \left[ G^{19,6}_{1,b19,AR} \right] -1 \right)}}{\sum_{AR} \left( \sum_{b19} \left[ G^{19,6}_{1,b19,AR} \right] -1 \right)}
\]

Where:
- TIG: Informational gain indicator for all branches
- AR: Branches of the regional datasets
- G^{19,AR}_{i}: Aggregator from regional to SAMNUTS2 branches for the respective datasets
- G^{19,6}_{i}: Aggregator from 19 to 6 branches

\[
\delta \left( \sum_{AR} G^{19,AR}_{1,b19,AR} \right) = \begin{cases} 
1 & \text{if } \sum_{AR} G^{19,AR}_{1,b19,AR} > 0 \\
0 & \text{if } \sum_{AR} G^{19,AR}_{1,b19,AR} = 0
\end{cases}
\]

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 SAMNUTS2 branches, and zeroes otherwise. The indicator can range between 0 (no informational gain if compared to Eurostat) and 1, which indicates a full coverage of SAMNUTS2 branches for the respective indicator. TIG measures only potential gain of information as it counts the number of usable branch classifications in the national datasets. An example for the calculation of TIG is provided in Annex 1.

The main purpose of screening the data supply of the national statistical organisations is to retrieve information on the availability of regional IOTs or SAMs. In this respect, the situation is modestly satisfying as survey-based regional IOTs could only be obtained for 12 Comunidades Autónomas (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 Eurostat, screening the national statistical departments’ supply results in a rather mixed picture. For freely available datasets, the informational gain as measured by the TIG-Indicator never reached a level above 0.92 (Figure 2).
The least informational gain is observed for Bulgaria (no gain) and France, Germany, and Ireland, for which the gain ranged below 0.3 points. The reason is the aggregate representation of agriculture, forestry, and fishery in the regional branch accounts. This is the case for 12 Member States (Table 9). As a disaggregation of these branches is crucial for this project, the availability of other indicators that could permit a split has been evaluated and summarized in Table 9.
<table>
<thead>
<tr>
<th>Regional branch accounts</th>
<th>A01, A02</th>
<th>A01 or (A+EAA) or (A_B+EAA +Forest Area)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>BG0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CZ0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DK0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DE0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>EE0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>IE0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>GR0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ES0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>FR0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>IT0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CY0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LV0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LT0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LI0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>HU0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MT0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>NL0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AT0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PL0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PT0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>RO0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SI0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SK0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>FI0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>SE0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>UK0000000</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>12</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

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

It is assumed that the 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 have been 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 9 shows the availability of forestry area data from Eurostat (black) or national statistical organisations (blue). Using these figures to derive average national employment and gross value-added coefficients per hectare of forestry area supplies the needed information, but is likely to yield negative values for fishery if calculated.
residually. Again, the procedures 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.

3.5 Multi-country IOT Projects

Numerous research projects have developed regional IOT or SAM-based models on a regional scale for single Member States, as discussed in the sections on Finland, Hungary, Italy, Poland, or Romania. The following sections provide an overview on three projects on multi-country regional CGEs or other IOT-based models.

3.5.1 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 Eurostat regional branch accounts. 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 SAMNUTS2 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.

3.5.2 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 Eurostat and supplementary information. The breakdown of branches is in general not compatible with the requirements of the SAMNUTS2 project. It could be possible that the DREAM database has meanwhile been updated to a more recent year, but the branch breakdown and the fact that it is an estimated database limits severely the potential use for the SAMNUTS2 project.

3.5.3 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 SAMNUTS2: 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.scnnet.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.

3.5.4 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. The 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.


The deliverables of this project are listed on an (apparently outdated) internet site as follows:

1) Regional Overview on Selection Criteria
2) Report on the Rural Case Study: Bulgaria
3) Report on the Rural Case Study: Romania
4) Report on the Rural Case Study: Slovenia
5) Report on the Rural Case Study: Croatia
6) Report on the Rural Case Study: Greece Regional I-O tables provided detailed information about the regional economic structure (deliverables 7-9).
7) Comparative Development of Rural Economy and Employment in the Balkan Applicant Countries
8) Input-Output Model Description and ReapBalk Common I-O Model
9) Regional Input-Output Tables

3.5.5 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 Eurostat 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. An inventory indicated that data relevant for the SAMNUTS2 project do not exceed the coverage of the Eurostat datasets.
4. Core Accounts

The first step in the development of the targeted regional input-output database is the exhaustion of available datasets for the most relevant accounts to characterise the regional economies. Due to the completeness of the branch-indicators “gross value-added” (b1g) and “employment” (emp), the regional branch account database from Eurostat (ESTAT) appears to be the logical anchor point for the subsequent steps. The main challenge from the structure of this database is the classification of economic branches according to the A6 format, which distinguishes six branches. As a default solution, the A6 branches are expanded into the targeted SAMNUTS2 format (B19) using the national shares of B19 in the A6 aggregates. Information from national statistical organisations is then added. A crucial implication of this approach is that A6-values from ESTAT have priority over A6-values from the NSOs. Alternatively, it would be possible to prioritise either NSO values or to calculate average values across all data sources. The advantage of the first alternative is that the ESTAT regional branch account figures are consistent with other national databases, particularly the national input-output figures, which is not guaranteed by the NSO data. The latter alternative would generate an additional datasets which does not necessarily comply with any other available information. As such, the qualitative decision which database should be used as benchmark is in favour of the ESTAT datasets. In the case of Spain, for which a most exhaustive database could be obtained, the NSO figures are in general in line with the ESTAT figures (Figure 3). Notably, the NSO figures are on average almost 9% higher than the ESTAT figures.

Figure 3  Comparison of Regional Branch Account Data (RAMA) from NSO and ESTAT for Spain

![Graph showing comparison between NSO and ESTAT figures for regional branch account data in Spain. The graph includes a linear regression line with the equation NSO = 1.0885ESTAT and R² = 0.9979.](image)

Source: NSO Spain [various departments], Eurostat, own presentation

Comparing the figures on a region-by-region scale (Figure 4), it can be seen that branch entries (A2B in Figure 4) deviate across sources, but also that the totals given by the NSO are not consistent with the corresponding ESTAT entries. Based on this
observation, and considering the fact that regional account figures from Eurostat are also in general consistent across countries, the decision to use Eurostat branch account as boundary condition for the further compilation steps appears to be reasonable.

Figure 3 Comparison of Gross Value-added for A6 Aggregates from NSO and ESTAT for "Cornwall and Isles of Scilly" (UKK3)

<table>
<thead>
<tr>
<th>Source</th>
<th>ESTAT</th>
<th>NSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td>mio_nac</td>
<td>mio_nac</td>
</tr>
<tr>
<td>Region</td>
<td>UKK30000</td>
<td>UKK30000</td>
</tr>
<tr>
<td>Year</td>
<td>2005</td>
<td>2005</td>
</tr>
<tr>
<td>Indicator</td>
<td>d1</td>
<td>d1</td>
</tr>
</tbody>
</table>

![Figure 3](image)

Sources: NSO (UK) and ESTAT

4.1 Expanding and Combining ESTAT and NSO Datasets

The following sections will provide an overview on the steps necessary to build a complete set of core accounts. We will start with the default solution to expand the A6 branch aggregates by using national shares. Then, the procedure to derive the remaining core accounts will be outlined.

Having made the decision to use ESTAT branch accounts as benchmark, the next question is how the available data can be combined. In general, there are three types of data available: First, the mentioned ESTAT regional branch accounts (RAMA\textsuperscript{ESTAT}), the national branch account tables (NAMA\textsuperscript{ESTAT}) and the branch accounts from the national statistical organisations (RAMA\textsuperscript{NSO}).

Figure 5 gives an example for the appearance of the NAMA and RAMA datasets for Cornwall and Isles of Scilly. It also 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 5: RAMA\textsuperscript{NSO} and NAMA\textsuperscript{ESTAT}). It has to be noted here that in some cases Members States' statistical organisations report regional data in national currency (here: Pound Sterling), which makes an immediate comparison also difficult as a decision on the appropriate aggregate average exchange rate has to be made.
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. The described procedures rely on the definition of appropriate aggregator matrices. The described procedures rely on the definition of appropriate aggregator matrices (Figure 6).
Using this aggregator matrix, the national shares of $b_{19}$ entries in $A_6$ are calculated for the national datasets (NASH):

$$NASH_{b_{19},j}^{ESTAT, MS} = \frac{NAMA_{b_{19},i,t}^{MS}}{\sum_{A_6} \left( G_{b_{19},A_6}^{19,6} \cdot \sum_{b_{19}} [G_{b_{19},A_6}^{19,6} \cdot NAMA_{b_{19},i,t}^{MS}] \right)}$$

Similarly, the shares of the regional datasets in $A_6$ are calculated (RASH):

$$RASH_{b_{19},j}^{NSO,R} = \frac{RAMA_{b_{19},i,t}^{NSO,R}}{\sum_{A_6} \left( G_{b_{19},A_6}^{19,6} \cdot \sum_{b_{19}} [G_{b_{19},A_6}^{19,6} \cdot RAMA_{b_{19},i,t}^{NSO,R}] \right)}$$

Combination into a core-account dataset $Y$ is then achieved by a simple expansion of the $RAMA_{b_{19}}^{ESTAT}$ datasets and multiplication with the shares, depending on the availability of datasets from NSO:

$$Y_{b_{19},j}^{MS,R} = \begin{cases} \sum_{A_6} [G_{b_{19},A_6}^{19,6} \cdot RAMA_{A_6,d,t}^{ESTAT,R}] \cdot RASH_{b_{19},j}^{NSO,R} & \forall RAMA_{b_{19},i,t}^{NSO,R} \neq 0 \\ 0 & \forall RAMA_{b_{19},i,t}^{NSO,R} = 0 \end{cases}$$
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 SAMNUTS2 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.

Comparing the deviation between NASH and RASH figures for Italy and Spain (Figure 6), 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. 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. 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 the A16 level a more reliable set of core accounts. In the case of the share of D in C2E, D (at least in Spain) has a particularly high value (Figure 7), such that biased estimates for C and E do not have a huge importance for the regional economies.
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 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 would significantly improve the picture of the regional economies.
4.2 Derived Core Accounts

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 Eurostat. To complete the regional core accounts and to derive regional SAMs, it is necessary to derive additional core indicators, such as gross output by branch, total intermediate demand by branch, or taxes on activities paid by branch. As shown in the inventory, these figures are available in some cases, but it is necessary to formulate a default strategy when 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, 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. The outcomes of these computations are compared with recorded figures from the Spanish NSO in Figure 8 and Figure 9. With respect to gross output (p1), the largest deviations occur for Fisheries, Forestry, and most notably Fuel industries (Figure 8).
Concerning intermediate demand (p2), the largest deviations are observed for Forestry, Fuel industry and Fisheries (Figure 9) – and to a larger extent than for the case of p1 as the intermediate demand of Forestry may be more than ten times larger as indicated by the national figures. These observations make clear that Forestry, Fisheries, and Fuel industry have a different inner structure on regional scale as if 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.
Figure 9  Deviation between National and Regional Coefficients for Intermediate Inputs (p2)

Largest deviations for AA02 B DF

A16 branches
5. Derivation of IO-Tables

After the derivation of regional core accounts, in the next step we derive the SAMs. As in previous steps, regional branch account data are combined with national coefficients, but with two major differences: First, the inventory showed that regional transaction matrices (the core SAM datasets) are in general not available. Second, it would be theoretically inconsistent to assume regional input coefficients to be equal to the national ones when distinguishing the intermediate demand by their origin. A highly specialised regional economy may cause the development of other industries in the same region that provide the needed inputs, such that the share of imported inputs would be significantly smaller than on national scale. In the context of the compilation of regional SAMs, this problem has been addressed frequently in the literature, leading to the establishment of a widely accepted sequence of procedural steps, the “Generation of Regional Input-Output Tables” procedure or GRIT. Next section summarizes the most relevant steps within GRIT and highlights the stages where the SAMNUTS2 approach presented here deviate. As GRIT draws on the application of location quotients and gravity models, the subsequent sections address these topics.

5.1 Generation of Regional Input-Output Tables: GRIT

In general, GRIT consists of a sequence of steps to derive regional IOTs based on national tables and regional account data, similar to the objective of the SAMNUTS2 project. The main approach is to incorporate superior information in the most efficient manner, whenever it becomes available. In this context, larger attention is devoted to larger coefficients than to smaller ones (“free from significant error”). It has to be noted that GRIT was developed in the context of a particular type of IOTs, which is not in all respects comparable to the IOT structure intended here. For instance, GRIT includes a variety of methods to derive the intermediate demand from sectoral transactions, while treating intermediate demand from other sectors as just one aggregate (comparable to the structure of the NAIO Tables 18). The SAMNUTS2 structure deviates from this as intermediates from other sectors are included in the form of sub-matrices, thus providing more information. Also, the typical IOTs developed with GRIT treats final consumption as a single aggregate, often calculated residually. Again, this is not sufficient for the needs of SAMNUTS2. Nevertheless, it is worthwhile to consider in more details the general approach of the inclusion of superior information (here: NSO data), as done in the next section.

GRIT Phase 1: Adjustment of National IOT

In Phase 1 of the GRIT procedure, a regional transaction matrix is derived based on the national coefficients. In Step 1, a national IOT is selected, which should be as disaggregated as possible. Intra-sectoral transactions are netted out to avoid the overestimation of regional coefficients. However, this step does not appear to be applicable to the SAMNUTS2 project as we have a clear interest in the inter-agricultural transactions. Step 2 involves the updating of the selected SAM to meet the target year
(2005 in the case of SAMNUTS2). Step 3 finally involves adjustments for international trade in case the underlying national table features only the intermediate demand from domestic origin. This is not the case in SAMNUTS2 as we use NAIO tables 1800 and 1900.

**GRIT Phase 2: Adjustment for Regional Imports**

Regional input coefficients and import coefficients are derived in this phase using national technical coefficients, again in case of a SAM structure that features only inputs from domestic origin while treating imported intermediates as an aggregate row. Consequently, Step 4 involves the calculation of non-competitive imports in case national sectors do not exist in a particular region (very important in the case of Fisheries and Mining sectors). The corresponding rows of coefficients are removed from the regional intermediate demand sub-matrix and allocated to the regional imports. In contrast, Step 5 involves the calculation of competitive imports. Regional input coefficients and regional (competitive) import coefficients are derived from location quotients (LQ).

**GRIT Phase 3: Regional Branches**

Within this phase of GRIT, superior data (like NSO statistics) is included in two stages. First, in Step 6, superior data is inserted with as much detail as possible, prior to the potential aggregation to the targeted branches in the following step 7. In practical terms for the SAMNUTS2 project, this is not possible as the available regional IOTs or SAMs are the most detailed source of information, such that no comparable information can be obtained at this stage. The decision on the regional branches has been also pre-defined for this project (b19 as target format). Aggregated (here: at b19 level) superior data is then supplemented within Step 8.

**GRIT Phase 4: Derivation of Prototype Transactions Table**

Based on the coefficient matrix derived in the previous steps, a regional transaction matrix is derived in basically four steps. Step 9: Derivation of initial transactions table by the multiplication of coefficients in each column with the derived core-account data. In the most ideal case, this would refer to regional gross output (p1). Then, in Step 10, adjustments of this prototype table are made by adding final consumption and primary inputs. Usually, three components of final demand are considered (household consumption, exports and other final demands). However, SAMNUTS2 requires at least a further distinction of regional and national governmental demand. Step 11 involves the further aggregation in case uniform tables are required. This is true for SAMNUTS2, but is done in our framework at an earlier stage. In case the regional IOTs are needed for multiplier models, the Leontief inverses may be calculated the next Step 12.
GRIT Phase 5: Derivation of Final Transactions Table

After the generation of a prior transaction table, further information can be included, e.g. expert’s knowledge on the reliability of the prior tables. Steps 13 to 15 involve these final adjustments and balancing of the prior tables.

GRIT Summary

The outline of GRIT-Phases in the previous sections makes clear that GRIT is not so much a set of well-defined rules but rather a set of guidelines on the most efficient use of additional information during various stages of the compilation procedure. Consequently, numerous modifications exist to meet the requirements of the respective project. It appeared that Location Quotients are a method applied at crucial points of the compilation stages. It seems that GRIT aims at the direct compilation of balanced IOTs, e.g. by calculating final uses as residuals. As such, it largely ignores the use of SAM estimation procedures.

5.2 Location Quotients

In the outline of the GRIT procedure, Location Quotients (LQs) 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. In the context of the SAMNUTS2 project, the question is how the core branch account data can be used to create the best a-priori estimate for the unknown regional SAMs. As the core accounts refer to the expenditure columns of the productive sectors, for the time being the expenditures of the institutions (households, government, and savings-investment account) are neglected and the following considerations refer only to the branch accounts, denoted as $S^A$. $S^A$ consists of four sub-matrices, distinguishing the origin of the intermediates: $Ar$ for regional origin, $Ad$ for domestic, and $Ai$ for imported. The total intermediate demand for branch is denoted $A$, while gross-total intermediate demand ($p2$) is denoted $At$. If gross output by branch is known from the ESTAT or NSO RAMA data, it would be possible to compute column-coefficients for $S^A$ for the national tables and multiply those with the regional supply data to create a set of regional prior matrixes. Given the information available from ESTAT RAMA and NAMA, a national coefficient matrix ($K$) can also be obtained by dividing each element of $S^A$ by the sectoral value-added ($VAD$, b1g in ESA notation).

$$S^A \equiv \frac{A}{VAD}$$

The regional tables $S^R$ may now be computed by multiplying the regional value-added data with a regional coefficient matrix $K^R$, which would be equal to the national one if no other information is available:
Location Quotients aim at the adjustments of the national coefficients according to the weight of a specific branch 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 (Flegg and Webber, 1995 or Bonfiglio and Chelli, 2008). The adjustment of the national coefficient matrix is then performed by multiplication with the respective LQ matrix:

\[ S_{i,b}^R = VAD_{i,b}^R \cdot K_{i,b}^R, \text{ with } K_{i,b}^R = K_{i,b} \]

The usage of the un-adjusted national coefficients is therefore a special case with \( LQ^R = 1 \) for all sectors and regions. There is a variety of location quotients discussed in the literature, which are summarized in 9. The Simple Location Quotient (SLQ) is calculated for the rows of the target matrix by relating the regional employment shares of a specific branch with the national employment shares. The cross-industry LQ (CILQ) is a further development as it also takes into account the relative weight of potentially supplying sectors. The two LQ suggested by Round (1978) (RLQ, ELQ) are a refined formulation of the CILQ by avoiding results larger than unity by using a log of base 2.

Of particular interest here are the two Flegg location quotients (FLQ and AFLQ) as they do not only depend on sectoral national and regional employment data (RE, NE), but also on the choice of the parameter \( \lambda (0 \leq \lambda < 1) \) that introduces an "element of flexibility" (Flegg et al. 1995). The choice of \( \lambda \) depends on empirical considerations (Flegg et al. 1995). Statistical properties were investigated by Bonfiglio and Chelli (2008), finding higher values for \( \lambda \) to yield better results based on a Monte Carlo Analysis.
Table 10  Location Quotients

<table>
<thead>
<tr>
<th>Name</th>
<th>Acronym</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple location quotient</td>
<td>SLQ</td>
<td>$SLQ_{b}^R = \frac{RE_{b}/TRE}{NE_{b}/TNE}$</td>
</tr>
<tr>
<td>Cross-industry location quotient</td>
<td>CILQ</td>
<td>$CILQ_{b}^R = \frac{SLQ_{b}^R}{SLQ_{b}^N}$</td>
</tr>
<tr>
<td>Round location quotient</td>
<td>RLQ</td>
<td>$RLQ_{b,b'}^R = \frac{SLQ_{b}^R}{\log_2(1 + SLQ_{b}^R)}$</td>
</tr>
<tr>
<td>Alternative Round location quotient</td>
<td>ELQ</td>
<td>$ELQ_{b,b'}^R = \frac{\log_2(1 + SLQ_{b}^R)}{SLQ_{b}^R}$</td>
</tr>
<tr>
<td>Flegg location quotient</td>
<td>FLQ</td>
<td>$FLQ_{b,b'}^R = CILQ_{b,b'}^R \cdot \left[ \log_2 \left( 1 + \frac{TRE}{TNE} \right) \right]^5$</td>
</tr>
<tr>
<td>Augmented Flegg location quotient</td>
<td>AFLQ</td>
<td>$AFLQ_{b,b'}^R = FLQ_{b,b'}^R \cdot \left[ \log_2 \left( 1 + SLQ_{b}^R \right) \right]$</td>
</tr>
</tbody>
</table>

Where:

- RE: Regional Employment by sector
- TRE: Total regional employment
- NE: National Employment by sector
- TNE: Total national employment

Sources: Adapted from: Bonfiglio and Chelli (2008), Flegg et al. (1995)

The performance of the alternative location quotients is tested in the following for the available Spanish SAMs. The root mean squared error (RMSE) is calculated for each LQ and divided by the RMSE of the default setting LQ=1. The results are shown in Figure 10. 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 $R_i$ of 0.4, which is well in line of the findings of Bonfiglio and Chelli (2008), who observed best performance of AFLQ with a $R_i$ between 0.3 and 0.5. The implication of these findings is clear: For the derivation of the coefficients to derive intermediate demand from regional origin (Ar), the augmented Flegg LQ with $R_i$ in the range of 0.3 to 0.5 is used. Concerning the derivation of the non-regional origin of inputs, an alternative approach is discussed in the next section.
5.3 Foreign or Domestic Origin of Used Commodities

Only the derivation of the Ar sub-matrix of intermediate demand from regional origin has been considered. The SAMNUTS2 database should include sub-matrices for intermediates from domestic and imported origin (Ad and Ai respectively). 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. Tinbergen (1962) pioneered the use of gravity equations in empirical specifications of bilateral trade flows in which the volume of trade between two countries is proportional to the product of an index of their economic size, and the factor of proportionality depends on measures of “trade resistance” between them. The equation below is a typical specification of this gravity model, as discussed by Bertrand (1989):

\[
PX_{b,i,j} = \psi_0 \left( \frac{Y_i}{Y_j} \right)^{\psi_i} \left( \frac{Y_j}{Y_i} \right)^{\psi_j} \\
\left( \frac{Y_i}{L_i} \right)^{\psi_i} \left( \frac{Y_j}{L_j} \right)^{\psi_j} \left( D_{b,i,j} \right)^{\nu_b} \left( A_{i,j} \right)^{\nu_a} \epsilon_{b,i,j}
\]

Where:
- \( PX_{b,i,j} \): Monetary flow for commodity b from region i to region j
- \( Y_i \) (\( Y_j \)): Income in region i or j
- \( L_i \) (\( L_j \)): Population in region i or j
- \( D_{b,i,j} \): Distance measure between economic centres in regions i or j, e.g. transportation cost for commodity b
- \( A_{i,j} \): Any other factor aiding or resisting trade between i and j
- \( \epsilon_{b,i,j} \): Error term
Parameters to be estimated

Modifications of the original model are discussed in various applications (e.g. Helpman et al. 2008, Bergstrand 1989), but they all rely on observation for the inter-regional trade flows PX. It may be possible to obtain valid estimates for these from the literature and use to derive regional trade flows of the NUTS2 regions of the EU. The problem is simplified on the one hand by the fact, that bi-lateral trade flows do not have to be estimated, but rather flows between region i and national or external markets. On the other hand, the problem is aggravated by the lack of distance measures between regional and domestic or foreign markets (D). A gravity model may help to derive total intermediate demands from domestic or imported origin (column-sums of Ad and Ai), but not to derive the full sub-matrices (Figure 11).

Figure 11   SAMNUTS2 Sub-matrices and Methods for Derivation

To derive the full set of needed sub-matrices, we derive total intermediate demand (A in Figure 11). In the absence of additional information (LQs only help to determine the origin of the intermediates from the same region), we use national total intermediate demand coefficients to derive sub-matrix A:

\[ A^R_{b19,b19'} = A^R_{b19'} \cdot \frac{A^MS_{b19,b19'}}{A^MS_{b19'}} \]

When plotting the thus derived entries of the total intermediate demand sub-matrix against the observed values for Spain (Figure 12), it appears that the derived values are on average 10% lower than the recorded values. The high measure of determination (R^2) is caused by the comparatively few entries with magnitudes above 5000 millions of Euro. When limiting the sample only to entries below 5000 millions of Euro, the derived values are on average 3% higher than recorded, with a measure of determination of 0.7.
In case the Transtools database appears to be suitable for an efficient estimation of the column-sum of Ai (Ait) in the NUTS2 regions, base values for Ai can be derived by using national coefficients as in the case of A:

\[ A_{b19,b19}'^{R} = A_{b19}'^{R} \cdot \frac{A_{b19,b19}'^{MS}}{A_{b19}'^{MS}} \]

Where Ait denotes the column-sum of the sub-matrix for imported intermediates. For illustrative purposes, the recorded values for Ait from the Spanish national organisations were used to examine the applicability of the national coefficients as proposed above (Figure 13). As in the case of total intermediate demand, the comparatively high \( R^2 \) of 0.93 is explained by few values with comparatively high magnitudes. The sub-sample of values below 2000 million Euro results in a \( R^2 \) of 0.62, the average deviations is then -21%.

Although the goodness of fit of the thus derived a-priori values for the two sub-matrices A and Ai are not entirely satisfying, particularly because of the systematic deviation between derived and recorded values, it has to be noted that the correlation is still high.
An alternative for the use of the Transtools database to derive Ait can be the usage of national coefficients for the derivation of core accounts. A preliminary test is illustrated in Figure 14. The largest deviations from regional to national coefficients can be observed for Hotel and Restaurant services (H000), lower but still significant ones for Construction (F000), Real Estate (K000) and Education (M000). All these branches are not distinguished in the commodity classification of the Transtools database, so it remains questionable if an improvement can be expected from there.
Having calculated the demand for regionally produced intermediates (Ar) based on location quotients (section 4.2), and A as well as Ai as illustrated above, the sub-matrix Ad may now be computed residually:

\[ Ad_{b19,b19}^R = A_{b19,b19}^R - Ai_{b19,b19}^R - Ar_{b19,b19}^R \]

The high variance of A, Ar, and Ai have to be taken into account to avoid negative a-priori values for Ad.

### 5.4 Taxes, Transfers, and Institutions

Having generated a-priori entries for the production accounts of the target SAMs, the next step is the distribution of value-added 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 (b2gb3g), while the local government receives the indirect taxes Ti paid by the production accounts (d29md39) and employers’ social contributions (d12). The thus obtained aggregate private income (d11 + b2gb3g) is then used to determine direct taxes and transfers based on national tax and transfer rates as obtained from the NASA dataset.

These direct taxes and transfers are then booked as revenues or expenditures, respectively, of the local government account. The difference between aggregate private income M (d11 + b2gb3g) and net direct taxes Td paid by private households is the disposable private income Md, which can either be saved or used for consumption expenditure. 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 Ct:

\[
Td_{s14s15}^R = M_{s14s15}^R \left( \frac{Td_{s14s15}^{MS}}{M_{s14s15}^{MS}} \right)
\]

\[
Md_{s14s15}^R = \left( M_{s14s15}^R - Td_{s14s15}^R \right)
\]

\[
Ct_{s14s15}^R = Md_{s14s15}^R \left( \frac{Ct_{s14s15}^{MS}}{Md_{s14s15}^{MS}} \right)
\]

Where “s14s15” denotes “households and non-profit organisations serving households”.

Revenues of the local government consist of collected direct and indirect net taxes (Td and Ti). As in the case of private households, we assume that governmental
expenditures have a uniform pattern across regions and apply national consumption share to derive total governmental consumption:

\[ Ct^{R}_{s1313} = \left( Td^{R}_{s1313} + Ti^{R}_{s1313} \right) \left( \frac{Cl^{MS}_{s1313}}{Td^{MS}_{s1313} + Ti^{MS}_{s1313}} \right) \]

Where “s1313” denotes “local government” and “s1311” “central government”

The next item to be derived is aggregate regional investment (It). The Eurostat dataset “reg_e2gfcf” provides information for 175 NUTS2 regions, in some cases overlapping with data from NSO RAMA datasets. In case neither pieces of information are available, we resort to the usage of investment shares in national income:

\[ I^R = \begin{cases} 
I_{NSO}^R & \forall I_{NSO}^R \neq 0 \\
I_{ESTAT}^R & \forall I_{ESTAT}^R \neq 0 \land I_{NSO}^R = 0 \\
\frac{GDP^{R}}{GDP^{MS}} I_{ESTAT}^R & \forall I_{ESTAT}^R = 0 \land I_{NSO}^R = 0 
\end{cases} \]

Having derived the regional exports as described in section 4.3, it is now possible to calculate regional savings residually and to close the sub-matrix on taxes and transfers.

### 5.5 Demand-side of the IO-Tables

The previous steps to generate the prior SAM entries focussed on the production and transfer sides, including intermediate demand and payment for fixed factors of production. The demand-side consists of private and governmental final consumption (C and Cw in the table below) as well as investment demand (I and Iw) and exports (E and Ew).

Based on the total values for these items (It, Ct, Et) derived in the previous section, it is now possible to compute prior entries for the target SAM. Clearly, it would be optimal if such information could be obtained directly from NSO, but as could be seen in the inventory, unlike production-side information, consumption side information is not always available. A notable exception is the “reg_e2gfcf” dataset from Eurostat, which provides gross-fixed capital formation (p5) for 175 NUTS2 regions in the A6 branch classification. For governmental and private final consumption (C), it is again necessary to rely on national expenditure shares. The procedure is similar to the procedure to derive intermediate demand from different origins. Starting with total
final demand for the considered institutions d ("government", "household"), the total consumption vectors Ct are distributed across the branches of the regional economies based on national shares:

\[ C_{b19,d}^R = C_d^R \cdot \frac{C_{MS}^{b19,d}}{C_d^{MS}} \]

Where index d denotes institutions with final consumption expenditures (s14s15, s1313, s1311)

Investment demand is computed accordingly:

\[ I_{b19}^R = I_d^R \cdot \frac{I_{MS}^{b19}}{I_d^{MS}} \]

In case regional investment demand is available from the Eurostat "reg_e2gfcf" dataset in the A6 classification scheme, we follow a procedure similar to the derivation of core accounts in section 3.2 by first calculating the shares of the b19 branches in A6 branches

\[ I_{b19}^R = \sum_{A6} \left[ G^{19,6}_{b19,A6} \cdot RAM_{ESTAT,R} \right] \]

Following the steps described in section 4.3, we split now aggregate demand (C, I) according to its origin (Cw, Iw, with w={r,d,i}), starting with final demand from regional origin Cr and Ir based on location quotients. Due to the fact that final demand, unlike intermediate demand, is not a bxb but rather a bx(d+1) matrix, AFLQ is not usable here and we have to apply the simple location quotient SLQ. The reasoning behind this procedure is based on the assumption that regions with an over-proportionally high share of some branches also supply the regional markets for final consumption and investment goods:

\[ C_{b19,d}^R = SLQ_{b19} \cdot C_d^R \cdot \frac{C_{MS}^{b19,d}}{C_d^{MS}} \]

\[ I_{b19}^R = SLQ_{b19} \cdot I_d^R \cdot \frac{I_{MS}^{b19}}{I_d^{MS}} \]

In the case of final consumption from imported origin, we assume that national shares on imported final uses can be applied to regional scales:

\[ C_{b19,d}^R = C_{b19,d}^R \cdot \frac{C_{MS}^{b19,d}}{C_{MS}^{b19,d}} \]

\[ I_{b19}^R = I_{b19}^R \cdot \frac{I_{MS}^{b19}}{I_{MS}^{b19}} \]

The derivation of final consumption from domestic origin follows the same logic, but takes into account the adjustment for regional origin:

\[ Cd_{b19,d}^R = (C_{b19,d}^R - C_{b19,d}^R) \cdot \frac{Cd_{MS}^{b19,d}}{Cd_{MS}^{b19,d}} \]
With the completion of the demand-side, the prior regional SAM is now fully populated. However, so far nothing guarantee that SAMs are balanced with respect to the regional accounting structure, nor that they add to the national tables. For this purpose, a balancing procedure, which is described in the next section, is applied.
6. Balancing Procedure

6.1 General Methodological Considerations

The previous steps to generate the prior SAM entries focussed on the production and transfer sides. The prior database developed throughout the previous steps does not necessarily fulfil the requirement that regional expenditures equal regional revenues, nor that the regional SAMs add up to the national ones. The core problem is to compile a number of regional \((r, (i,j))-\)dimensional tables \(S\) that include all monetary transactions between productive sectors, commodity and factor markets, and institutions within the national and regional economies.

The first requirement is the equality of row- and column-sums, formalised as:

\[
\sum_{j} S_{r,i,j} = \sum_{j} S_{r,j,i}
\]

Secondly, the regional SAM has to add up to the respective national table \(S^{MS}\):

\[
\sum_{r} S_{r,i,j} = S^{MS}_{i,j}
\]

Furthermore, the targeted tables \(S\) has to be consistent with control-totals \(\Gamma\) (e.g. macro-economic indicators like gross value-added or other information obtained from either NSO or Eurostat, which correspond to certain sub-totals of \(S\)).

\[
\sum_{j} \left[ \sum_{i} G_{k,j} S_{r,i,j} \right] = \Gamma_{r,k,l}
\]

Where \(G\) is the aggregator matrix mapping the elements of \(S\) into the corresponding control totals.

Based on the available information described in the previous sections, it is possible to create a set of prior matrices \(S^0\), which are not balanced because of inconsistencies in the used datasets or because of the different procedures to derive the prior entries, and do not comply with the available control totals:

\[
\sum_{j} S^0_{r,i,j} \neq \sum_{j} S^0_{r,j,i}, \sum_{j} S^0_{r,i,j} \neq S^{MS}_{i,j}, \text{ and } \sum_{j} \left[ \sum_{i} G_{k,d} S^0_{r,i,j} \right] \neq \Gamma_{r,k,l}
\]

As it is rarely possible to observe the true values of \(S\) but rather the distorted ones \(S^0\), a robust estimator must be used to consolidate the SAM. In general, \(S\) should be as close as possible to \(S^0\), but should fulfil at least conditions (1) and (2) and also condition (3) if control totals are available. Having identified these boundary conditions, the next step is to specify a statistical criterion that allows estimating \(S\) which is as close as possible to \(S^0\) subject to these conditions.

Round (2003) provides an overview on SAM and IO estimation and balancing approaches, including Generalised Cross Entropy (GCE, Golan et al. 1994, Robinson et al. 2001), Stone-Byron (Stone 1977, Byron 1978), and RAS (Bacharach 1970). In practice,
many variants and combinations of the basic methods are applied, some building on column-coefficients (Breisinger 2005, Robinson et al. 2001, and various related applications at the International Food Policy Research Institute (IFPRI)), some on the actual table entries (e.g. Mueller 2006, Nakamura 1998). Control totals on sub-matrices of the tables are sometimes imposed rigorously (Robinson et al. 2001), or can be associated with an error term (Breisinger et al. 2005). Ultimately, the choice of the appropriate procedure to estimate balanced tables depends on the available data and the way in which the researcher wants to express the trust he has in the information at hand. For the SAMNUTS2 project, the procedure builds on the actual entries rather than column coefficients to ensure compliance with control-totals and prior information available at regional level. Formally, the objective function $Z$ of the estimation model should be the minimization of a distance measure $d$ between $S$ and $S^0$:

$$
\min_{S_{ij}} Z = d\left( S_{r,i,j} \mid S^0_{r,i,j} \right)
$$

In a project on development of large-scale, economy-wide databases (AgroSAM, Mueller et al. 2009), the distance measure between $S$ and $S^0$ (equation (5)) is built on the declaration of a multiplicative distortion term with an optimum value of 1 and a minimum of 0. The decision to use multiplicative instead of additive distortion terms is based on the consideration that they guarantee the preservation of signs when defined as strictly positive, and maintain zero-entries. Instead, it is also possible to express $S$ in terms of support points $B$ and associated weights $W$:

$$
S_{r,i,j} = \sum_s B_{s,r,i,j} W_{s,r,i,j}
$$

The support points can be chosen according to an assumed variance around the prior values $S^0$. Additionally, it has to be imposed that the weights range between 0 and 1 and add up to 1:

$$
\sum_s W_{s,r,i,j} = 1, \quad 0 \leq W_{s,r,i,j} \leq 1
$$

The GCE approach permits the inclusion of prior weights $W^0$, which are chosen such that the sum-product of $W$ and $B$ equals $S^0$ when the GCE objective function is at its minimum:

$$
S^0 = \sum_s B_{s,r,i,j} W^0_{s,r,i,j}
$$

Weights $W$ and prior weights $W^0$ enter the GCE objective function $Z$ in the following way:

$$
\min_{W_{s,r,i,j}} Z = \sum_{s,r,i,j} W_{s,r,i,j} \ln \left( \frac{W_{r,i,j}}{W^0_{r,i,j}} \right)
$$

Objective function (9) is then minimized subject to constraints (1) to (3) and (6) to (8). A GCE application requires for each estimate matching priors and weights, and in the case of more than two supports, additional constraints. The resulting high number of variables and constraints may cause computational difficulties for large-scale datasets during estimation and increase estimation time. Additionally, the implicit posterior
density depends on the interaction between the choice of supports, the a priori probabilities and the entropy criterion. Both problems were addressed by Heckelei et al. (2008) and Witzke and Britz (2005) by motivating a Highest Posterior Density (HPD) estimator which refrains from discrete support points but still allows to express confidence by using informative priors on the variance of each estimate.

Translating the discussed estimation problem into the HPD framework yields:

\[
\min_{S_{r,i,j}} \left( \sum_{r,i,j} \left( \frac{S_{r,i,j} - S_{0,r,i,j}}{S_{r,i,j}^\sigma} \right)^2 \right)
\]

Where \( S^u \) denotes the standard deviation of \( S \). Objective function 10 is then minimized subject to constraints (1) to (3), which is similar to the Stone-Byron approach, although motivated by Heckelei et al. (2008) from a different perspective. This setting creates a substantially smaller computational burden, not only because of the smaller number of constraints and variables, but also because of the linearity of the first derivative of (10) with respect to \( S \).

Because of this reason, we resort to the application of a HPD rather than a GCE procedure for the SAMNUTS2 database.

### 6.2 Balancing Approach for Core Accounts

The magnitude of the task to balance 271 SAMs with numerous internal (accounting identities in each table) and external (adding-up to national total) constraints made it difficult to implement the balancing procedure as a single optimization problem. Apart from long solving time for each Member State, the tractability of potential sources for infeasibilities and implausible values is compromised due to the large number of variables and equations in each problem. The balancing procedure is thus decomposed into several steps, following the steps involved in the construction of the datasets. Starting point is the balancing of national core accounts for all target sectors and sector-aggregates. We minimize deviations from observed or derived core account entries \( S^0 \) for the b19 target sectors and the respective sector aggregates at A6 and A16 levels:

\[
\min_{MS_{r,i,j}} \left( \sum_{MS_{r,i,j}} \left( \frac{S_{MS_{r,i,j}} - S^0_{MS_{r,i,j}}}{S_{MS_{r,i,j}}^\sigma} \right)^2 \right)
\]

Where \( S^u \) again denotes the standard deviation of the balanced core account entries \( S \), index \( i \) refers here to the items considered as core accounts (Table 11) and index \( j \) to all economic branches and -aggregates (b19, A6, A16). In this first step, the index MS identifies the 27 Member States at national level (NUTS0).
Table 11  Core Accounts in First Balancing Step

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2pp</td>
<td>&quot;Total intermediate consumption/Final use at purchasers prices&quot;</td>
</tr>
<tr>
<td>d1</td>
<td>&quot;Compensation of employees&quot;</td>
</tr>
<tr>
<td>d11</td>
<td>&quot;Wages and salaries&quot;</td>
</tr>
<tr>
<td>d12</td>
<td>&quot;Employers social contributions&quot;</td>
</tr>
<tr>
<td>k1</td>
<td>&quot;Consumption of fixed capital&quot;</td>
</tr>
<tr>
<td>b2npb3n</td>
<td>&quot;Operating surplus, net&quot;</td>
</tr>
<tr>
<td>b1g</td>
<td>&quot;Value added at basic prices&quot;</td>
</tr>
<tr>
<td>p1</td>
<td>&quot;Output at basic prices&quot;</td>
</tr>
<tr>
<td>Emp</td>
<td>&quot;Total employment&quot;</td>
</tr>
<tr>
<td>Sal</td>
<td>&quot;Employees&quot;</td>
</tr>
<tr>
<td>Self</td>
<td>&quot;Self-employed&quot;</td>
</tr>
</tbody>
</table>

The reliability of the prior core accounts depends on the source of the respective entries. While it is possible to obtain data on A6 and A16 level for the majority of countries, at b19 level the available information is scarcer, in some cases only available from national input-output tables for some years. Therefore, different values for \( S^n \) were chosen to reflect the data availability. To ensure consistency across the different aggregation levels, the following conditions were imposed:

\[
\sum_{b19} \left[ G^{19,16}_{b19,A16} \cdot S^{\text{MS},b19}_{i,A16} \right] = S^{\text{MS},i,A16}
\]

(12)

\[
\sum_{A16} \left[ G^{16,6}_{A16,A6} \cdot S^{\text{MS},i,A16} \right] = S^{\text{MS},i,A6}
\]

Here, \( G \) denotes the aggregator matrices between levels of branch aggregation. The condition above only imposes consistency of the i core accounts within the respective aggregates, but do not ensure that certain sub-items add up. For instance, "gross output – p1" should be equal to "gross value added at basic prices – b1g" plus "intermediate demand at purchaser’s prices – p2pp" or "employment – emp" should be the sum of "employees – sal" and "self-employment – self". For those identities, a set of additional constraints is imposed on sub-sets of i and the corresponding items that should add up (here: ii), in general expressed as:

\[
\sum_{ii} \left[ G^{i,ii}_{i,ii} \cdot S^{\text{MS},ii,j}_{i,j} \right] = S^{\text{MS},i,j}
\]

(13)

Solving equation 11 with respect to equations 12 and 13 for the core accounts at national level results in a set of balanced core accounts \( S^n \), from which average national wages "wage" and coefficients "incf" for a number of inputs ("inp") are obtained:

\[
\text{wage}^{\text{MS}}_{j} = S^{\text{MS},i,d1,1}_{i,j} / S^{\text{MS},i,\text{emp},i}_{i,j}
\]

\[
\text{incf}^{\text{MS}}_{\text{inp},j} = S^{\text{MS},i,\text{inp},j}_{i,j} / S^{\text{MS},i,p1,1}_{i,j}
\]

Wages and input coefficient are important for the subsequent step in which the regional core accounts are balanced. For the regional balancing procedure, equations 12 and 13
are also imposed like on national level, but equation 14 also enters as a constraint (this requires to replace the index MS by r for the regions of the individual Member States). Due to the fact that information on Member State level could in many cases only be obtained at the A6 level of aggregation, it is necessary to derive the regional core accounts based on national shares (as outlined in section 3). Whenever regional entries have to be derived in such a way, the assumed variance \( S^\sigma \) is set at a much higher level than in those instances where information could be obtained e.g. from NSOs. Applying the same balancing procedure as on national level results in many cases in values that fulfilled adding-up conditions but deviated substantially from national wages or input coefficients. While this is tolerable for reported data, is unsatisfying for purely derived data. Therefore, the objective function 11 is modified for regional core accounts by adding two terms that penalize overly large deviations from national coefficients:

\[
\begin{align*}
\min_{S_{r,i,j}, \text{wage}_{r,i,j}, \text{inff}_{r,i,j}} Z &= \\
\sum_{r,i,j} \left( \frac{S_{r,i,j} - S^0_{r,i,j}}{S^\sigma_{r,i,j}} \right)^2 + \\
\sum_{r,i,j} \left( \frac{\text{wage}_{r,i,j} - \text{wage}^\text{MS}_{r,i,j}}{\text{wage}^\sigma_{r,i,j}} \right)^2 + \\
\sum_{r,i,j} \left( \frac{\text{inff}_{r,i,j} - \text{inff}^\text{MS}_{r,i,j}}{\text{inff}^\sigma_{r,i,j}} \right)^2
\end{align*}
\]

The variances for wage and inff \((\text{wage}^\sigma, \text{inff}^\sigma)\) are set such that they are larger than those for reported data but smaller than those for purely derived data:

\[
S^\sigma_{r,i,j,\text{reported}} < \text{wage}^\sigma_{r,i,j}, \text{inff}^\sigma_{r,i,j} < S^\sigma_{r,i,j,\text{derived}}
\]

This combination of variances ensures that reported bits of information are maintained and implausible values for derived accounts are avoided. Consistency of the regional core accounts is also imposed:

\[
\sum_r \left[ G_{r,MS} \cdot S_{r,i,j} \right] = S_{MS,i,j}
\]

Where \( G \) here denotes the mapping of NUTS2 regions with the respective Member States.

The results of the two-step balancing procedure for the core accounts are satisfying. As shown in Figure 15 for the case of p2pp, the ratio between balanced and prior values centres around one, thus indicating unbiased and efficient estimates of the balanced core account data. When comparing the deviations between national and regional balanced wages, the results show a comparable pattern (Figure 16). Based on the evaluations of the core account balancing procedure, the estimates are sufficiently accurate to form the basis for the subsequent steps.
6.3 SAM Balancing Approach

6.3.1 National SAM

After the consolidation of the core accounts, the transaction matrices national input-output tables from Eurostat or other sources are adjusted with respect to the obtained values for intermediate demand at purchasers’ prices (p2pp). Other entries are replaced by the balanced core accounts. This ensures that the SAMs for different years could be used to generate a full set of national for one base year (e.g. 2005). More specifically:
The balancing is carried out similarly to equation 11, with adding up conditions as in equation 13. When comparing the results for UK in 2005 (Figure 17), for which the latest reported IOT dated from 2000, it appears that the deviations from the prior are centred on one, although with a slight bias to the right. This indicates that the values in the balanced SAM are increased on average due to imposing adding-up conditions and compliance with core-account entries.

Figure 17 Deviation between prior and balanced national SAM for UK 2005 (S/S0)

6.3.2 Total Transaction Matrices on NUTS2 Level

Equipped with a full set of national SAM and balanced regional core accounts, the next step in the sequence is the compilation of prior regional total transaction matrices without distinction origins. This step ensures that total intermediate demand coefficients for the industries as well as for final consumption do not deviate substantially from the national averages, unless regional datasets indicate that this may be the case. It has to be noted that the usage of national coefficients still ensures a higher plausibility of the regional demand structure as opposed to the only usage of adding-up conditions. As in the case of balancing regional core accounts, the variances are set in such a way that deviations from observed data are stronger penalized than deviations from national coefficients – the highest variance is again associated with prior entries which were purely derived from national statistics. The objective function for this compilation step includes therefore two terms: one that penalized deviations from SAM values, the other from national input coefficients incf (note the new indexes). The indexes i and j refer now to the full SAM accounts, not anymore to the core accounts only:
\[
\min Z = \sum_{r,i,j} \left( \frac{S_{r,i,j} - S_{r,i,j}^0}{S_{r,i,j}^\sigma} \right)^2 + \sum_{r,i,j} \left( \frac{incf_{r,i,j} - incf_{r,i,j}^{MS}}{incf_{r,i,j}^\sigma} \right)^2
\]

Objective function 18 is then minimized subject to the condition that the regional total transactions add up to the national level:

\[
\sum_r \left[ G_{r,MS} \cdot S_{r,i,j} \right] = S_{MS,i,j}
\]

One example for available regional SAM is Andalucía (ES61), for which the ratio between prior data (the original tables, adjusted for core account consistency) and the balanced values is shown in Figure 18. It appears that the distribution centres narrowly around one, so that the balanced values are sufficiently close to the observations.

**Figure 18** Deviation between prior and balanced regional total transaction matrix for Andalucía ES61 (S/S^0)
6.3.3 Detailed Transaction Matrices on NUTS2 Level

Having derived a full set of regional total transaction matrices, the totals should be distributed by three potential origins: regional, domestic, and imported. Of particular importance are the inter-regional trade flows, which do not appear in the national table, such that the accounting identities have to be adjusted accordingly. Namely “total uses” (tu) of regional origins include now “domestic exports” (p6s1), which have to sum up to “total uses” of domestic origin across all regions. This means in turn, that regional and domestic total uses for each region do not add up to the respective sub-totals from the previous step because of the additional entries in the “domestic exports” column. All other identities remain unchanged.

Figure 19 illustrates the distribution of relative deviations between balanced and base values for all matrix entries in the case of Andalucía, for which a regional SAM is available from NSO. The deviations are centred on one, with additional peaks in the ranges of plus and minus 40% from the prior. Here, adjustments of comparatively small entries are necessary to meet the adding-up conditions imposed at this stage. The high share of zero-values (16% of all cases) refers to cases in which entries in the regional table are deleted because they are not consistent with entries in the national SAM.
7. Implementation

The flow of operations to generate the final balanced SAMs at NUTS2 can be separated into four distinct steps. First, data are imported from different formats (file types, e.g. csv, tsv, xls ...) and converted into gdx format, while maintaining their original structure of economic branch classifications, country-classifications, and so forth. In the case of Eurostat and IPTS data, this is done within the boxes ("procedures") starting with the prefix "import_" in 0. The files "import_estat_naios.tsv.gms" and "import_estat_naios.xls.gms" for instance load national input-output tables from either bulk-download (tsv files) or separately downloaded spreadsheets (MS xls). The spreadsheet-files were included as they are more frequently updated than the database for bulk-downloads. Other datasets from Eurostat are also loaded and converted into gdx. The important property of these "import_" files is that the datasets are kept in their original format (usually ESA95). In a next step, the gdx-files are re-arranged to follow the target classifications used for the SAMNUTS2 database (mapping in files "mapdat_estat_XXX.gms"). This strict separation of importing and data transformation permits flexible changes of target classifications without the need to re-run rather time consuming import procedures. In case of national statistics, this procedure is changed as the multitude of different classifications called for a merged procedure. In such cases, importing and mapping is done in one step ("impmap_XXX_XXX.gms""). All loaded and converted datasets are then combined into a single parameter "p_data_raw" that has to be used in the further steps (combdat_main.gms). This step serves also as a structural break between data-handling and actual prior-construction. Steps before combdat_main.gms need only be repeated if new datasets are added; subsequent steps do not refer to files generated before combdat_main.gms.

Based on this combined raw database, the core accounts are constructed in a next step (build_core.gms) and balanced on national level (balance_core_national.gms). The results from the national balancing procedure are then fed into the balancing of regional core accounts. Based on those and the raw data, national SAMs are constructed and balanced (build_nsam.gms) as described in section 5.3.1. Regional core accounts, regional statistics, and balanced national tables are then used to create a complete, yet unbalanced, set of SAMs at regional level (build_rsam.gms). The regional SAMs are then first balanced for the totals of the transaction matrices (balance_tsect2.gms), then by their origin (balance_idrsect2.gms).

The all steps are implemented in the software GAMS, the balancing procedures are set-up as NLP and solved with the numerical algorithm CONOPT3. All balancing steps are coded in loops over Member States and years, which permits the parallel processing of the respective sub-problems. This feature becomes particularly important when considering the computer time needed for solving the respective problems, which depends mainly on the number of NUTS2 regions in a Member State (Figure 23). Particularly balancing the core accounts on regional scale turned out to be the most time-consuming step with single countries requiring 3000 and more CPU seconds on a normal desktop PC with two cores. The following balancing the regional total transaction matrices took considerably less time, but the detailed transaction matrices required again up to 2500 CPU seconds. When these steps have to be performed
sequentially, the user is likely to wait extensively for results, while the largest single problem is the limiting factor in case of parallel processing.

Figure 20  Flow of Raw-Data Import in the SAMNUTS2 Procedure

Figure 21  Flow of Compilation Steps in the SAMNUTS2 Procedure
Figure 22  Directories of the SAMNUTS2 Main Folder

Figure 23  Number of NUTS2 regions and resource use of compilation steps (in CPU seconds on normal desktop PC)
8. Summary and Conclusion

The purpose of the SAMNUTS2 project is to create a database that permits the implementation of regionalized general equilibrium models for rural development policy analysis. To meet this purpose, it is necessary but not sufficient to estimate balanced regional/national SAMs. Policy parameters (e.g. tax rates) and technology coefficients have to permit the calibration of computational models, which can be impaired by overly large (i.e. in comparison with related indicators) or small (i.e. close to zero) entries. Some constellations of estimated SAM coefficients may render a calibration process infeasible or at least numerically demanding. These problems originate from the interactions of SAM entries and the definition of equations in the models which have to be calibrated or fitted to the SAMs. Even if it is possible to calibrate or fit the model in question to the SAMs, the simulation behaviour may be implausible (e.g. due to overly high savings rates of some regional households). As such problems are only revealed in the interaction of SAM and model, it is not always possible to eliminate a-priori SAM entries that may cause problems during calibration or simulation stages. Currently (October 2012), a regional CGE (Britz 2012) has could be calibrated to the SAMNUTS2 database. This calibration required frequent interaction between SAM and model developers and also substantial changes in the SAMNUTS2 estimation process, e.g. regarding the constraining of tax parameters and input coefficients to plausible ranges, or the final demand for products of regional origin. We considered at some point to include some – if not all – of the CGE accounting identities and calibration constraints in the estimation procedure, but this would have added a new set of non-linear constraints at the expense of computation time and tractability of estimation results. In the current SAM balancing procedure, it is comparably straightforward to trace the source for implausible entries back to the raw data, if necessary. This process would become more difficult once an additional layer of constraints is added. Also, as some accounting or calibration constraints may be unique for certain models, it seems at this stage that the current procedure (direct, personal, interaction between database and model developers) is preferable to a more automatized procedure. Once a set of models using the SAMNUTS2 database has been identified, it may be possible to include generic model constraints in the estimation procedure.

The above discussion highlights some important aspects of the current state of the SAMNUTS2 project and the strategy for further development: First, the SAMNUTS2 estimation procedure generates a database that can be used to put regional CGEs for all EU Member States at NUTS2 into operation. To our knowledge, it is the first database of this kind. Second, the SAMNUTS2 database uses all information available in an efficient manner and preserves the recorded data structure as long as accounting and modelling constraints are not violated. This emphasizes the nature of SAMNUTS2 as a model and not a statistical database. Third, the actual entries in the SAMNUTS2 database are not static. Depending on the targeted models, the database may change due to new accounting or calibration constraints. Should the number of models increase, this may cause the parallel existence of different versions of the SAMNUTS2 database. To which extent such a development would be desirable is questionable and should be taken into account in further stages. Finally, the quality of the SAMNUTS2 database can only be
improved through actual usage as many implausible entries can only be detected through calibration of and simulation with a variety of computational models
References


Britz W. (2012): RegCgeEU+ in GAMS, documentation including the Graphical User Interface. CAPPRI-RD Deliverable 3.2.4 (http://www.ilr.uni-bonn.de/agpo/rsrch/caprird/docs/d3.2.4.pdf)


Annex 1: Example for the Computation of TIG

The following example is meant to illustrate the calculation of the informational gain indicator (TIG) as introduced in the beginning of section 3. It does not refer to any particular classification of branches in the regional account datasets. The target classification for the SAMNUTS2 database features 19 economic branches (b19), whereas the regional datasets available from Eurostat provide only information for 6 branch aggregates. The mapping between these branches may be presented in the form of an aggregator matrix $G_{19,6}^{b19,A6}$, that indicates which items in the b19-framework have to be added up to the corresponding entries in the A6 classification:

Adding-up the row-dimension of this aggregator matrix yields the number of elements of b19 in A6:

$$\sum_{b19} G_{b19, A6}^{19,6} =$$

As the Eurostat datasets are available for all NUTS2 regions in this A6 classification, it would be only necessary to collect the number of elements of b19 in A6 less one for each A6 category. For example in the case of the category A2B, it would be sufficient to collect 2 data-points for the corresponding elements in b19 (AA01, AA02, or B000) as it is possible to calculate the remaining value residually. Therefore, it would be sufficient to collect only 13 data-points (19-6), provided that they can be mapped accordingly into the b19 classification. The needed data-points are therefore:

$$\sum_{b19} G_{b19, A6}^{19,6} - 1 =$$

Consider now the case of a obtained regional dataset that provides information for branches in a specific regional format (AR). The correspondence between the regional and SAMNUTS2 classification may also be represented in the form of an aggregator matrix $G_{19,AR}$.

\[
G_{b19,AR}^{19} \equiv
\]

This regional classification may provide aggregate figures in the case of branch A, which corresponds with the branches AA01 and AA02 in the b19 format. In other cases, AR is more detailed than required as the data on CA and CB would anyway be added to the category C in the b19 format. Counting the number of usable entries in the AR classification with respect to b19 is expressed as a function $\delta$ of the row-sum of $G_{19,AR}$:

\[
\delta \left( \sum_{AR} G_{19,AR}^{19} \right) = \begin{cases} 
1 & \text{if } \sum_{AR} G_{19,AR}^{19} > 0 \\
0 & \text{if } \sum_{AR} G_{19,AR}^{19} = 0
\end{cases}
\]

To compare the gain of usable information obtained from this regional dataset with the information provided by Eurostat, one has to re-arrange the dataset into the Eurostat A6 classification. This is here done by first multiplying the $d$-values with the aggregator matrix $G_{19,6}$ that combines SAMNUTS2 target branches with the A6 classification.
Adding up this intermediate matrix for each column shows the number of usable entries, corresponding to the Eurostat A6 classification. In the given example, the regional dataset provides only 2 data-points for A2B instead of the 3 entries that would result from mapping a complete b19 dataset into the A6 categories:

\[
\sum_{b19} G_{b19,A6}^{19,6} \cdot \delta \left( \sum_{AR} G_{i,b19,AR}^{19,AR} \right) = \begin{bmatrix}
\text{Eurostat Classification (A6)}
\begin{array}{cccccc}
A2B & C2E & F00 & G21 & J2K & L2P \\
AA01 & 1 & 1 & 1 & 1 & 1 \\
AA02 & 1 & 1 & 1 & 1 & 1 \\
0000 & 1 & 1 & 1 & 1 & 1 \\
C000 & 1 & 1 & 1 & 1 & 1 \\
DA15 & 1 & 1 & 1 & 1 & 1 \\
DF23 & 1 & 1 & 1 & 1 & 1 \\
D020 & 1 & 1 & 1 & 1 & 1 \\
E080 & 1 & 1 & 1 & 1 & 1 \\
F000 & 1 & 1 & 1 & 1 & 1 \\
G000 & 1 & 1 & 1 & 1 & 1 \\
H000 & 1 & 1 & 1 & 1 & 1 \\
I000 & 1 & 1 & 1 & 1 & 1 \\
J000 & 1 & 1 & 1 & 1 & 1 \\
K000 & 1 & 1 & 1 & 1 & 1 \\
L000 & 1 & 1 & 1 & 1 & 1 \\
M000 & 1 & 1 & 1 & 1 & 1 \\
N000 & 1 & 1 & 1 & 1 & 1 \\
P000 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\end{bmatrix}
\]

As Eurostat provides already one data-point for each A6 category in the form of a subtotal that would allow calculating one entry residually, the gain of information is again the number of entries less one:

\[
\sum_{b19} G_{b19,A6}^{19,6} \cdot \delta \left( \sum_{AR} G_{i,b19,AR}^{19,AR} \right) - 1 = \begin{bmatrix}
\text{Eurostat Classification (A6)}
\begin{array}{cccccc}
A2B & C2E & F00 & G21 & J2K & L2P \\
2 & 3 & 1 & 2 & 1 & 0 \\
\end{array}
\end{bmatrix}
\]

So, for instance in the case of category A2B, only 1 data-point could be effectively gained if compared to the Eurostat regional datasets. To derive an indicator to measure the total informational gain across all branches, the previous operations with the aggregator matrices \(G^{19,6}\) and \(G^{19,AR}\) are combined to derive the total informational gain indicator (TIG) used throughout this inventory:

\[
TIG_{i}^{MS,R} = \sum_{A6} \left( \sum_{b19} G_{b19,A6}^{19,6} \cdot \delta \left( \sum_{AR} G_{i,b19,AR}^{19,AR} \right) \right) - 1 \right) / \sum_{A6} \left( \sum_{b19} G_{b19,A6}^{19,6} \right) - 1
\]
Abstract

The quantitative assessment of agricultural and rural development policies requires consolidated, multi-sector databases, which take the economic heterogeneity between regions into account. Regionalized Social Accounting Matrices represent a suitable and convenient data format for such a database, which, to our knowledge, does not yet exist for the 27 current EU Member States. To address this informational gap, the European Commission's IPTS launched the SAMNUTS2 project to obtain a complete and consistent set of regional Social Accounting Matrices. This report summarizes the achievements of the SAMNUTS2 project to construct such a database.
As the Commission’s in-house science service, the Joint Research Centre’s mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.