Thissen, Mark; Lankhuizen, Maureen; van Oort, Frank G.; Los, Bart; Diodato, Dario

Working Paper

EUREGIO: The construction of a global IO DATABASE with regional detail for Europe for 2000-2010

Tinbergen Institute Discussion Paper, No. TI 2018-084/VI

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Thissen, Mark; Lankhuizen, Maureen; van Oort, Frank G.; Los, Bart; Diodato, Dario (2018) : EUREGIO: The construction of a global IO DATABASE with regional detail for Europe for 2000-2010, Tinbergen Institute Discussion Paper, No. TI 2018-084/VI, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/185603

TI 2018-084/VI
Tinbergen Institute Discussion Paper

EUREGIO: The construction of a global IO DATABASE with regional detail for Europe for 2000-2010

Mark Thissen[1]
Maureen Lankhuizen[2]
Frank (F.G.) van Oort[3]
Bart Los[4]
Dario Diodato[5]

[1] Netherlands Environmental Assessment Agency (PBL)
[2] Vrije Universiteit Amsterdam
[3] Erasmus University Rotterdam
[4] University of Groningen
[5] Harvard Kennedey School

EUREGIO: The construction of a global IO DATABASE with regional detail for Europe for 2000-2010

Mark Thissen (PBL), Maureen Lankhuizen (PBL, VU), Frank van Oort (EUR, UU), Bart Los (RUG) and Dario Diodato (HKS)

October 2018

CONTENTS

1 INTRODUCTION: RATIONALE AND METHODOLOGY FOR CONSISTENT INTERREGIONAL IO TABLES IN EUROPE

Regions matter for economic performance. In practically all countries, regions vary in terms of their economic specialization patterns and hence in their contribution to production and productivity growth (OECD, 2009). Differences in population density are related to differences in the importance of agricultural activities, while differences in the importance of agglomeration externalities also cause variation in specialization patterns of regions and cities (Combes et al., 2010). Differences in specialization patterns across regions have implications for the extent to which they are affected by shocks, of various sorts. Autor et al. (2013) show that workers in regions in the United States with a specialization in specific manufacturing industries were more vulnerable for the emergence of China as a giant exporter of manufactured products than workers in other region-sector combinations. Similar differences were also revealed by Acemoglu and Restrepo (2017) with respect to the vulnerability of workers to the consequences of robotization. Regional-economic resilience analyses of the recent global economic crisis shows a large degree of regional heterogeneity according to specialization patterns and sorting effects (Groot et al., 2011; Fingleton et al., 2012). Interestingly, analyses of voting patterns in the Brexit referendum and the US Presidential elections also show that regional outcomes correlate strongly with variables related to regional sectoral specialization patterns (Becker et al., 2017; Autor et al., 2017).

Differences in the regional compositions of economic activities are of course not the only determinant of within-country differences in economic performance. Among many other relevant determinants, variations in supplier-user linkages to industries in other regions also play a role. A region might be seriously affected by an economic downturn in another region if it sells much of its output to that region, while regions less dependent on that region might be hurt to a much lesser extent when in crisis (Thissen et al., 2016).

These arguments reinforce a longstanding demand for consistent interregional input-output (IO) tables, which contain information on both sectoral specialization, and linkages within and across regions (e.g., Isard, 1953; Hewings and Jensen, 1987). Such tables could be used for regional economic analyses varying from impact studies (of global or localized economic shocks) to general equilibrium modelling to evaluate alternative policy options. The world has changed since the earliest pleas for such data. The fact that the shares of both final products and intermediate inputs sold across national borders increased with the emergence of Baldwin's (2006) "second wave of

globalization", implies that the effects of supplier-user linkages are no longer restricted to interregional transactions within countries. Incorporating such effects in regional analyses requires global IO-tables in which (at least some neighbouring) countries are spatially disaggregated.

In recent years, researchers have made great strides in constructing global IO tables, by linking data on national production structures (national accounts, supply and use tables and/or IO tables) to data on bilateral trade in goods and services.[1] The construction of such tables with regional detail is still in its infancy, though. Dietzenbacher et al. (2012) included Brazilian regional detail in the World Input-Output Database (WIOD). Cherubini and Los (2012) similarly analyzed four Italian NUTS1 regions. Wang et al. (2017) incorporated regional data for China in Eora's global IO tables, and Meng and Yamano (2017) presented analysis based on spatial disaggregations of China and Japan (in turns, not simultaneously) within the OECD's Trade in Value Added tables. Despite the good quality of data generally available in (countries of) the European Union, there has never been a longitudinal and consistent dataset on trade across EU-regions. Given the hypothesized importance of European national and regional integration for overcoming disadvantages of sectoral and spatial fragmentation, as well as the need for consistent evaluation of a burgeoning number of policy initiatives on cohesion, competitiveness, resilience, growth and smart specialization (McCann, 2015; Bachtler et al., 2013; Foray, 2015; Piattoni & Polverari, 2016), this absence is remarkable and is felt to full impact.

This paper introduces the EUREGIO database: the first time-series (annual, 2000-2010) of global IO tables with regional detail for the entire large trading bloc of the European Union. The construction of this database, which allows for regional analysis at the level of so-called NUTS2 regions, is presented in detail for its methodology and applications.[2] The tables merge data from WIOD (the 2013 release) with, regional economic accounts, and interregional trade estimates developed by PBL Netherlands Environmental Assessment Agency (PBL/RT, see section 3.2), complemented with survey-based regional input-output data for a limited number of countries. All used data are survey data and *only* non-behavioral assumptions have been made to estimate the EUREGIO dataset. These two general rules of data construction allow empirical analyses focused on impacts of changes in behavior (of economies, firms, policies) without endogenously having this behavior embedded

[1] See, e.g. the contributions to a special issue of *Economic Systems Research* (Tukker and Dietzenbacher, 2013), with contributions by the consortia that constructed GTAP-based global IO tables (Andrew and Peters, 2013), Eora (Lenzen et al., 2013), EXIOPOL (Tukker et al., 2013), the World Input-Output Database (Dietzenbacher et al., 2013) and IO tables covering South East Asia and its main trading partners (Meng et al., 2013). Another popular source of global IO tables is OECD's Trade in Value Added database.

[2] Details on the NUTS2 classification are presented in Appendix A.

already by construction. The tables are publicly available to the research community, from the Dutch government open data website.

The construction of the time series of multiregional NUTS2 input-output tables is based on a top-down approach where national accounts in the format of national supply and use tables have been taken as given. A supply and use framework is used rather than an input-output (IO) framework. An input-output framework uses the assumption that every sector produces only one good. There are two types of input-output matrices: product-by-product matrices and sector-by-sector matrices. Product-by-product input-output matrices are generally constructed around the product classification, and sectors are therefore adjusted or mixed in such a way that only one sector makes only one product. Sector-by-sector input-output matrices are generally constructed around the sector classification, and products are therefore adjusted or mixed in such a way that only one sector makes only one product. This implies that, depending on the type of IO table, either sectors are not comparable across countries and not comparable with regional sector statistics, or products are not comparable across countries and not comparable with trade statistics. In this project, both a regional trade database, and production and consumption data of different actors in different regions, were used. The focus is thus intentionally on the regionalization of both trade and the regional use and supply of products by different economic actors. The regionalization of a complete supply and use framework is then the only option available.

The paper is organized around the successive steps of the data construction. These are outlined below and will be explained in more detail in the remaining sections in the paper.

1. Adjustment of WIOD (section 2). The WIOD international supply and use tables were taken as the starting point of the analysis. The WIOD database (Timmer et al., 2012; Dietzenbacher et al., 2013) makes a detailed distinction between final and intermediate goods trade. The supply and use tables provide detailed information on bilateral trade for 40 countries and the rest of the world. The data include 59 product categories, among which services, according to the European Statistical Classification of Products by Activity (CPA) 2002. The data are consistent with countries' national accounts.[3] The WIOD international supply and use tables are first adjusted so as to (1) account for the distribution of the re-exports over (most likely) origin and destination countries, and (2) to ensure consistency in bilateral trade flows (i.e., trade matching: exports from *i* to *j* equal imports of *j* from *i*), and (3) that exports and imports of each country add up to their national accounts totals as presented in the WIOD

[3] See Appendix A for a list of countries, product categories and regions identified in the data base.

database. Both adjustments have to be done before the regionalization because otherwise inconsistencies would have to be regionalized as well. The regionalization of inconsistencies is theoretically not possible since they do not exists in reality and therefore cannot be based on actual information.

2. Regional information (section 3). Subsequently, information on sector production, investment and income development from the Eurostat regional accounts was added. After these have been made consistent with the above mentioned national accounts, the data were used to regionalize the national tables. As the outcome of this regionalization procedure, regional supply and use tables for each of the 256 European NUTS2 regions, for 14 sectors and 59 product groups for the years 2000 to 2010, are obtained. Where available, regional survey based information on supply and use of different sectors was added. In particular, regional supply and/or use tables are available for Scotland and Wales, as well as Italy (five NUTS1 regions), Finland (21 NUTS3 regions) and Spain (15 NUTS2 regions). These tables were added as additional priors to the estimation. Regional trade is taken from the PBL Netherlands Environmental Assessment Agency regional trade data for the year 2000 as a prior to the estimations for 2000-2010.[4]
3. Construction of tables (sections 4 and 5). Taking the regionalized supply and use tables, the PBL regional trade data and the survey based regional supply and use tables as a prior, the EUREGIO supply and Use tables are estimated for the years 2000-2010. The estimation approach is based on a constraint non-linear minimization approach that guarantees consistency of the regional tables with the national tables (the WIOD database).[5] This consistency implies that adding up the regionalized supply and use tables results in the corrected national WIOD supply and use tables. The interregional supply and use tables that have trade, matched bilateral trade flows and no re-exports.
4. We conclude (in section 6) on the usefulness of this type of regional IO tables with an overview of current applications of the EUREGIO database.

[4] The only fully consistent database on trade in goods and services at the NUTS2 regional level was constructed by PBL Netherlands Environmental Assessment Agency (Thissen et al., 2013a, 2013b and 2013c). This database, based on a parameter-free estimation of trade data (Simini et al., 2012), was the basis for the tailor-made bilateral trade data set used here.

[5] A quadratic minimization function is to be preferred over a logarithmic function (often used in entropy minimization, e.g., Thissen and Lofgren, 1998) because of its mathematical properties that reduce computation time significantly.

2 TRADE MATCHED AND RE-EXPORT CORRECTED INTERNATIONAL SUPPLY AND USE TABLES

Before inter-country trade could be regionalized in the EU/REG/IO database, some adjustments in the WIOD trade data had to be made.

First, in the original WIOD database, exports were not assigned to countries of destination. As a precondition, the total exports in the tables cannot be less than the imports from all other countries after the correction for re-exports. Hence, bilateral trade flows in WIOD had first to be made consistent, i.e. fully trade matched. This is part of a wider adjustment of the WIOD data to correct for the actual origin of re-exported trade discussed below. Before any differences between the values of exports and imports can be evaluated, they should both be valued in the same prices. The WIOD tables follow Eurostat in having both exports and imports in fob (free on board) prices. However, not all countries present their exports at the product level in fob, since the Eurostat manual leaves the choice between exports in fob at the product level and total exports in fob open to the bureaus of statistics. Moreover, countries have changed the way they report the exports over time. In the year 2010 there are 18 out of 40 countries that use a correction term to have only the total value of exports in fob prices. These countries are Austria, Bulgaria, Cyprus, Germany, Spain, Estonia, Finland, France, Greece, Hungary, Italy, Lithuania, The Netherlands, Poland, Portugal, Romania, Slovenia and the United States. This correction factor had to be applied to the different products to obtain fob prices at the product level. Since there is no information available about these margins, the domestic trade and transport margins are used as a proxy. The correction factor has therefore been applied over the products proportional to the domestic trade and transport margins.

Second, trade flows between countries in WIOD were not corrected for re-exports. Re-exports are goods imported by a resident, who assumes (short-term) ownership of the goods.[6] The goods are subsequently exported without having received any significant industrial processing.[7] Re-exported goods may actually never have been in the intermediate country, as trade is only determined by ownership changes. Re-exports affect the measured patterns of bilateral trade. Re-exports are recorded as exports of the intermediate country in the system of national accounts. This implies that trade flows are registered to a different country than the true origin or final destination. Our

[6] Re-exports are important foremost in goods trade and much less in services trade.

[7] If there is no transfer of ownership at any stage, the goods are considered to be in transit. Transit goods are not part of the system of national accounts and are therefore not part of the (nationally) reported exports and imports.

correction involves estimating complete origin-destination matrices of re-exports and then restoring bilateral trade to its 'proper' origin and destination.[8]

The WIOD supply and use tables include estimates of the size of total re-exports per intermediate country. The re-exports in the WIOD database are minimum estimates derived from the accounting principle that exports cannot be larger than production. Re-exports in WIOD are therefore determined as the exports minus the production if this results in a positive number. In mathematical form this can be written as:

$$RE_{q,p} = Max\left[0,\left(Ex_{q,p} - Y_{q,p}\right)\right], \quad [1]$$

where $RE_{q,p}$ denotes re-exports *RE* by country *q* and product *p*, excluding trade margins, $Ex_{q,p}$ denotes exports of *p* from country *q* excluding trade margins, and $Y_{q,p}$ is the production of product *p* in country *q*.

Some provisions have been made in WIOD to account for re-exports. The WIOD trade database leaves out re-exports from the imports by intermediate country *q* of product *p*. However, it does not address re-exports from the intermediate country to the final destination. Our correction then entails the following two steps. We first subtract from the imports by destination *j* of product *p* in the original tables the re-exports from intermediate country *q*. Subsequently, we add to trade in *p* between origin *i* and *j* re-exports via any intermediate *q*. This is explained by the following equation:

$$TRE_{i,j,p} = T_{i,j,p} - RED_{q,j,p} + \sum_{q} REOD_{i,q,j,p} + \hat{t}^{adj}_{i,j,p} \quad [2]$$

Here, $TRE_{i,j,p}$ denotes bilateral trade between *i* and *j* in product *p* corrected for re-exports; $T_{i,j,p}$ designates trade in *p* between origin *i* and *j* in the original tables; $RED_{q,j,p}$ denotes re-exports of *p* from intermediate country *q* to final destination *j* ; and $REOD_{i,q,j,p}$ is the origin-destination matrix of re-exports in *p* via *q*. The adjustment term $\hat{t}^{adj}_{i,j,p}$ ensures consistency of bilateral trade, i.e. that total imports from a certain origin cannot exceed the exports of that origin.

[8] See Lankhuizen and Thissen (2018) for a detailed description of the methodology and implications of the corrections.

Box 1: Optimisation and constraints

Objective function – minimization of errors

$$Z = \sum_{i,q,p} \left(\hat{e}_{i,q,p}\right)^2 + \sum_{q,j,p} \left(\hat{e}'_{q,j,p}\right)^2 + \sum_{i,q,j,p} \left(\hat{e}_{i,q,j,p}\right)^2 + \sum_{i,j,p} \left(\hat{t}^{adj}_{i,j,p}\right)^2 \quad [3]$$

Under the following constraints:

- Total re-exports from origin *i*, $\sum_q REO_{i,q,p}$, cannot exceed its total exports $Ex_{i,p}$;
- Total re-exports to destination *j*, $\sum_q RED_{q,j,p}$, cannot exceed its total imports $IM_{j,p}$;
- Trade consistency: exports from *i*, $Ex_{i,p}$, cannot exceed total imports from *i*, $\sum_j TRE_{i,j,p}$;
- Trade consistency: imports into *j*, $Im_{j,p}$, cannot exceed total exports into *j*, $\sum_i TRE_{i,j,p}$.

Regarding the errors the following definitions apply:

- Estimated origin of re-exports: $REO_{i,q,p} = PI_{i,q,p} RE_{q,p} + \hat{e}_{i,q,p}$;
- Estimated destination of re-exports: $RED_{q,j,p} = PE_{q,j,p} RE_{q,p} + \hat{e}'_{q,j,p}$;
- Origin-destination matrix of re-exports: $REOD_{i,q,j,p} = \frac{REO_{i,q,p}}{RE_{q,p}} RED_{q,j,p} + \hat{e}_{i,q,j,p}$

where,

$PI_{i,j,p} = \frac{T_{i,j,p}}{\sum_{i'} T_{i',j,p}}$, is the probability of the origin *i* of imports of product *p, and*

$PE_{i,j,p} = \frac{T_{i,j,p}}{\sum_{j'} T_{i,j',p}}$, is the probability of the destination *j* of exports of product *p.*

The probabilities of origin and destination of re-exports reflect our assumption that re-exports have a similar country pattern as regular trade. That is, important origins (destinations) of a country's imports (exports) are also important origins (destinations) of its re-exports. See Lankhuizen and Thissen (2018) for a further discussion of this assumption.

$RED_{q,j,p}$, $REOD_{i,q,j,p}$ and $\hat{t}^{adj}_{i,j,p}$ are estimated through a constrained non-linear optimisation procedure. The procedure is summarized in Box 1. The minimization results in the complete re-

export matrices $REOD_{i,q,j,p}$ and new, fully matched trade flows between countries for 59 product categories.

3 REGIONAL INFORMATION TO REGIONALIZE SUPPLY AND USE TABLES[9]

Central in the procedure of constructing EUREGIO is regionalization of the WIOD supply and use framework. The regional supply and use tables are organised according to the standard structure of use and supply tables (see Tables 1 and 2). This means that total use in the region equals total regional supply. Also, column totals in the regional supply and use tables are equal because total output of every regional industry equals this industry's total input and value-added.

Table 1: Use table

	input industry 1	input industry 2	...	input industry s	household demand	non-profit demand	government demand	fixed capital formation	changes in inventories	changes in valuables	exports	total
product 1												
product 2												
..												
product p												
total												

Table 2: Supply table

	input industry 1	input industry 2	...	input industry s	trade and transport margins	+ taxes -subsidies on production	imports	total
product 1								
product 2								
..								
product p								
total								

The regionalization of supply and use was achieved as a first estimate using the Commodity Balance (CB) method, first suggested by Isard (1953). By multiplying the national values of value-added, investment, government demand and consumer demand in WIOD with the respective shares of regions in the national totals, consistent regional values were obtained. This part of the analysis did not involve the estimation of regional trade and the assumption of the same production technologies and consumption patterns was later relaxed using additional regional supply and use information where available. Regional data on value-added, regional investment, government demand and

[9] Section 3 and 4 are in part based and elaborates on Appendix A of Thissen et al (2013), and are used here with permission.

consumer demand (both with regional income as a proxy) are from Eurostat (2014). The Eurostat data provide information on regional totals, without a distinction between different products. The structure of the national supply and use tables is assumed to give a good approximation for the regional tables. More formally, consumers are assumed to have homogenous preferences throughout the country concerned, homogenous government spending is assumed over the regions, and industries are assumed to use the same production technology, irrespective of their location within the country concerned.

3.1 REGIONAL PRODUCTION AND CONSUMPTION

With respect to the supply tables, regional output has been broken down by industries, trade and transport margins and net taxes. First, the data on production were determined per industry as well as the intermediate demand. No information is available on output at the European NUTS2 level. However, data on value-added (*VA*) for 14 economic sectors is available from Eurostat. The number of sectors in WIOD have therefore been aggregated accordingly. Table 3 presents the classification for these 14 economic sectors.

Table 3: Industry classification in 14 sectors

S1	Agriculture
S2	Mining, quarrying and energy supply
S3	Food, beverages and tobacco
S4	Textiles and leather
S5	Coke, refined petroleum, nuclear fuel and chemicals etc.
S6 & S7	Electrical, optical and transport equipment
S8	Other manufacturing
S9	Construction
S10	Distribution
S11	Hotels and restaurants

S12	Transport, storage and communications
S13	Financial intermediation
S14	Real estate, renting and business activities
S15	Non-market services

Regional supply and use tables were constructed using the commodity balance approach where different columns of the tables are proportionally distributed to the regions using regional information on a column total or a proxy for this column total Regional value added was used to distribute the column of industries (output) over the regions. Regional income statistics were used to distribute the demand categories (household demand where x and x were taken together and government demand) over the regions. Gross capital formation is divided into three items: gross fixed capital formation, changes in inventories and changes in valuables. With respect to the gross capital formation regional information on investments is used to regionalize the national total capital formation. Changes in inventories and valuables were constructed using regional value-added as a proxy.

In a similar fashion, this last consideration could also be applied to the two remaining columns in the supply table: trade and transport margins and taxes and subsidies - since their regional variation is also assumed to be proportional to production. All data in the regionalized supply and use tables now are defined, except domestic trade, i.e., trade between regions of the same country. The PBL trade data set (Thissen et al. 2013b,c – see section 3.2 in this paper) were used as a first estimate. For example, the use of a product by an industry is allocated to regions in a country according to the intraregional trade of this product that is available in the original PBL trade database. Again, this is only a first proxy for the trade, since this is endogenously determined in the estimation procedure described in the next section after introducing the PBL regional trade dataset in the subsequent subsection.

3.1 CONNECTING REGIONAL TABLES: REGIONAL TRADE[10]

An important ingredient for the construction of the EUREGIO database set out in this paper, is a unique dataset on regional bilateral trade assembled by PBL (Thissen et al., 2013b, 2013c). This PBL regional trade (PBL/RT) database contains information on trade in 59 product categories (according

[10] This section is adapted from the Appendix in Thissen et al. (2013a, pp.179-232) to explain its use in the construction of the regional tables used in this paper with permission of Edgar Elgar Publishing.

to CPA) between 256 European NUTS2 regions for the year 2000, and an update of this dataset for the period 2000-2010. The PBL/RT dataset was constructed by bringing together data from several sources as no complete trade surveys exist at this detailed geographical scale. The main sources used are (1) the national accounts of the 25 selected countries in a supply and use format (these are also used in the EUREGIO database described in this paper), (2) international trade data on goods from Feenstra et al. (2005) and on services from Eurostat (2009), (3) regional information on production, investment and consumption made available in Cambridge Econometrics (2008) and Eurostat's regional accounts, (4) information on freight transport among European regions from the Dutch Ministry of Infrastructure and the Environment (2007), and (5) first and business class airline ticket information from MIDT (2010)[11]. Additional regional and national information is gathered from Eurostat to update the data for the period 2000 to 2010. The PBL bi-regional panel trade dataset describes the most likely trade flows between European regions given all available information. As in the EU/REG/IO database, no spatial structure has been imposed on the data. In other words, no specific model is used to estimate trade patterns[12]. For a full understanding of these data in the light of the EUREGIO database, we discuss the elementary ingredients of it in the four steps of its construction.

3.1.1 Trade flows at the country level

First, a consistent international trade matrix of flows in goods and services was created between all 25 European countries and with the rest of the world divided into several blocks[13]. International trade in goods is based on the data collected by Feenstra et al. (2005). Trade in services is based on Eurostat trade statistics taken from the balance of payments (Eurostat, 2009). These two sources were the best available for international trade at the time of construction. However, they are not always consistent with the national accounts or the national use and supply tables. Moreover, trade in goods (Feenstra et al., 2005) are based on the 4-digits SITC and therefore require a conversion and an aggregation and corrections for c.i.f./f.o.b. inconsistencies and re-exports had to be done. Trade in services, instead, is divided in only four macro-categories and required a disaggregation.

[11] MIDT data includes *all* tickets booked via Global Distribution Systems. Tickets directly booked with airlines are missing, but only a very small part of the total market in the year 2000.

[12] For instance, research on the validity of a gravity model based on data generated by a gravity model will by definition result in the confirmation of the validity of the model.

[13] For two out of 25 considered countries (Latvia and Greece), national accounts were not available. Accounts from the year 1998 have therefore been updated using the commonly applied RAS method (or bi-proportional updating method). The necessary row and column sums for the supply and use tables of Latvia and Greece were taken from Eurostat.

3.1.2 Cross-hauling and regional imports and exports

The second step in the construction of the PBL regional trade database concerned estimating regional exports and imports for all distinguished NUTS2 regions in Europe. Regional trade also has to take into account *intranational* trade, i.e., products sold outside the producing region but within the same country, or consumption by a region of products produced elsewhere in the same country. These estimates of *intranational* trade incur in a difficulty when regions simultaneously import and export the same type of goods. As this phenomenon, known as cross-hauling, is empirically relevant, it invalidates regionalization procedure that do not account for it (see Kronenberg, 2009).

As we stress the importance of avoiding imposing a structure on trade patterns, particular attention has been dedicated to solving the problem of cross-hauling.

The Krugman trade model (Kruman 1980, 1991) is a convenient international trade model, as it allows for cross-hauling. As Krugman noted in its seminal work in 1980, previous trade models could not explain two-way exchange of differentiated products. We define an approach, then, that bases itself on the Krugman model and guarantees consistency with the national accounts.

The core of the model is built around a Dixit-Stiglitz-Krugman framework, which assumes the different varieties of a good are not identical (perfectly substitutable). Consumers have in fact a preference for consuming many varieties of a good ('love for variety'), a feature that is used to explain why a location would simultaneously export and import the same type of good . A CES demand function (constant elasticity of substitution) with iceberg transport costs (originally from Samuelson 1954, but included in several trade models including Krugman 1991) determines demand intensity from different locations. . As we want to avoid imposing model assumptions on the definition of trade patterns, we only use a two-region variant of the Krugman model to determine the size of cross-hauling in every region. *not* to determine trade patterns.

The two-region model we propose iteratively focuses on a single region under investigation (the focus region *r)* and the rest of the country *h* (the second region). The second region is a fictional aggregation of all regions in the country except the focus region and, therefore, its definition changes for every focus region. . For each product *g* firms are distributed over both regions. Consumers pay a higher price when the buy a product imported from the other region, as transport costs are inevitably larger compared to those of local goods[14] Given the theoretical model, trade from region *r* to the rest of the country *h* equal to

[14] This is known in the literature as the common assumption of mill pricing.

$$IX^r = \frac{(P^r T_{rh})^{1-\sigma}}{n^h (P^h T_{hh})^{1-\sigma} + n^r (P^r T_{rh})^{1-\sigma}} ID^h n^r \quad [4]$$

Where *P* is price, *n* is number of firms, □ is elasticity of substitution and *T* are iceberg transport costs. The index *g* for the product category was omitted to simplify the equations. *IX* is the production of region *r* sold to other regions of the same country (intranational exports) and *ID* is the total demand in the composite region *h*[15], which is satisfied either by local firms or by the production in other regions of the same country (intranational imports, *II*).

Mirroring exports, intranational imports of region *r* from the rest of the country *h* are:

$$II^r = \frac{(P^h T_{hr})^{1-\sigma}}{n^r (P^r T_{rr})^{1-\sigma} + n^h (P^h T_{hr})^{1-\sigma}} ID^r n^h \quad [5]$$

With homogeneous technology we have that:

$$n^r = \frac{IY^r}{\alpha P^r} \quad [6]$$

with α equals optimal output per variety, which we model to be constant across locations. Substituting the number of varieties [24] in equations [22] and [23] and then combining them allows us to express intranational exports and internal imports as a function of values (instead of prices or quantities). This is convenient as typically, as well as in our case, data on quantities and prices are not available.

$$IX^r = f(T, \sigma, ID^r, ID^h, IZ^r) \equiv II^h$$

And [7]

$$II^r = f(T, \sigma, ID^r, ID^h, IZ^h) \equiv IX^h$$

[15] Note that, to lighten the text, the index for goods has been omitted

where IZ is the trade balance of the region (internal trade balance). Internal imports (*II*) and exports (*IX*) now depends on only two unknown quantities: transport costs and the elasticity of substitution[16]. We take the elasticity of substitution from the literature where it is commonly assumed to be equal to 1.5 (McKitrick, 1998). Transport costs are estimated simultaneously with internal trade using non-linear programming. The values of the transportation costs paramets t_{rh} are determined in the nonlinear optimization, imposing that the national transport costs by product are as close as possible to the national accounts' data on trade and transport margins. We further assume a common transport costs function, which is declining in transported distance). The cross-hauling, equivalent to the intranational trade of a region - is endogenously determined in the procedure.

Since the data on trade and transport margins by product is only available at the national level,the methodology is strengthened by extending the nonlinear optimization procedure with a second objective, based on freight data of the Dutch Ministry of Infrastructure and the Environment (2007). We derive cross-hauling from freight data as the share of goods staying within regional borders over the goods sold to other regions. The share for services, instead, is calculated as follows

$$ChS_r^{services} = ChS_r^{goods} \frac{\sum_{g \in services} Supply_g^N \Big/ \sum_{g \in services} \left(X_g^N + Supply_g^N\right)}{\sum_{g \in goods} Supply_g^N \Big/ \sum_{g \in goods} \left(X_g^N + Supply_g^N\right)} \qquad [8]$$

The correction is based on the relative propensity of exporting services in comparison to goods. Taking all elements together leads to the following nonlinear minimization problem.

[16] The explicit functional forms is:

$$IX_r = \frac{-(IZ_r t_{rr} t_{hh} + ID_h t_{rh}^2 - IZ_r t_{rh}^2 + ID_r t_{rh}^2) + \sqrt{\begin{array}{l}[IZ_r t_{rr} t_{hh} + ID_h t_{rh}^2 - IZ_r t_{rh}^2 + ID_r t_{rh}^2]^2 + \\ -4[t_{rr} t_{hh} - t_{rh}^2][ID_h IZ_r t_{rh}^2 - ID_h ID_r t_{rh}^2]\end{array}}}{2(1 - t_{rh}^2)} \equiv I_h$$

where $t_{ij} = T_{ij}^{1-\sigma}$.

$$Min \quad Objective = Z_1 + Z_2$$

$$s.t.$$

$$Z_1 = \sum_c \left[TTM_c - \sum_{r \in c} \left(T_{rh} - 1\right) IX^r - \sum_{r \in c} \left(T_{rr} - 1\right)(IY^r - IX^r) \right]^2$$

$$Z_2 = \sum_r \left[ChS_r IY_r - (IY^r - IX^r) \right]^2$$

$$t_{rh} = \left(1 + \ln\left[\beta + \gamma dist_{rh}\right]\right)^{1-\sigma}$$

$$(t_{rr}t_{hh} - t_{rh}^2)IX_r^2 + (IZ_r t_{rr} t_{hh} + ID_h t_{rh}^2 - IZ_r t_{rh}^2 + ID_r t_{rh}^2)IX_r + (ID_h IZ_r t_{rh}^2 - ID_h ID_r t_{rh}^2) = 0 \qquad [9]$$

In [27] there are seven free variables: the objective variables (Z_1 and Z_2), the transportation variables (, , *t* and *T*) and intranational exports (*IX*). All remaining elements are parameters, whose values can be taken from the available data sources. Solving the minimization gives us the amount of cross-hauling per region.

3.1.3 Interregional trade flows of goods and services

The third step in the construction of the PBL/RT database concerned the approximation of interregional trade flows. These have been determined using freight transport data from the Dutch Ministry of infrastructure and the environment (2007) based on the Eurostat micro data and business flight ticket information from MIDT (2010). The former is used to estimate the network of trade of goods, while the latter is used to estimate the network of trade in services. The procedure used to determine the trade flows, distributes the trade over the regions given the amounts produced and consumed in every region. The amount of goods and services that are produced and consumed in the same region have been determined in the previous section. Thus, the diagonal of the trade matrix is known. The amount of produced goods and services leaving a region and the amount of produced goods and services entering a region are therefore also determined in the regionalized Supply and Use tables. These regional 'exports' are divided into international exports (those that leave the country) and intranational exports (those that go to different regions in the same country).

In the used estimation methodology both the international as well as the intranational trade were distributed simultaneously. This implies that both 'compete' for a destination in the export estimate (or an origin in the import estimate). The international and the intranational quantities are therefore not predetermined. They are endogenously determined in the process based on the probabilities of trade flows between regions.

The existence of transshipment locations makes the derivation of trade from transport data a complex procedure. The consequence of transshipment locations are that goods transported may be going to a transshipment location instead of going to their final destination. Therefore there may be a large difference between transport and trade data. Survey Data for the Netherlands (Dutch Ministry of infrastructure and the environment, 2007) reveal that only 40 per cent of *all* goods traded does not make use of any transshipment location and reaches its final destination directly. The rest of the goods uses at the least one transshipment location before reaching their final destination. Especially in the case of international trade we expect more than one transshipment location: one transshipment location in the country of origin and another transshipment location in the country of destination. The existence of transshipment locations is therefore explicitly taken into account in our procedure and since only in the Netherlands data was collected on the share of direct flows relative to indirect flows using transshipment locations this number of 40 percent is also used for other regions in Europe.

The procedure followed in the PBL/RT data is based on a combination of two estimates for the international trade between nuts2 regions. The first estimate is based on the export of goods (the destination), while the second estimate is based on the imports of goods (the origin). Both estimates are given a weight of a half and then the quadratic difference between the final trade matrix and the two estimates is minimized. The methodology to determine the two estimates consists of three steps. All steps can be viewed either from the export or import perspectives of regions.

In the first step the direct flows are determined. These are the traded goods and services that are directly transported from the region of origin to the region of destination without the use of any transshipment location. In this step, the production of every region is distributed over the destination regions. In this trade flow estimation the distribution of intra and inter-national export trade flows are not pre-determined with respect to their destination regions. Only the overall intercountry flows are predetermined. In order to determine direct trade flows from an export perspective, one needs the direct probability $P_{i,k}^{0}$ of goods being exported from a region i to a region k .This probability can be described as follows.

$$P_{i,k}^{0} = \frac{T_{i,k}}{\sum_{k} T_{i,k}} \qquad [10]$$

Where $P_{i,k}^0$ is equal to the probability that a good is transported from *i* to *k* without using a transshipment location and $T_{i,k}$ is the data on the amount of goods (or number of trips) transported from origin i to destination k derived from the earlier mentioned survey data of the Dutch ministry of Infrastructure and the Environment. The probabilities for services trade are derived from first and business class ticket information on airline flights (MIDT).

Given the (intranational and international) exports X_j that are exported from region *j* one can describe the direct flows of exports X_{ik}^0 from origin i to destination k in the following way.

$$X_{ik}^0 = \lambda P_{i,k}^0 X_i \qquad [11]$$

Where λ is a fraction of the goods which on average are transported directly (this is 40 percent of the goods). However, in some regions, and especially in transhipment locations, there will be not enough goods imported (I_k) if compared to the amount of goods exported to these regions. We therefore introduce the following parameter Z_k^0.

$$Z_k^0 = \max\left(-I_k + \lambda \sum_i P_{i,k}^0 X_i, 0\right) \qquad [12]$$

The exports to an area *k* is larger than the imports in this area if Z_k is positive. In those cases we have to adjust our equation [30] to determine the trade flows as follows.

$$X_{ik}^0 = \lambda P_{i,k}^0 X_i - \frac{Z_k^0}{\lambda \sum_i P_{i,k}^0 X_i} \lambda P_{i,k}^0 X_i \qquad [13]$$

The two equations [30] and [31] are sufficient to determine the direct trade flows between all regions.

In the second step, the indirect flows are determined. These indirect trade flows are goods and services that use at least one transshipment location when being transported from the producer to the

consumer. Again, looking from the export perspective, the probability $P_{i,k}^{1}$ that a good is transported from *i* to *k* using a transshipment location, the export amount X_{j}^{1} which still have to be distributed over the regions of destination and the import amount I_{k}^{1} that has still to be imported in the region are calculated as follows.

$$P_{i,k}^{1} = \sum_{j} P_{i,j}^{0} P_{j,k|k \neq i}^{0} \quad [14]$$

$$X_{i}^{1} = X_{i} - \sum_{j} X_{ij}^{0} \quad [15]$$

$$I_{k}^{1} = I_{k} - \sum_{i} X_{ik}^{0} \quad [16]$$

Where a condition is imposed such that goods cannot be transported back to a region where they have been produced or reloaded. The indirect trade flows for services are now based on a somewhat adjusted probability matrix. This probability matrix is now added with flows to other neighboring regions that do not have an airport based on regional goods trade.[17] Analogue to the direct trade flows one then can determine whether the estimate of the exports using a transshipment location will generate more exports to regions than Imports available in that region.

$$Z_{k}^{1} = \max\left(-I_{k}^{1} + \sum_{i} P_{i,k}^{1} X_{i}^{1}, 0\right) \quad [17]$$

The exports to an area *k* are larger than the imports in this area if Z_{k} is positive. In those cases the exports are adjusted such that the exports to a region are always smaller or equal to the imports in that region. Indirect trade flows are therefore determined by the following equation.

[17] Note that the stepwise procedure does not need to have a full matrix with destinations during every step and not all production has to be allocated in every step. Moreover, since full consumption is enforced in higher order steps it implies that a missing destination in the probability matrix of direct flows is simply filled up in the second step with the second order probabilities. The approach is therefore completely different from a "common" econometric estimation based on a quadratic error to be minimized. It is better to compare the approach to an allocation mechanism distributing production over its possible destinations.

$$X_{ik}^{1} = P_{i,k}^{1} X_{i}^{1} - \frac{Z_{k}^{1}}{\sum_{i} P_{i,k}^{1} X_{i}^{1}} P_{i,k}^{1} X_{i}^{1} \quad [18]$$

In the third step the higher order indirect flows are determined. Higher order trade flows are traded goods or services that are transported via one or more transshipment locations before they reach their final destination. Higher order indirect trade flows use multiple transshipment locations transporting goods from the production to the consumption location. The amount of goods that is using these higher order transshipment locations is simply the amount that cannot be allocated using less transshipment location. We therefore use a minimum estimate to the amount of goods using multiple transshipment locations. Using the same methodology as described with the indirect flows, one can describe the higher order probabilities and trade flows by the following set of equations (again from the export perspective):

$$P_{i,k}^{2} = \sum_{j,l|l\neq i} P_{i,j}^{0} P_{j,l}^{0} P_{l,k|k\neq i,k\neq j}^{0} \quad [19]$$

$$X_{i}^{2} = X_{i} - \sum_{j} X_{ij}^{1} - \sum_{j} X_{ij}^{0} \quad [20]$$

$$I_{k}^{2} = I_{k} - \sum_{i} X_{ik}^{1} - \sum_{i} X_{ik}^{0} \quad [21]$$

$$Z_{k}^{2} = \max\left(-I_{k}^{2} + \sum_{i} P_{i,k}^{2} X_{i}^{2}, 0\right) \quad [22]$$

$$X_{ik}^2 = P_{i,k}^2 X_i^2 - \frac{Z_k^2}{\sum_i P_{i,k}^2 X_i^2} P_{i,k}^2 X_i^2 \quad [23]$$

Any subsequent transshipment location is treated in the same way. The procedure is continued including more transshipment locations until a maximum of 5 transshipment locations. The remaining X_k^n is bi-proportionally distributed such that all exports match with all imports. Total trade between the regions will be the aggregate over all subsets of trade using *n* different numbers of transshipment locations. Thus the following equations describing total exports between regions is then in place

$$X_{ik}^{prior} = \sum_n X_{i,k}^n \quad [24]$$

After estimating the export trade flows we can perform a similar procedure and get the estimated trade flows from the imports perspective. Thus, we try to find the origin of product consumed in a region instead of finding the destination of product produced in a region.

3.1.4 Minimization of distance between trade estimates

The information gathered in the previous steps resulted in two different estimates of the bilateral trade flows between all European regions including those in the same country. One estimate was derived from the export side and the other estimate was derived from the import side. The final PBL/RT matrix is the one that minimizes the quadratic distance between these two estimates. The purpose of this last step is to bring to consistency between our two estimates of regional bilateral trade

3.2 ADDITIONAL INFORMATION FROM SURVEY-BASED REGIONAL SUPPLY AND USE TABLES

Regional supply and/or use tables are available for Scotland and Wales, as well as Italy (five NUTS1 regions), Finland (21 NUTS3 regions) and Spain (15 NUTS2 regions). This additional information has been used as a prior and not as a given constraint to the estimation in the next section because of the following three reasons. First, implementing the regional information as a constraint would automatically allocate all possible errors in the additional regional tables in the regions for which regional information in the same country is lacking. This may cause erroneous results. These errors may be exacerbated for a second reason: there are often definition differences between the “extra” tables at the regional level in comparison to the tables at the national level. This makes the tables only comparable to a certain extent. Third, the regional information does not always coincide with the

NUTS2 aggregation level. Using the additional information as a prior is a transparent approach to incorporate different types of information in one estimation. The following adjustments were made in order to make these tables consistent with the WIOD and the regional tables.

In the case of Scotland, Wales and Italy, the regional supply and/or use tables have a more aggregated regional level than NUTS-2. The set of constraints in the minimization procedure was therefore extended. The additional constraint in the case of Scotland, Wales and Italy is that all the estimated elements in the NUTS-2 SAMs add up to the corresponding elements in the more aggregated Scottish, Welsh and Italian 5-region SAMs.

Below, subscript *g* denotes products and *s* denotes industries.[18] Subscript *k* is used to denote the (number of) NUTS2 regions in a particular region. *x* denotes the value in a given cell in the SAM. Regional use tables for Scotland are the aggregate of (the use tables for) the underlying Scottish Nuts-2 regions. There are four NUTS-2 areas in Scotland.[19] Hence, *k* = 4.

$$x_{gs}^{S\,\mathrm{cot}\,land} = \sum_{k=1}^{4} x_{gs}^{k} \qquad \text{(use)} \qquad g=sp1, \ldots sp65;\ s=ss1, \ldots ss23$$

The Welsh table is the aggregate of (the use tables for) two underlying Nuts-2 regions: West Wales and The Valleys, and East Wales. Hence, *k* = 2.

$$x_{gs}^{Wales} = \sum_{k=1}^{2} x_{gs}^{k} \qquad \text{(use)} \qquad g=sp1, \ldots sp65;\ s=ss1, \ldots ss23$$

For Italy, supply and use tables for North-East Italy, North-West Italy, Central Italy, South Italy and Insular Italy (NUTS1) are available. The same methodology is adopted as for Wales and Scotland. So, e.g., the use table for North-East Italy is the aggregate of (the use tables for) four underlying Nuts-2 regions. Hence, *k* = 4.

$$x_{gs}^{North-East\ Italy} = \sum_{k=1}^{4} x_{gs}^{k} \qquad \text{(use)} \qquad g=sp1, \ldots sp65;\ s=ss1, \ldots ss25$$

Similar treatment is given to North-West Italy, Central, South and insular Italy.

[18] Product classifications in the regional supply and use tables did not always match the NACE Rev.1 classification used in WIOD. In some cases products in the regional tables had to be aggregated and in some cases products in the regional tables had to be split. Regarding the latter, more aggregate product groups in the regional tables were split into the respective WIOD products using shares obtained from the national table for that region.

[19] Eastern Scotland, Highlands and Islands, North Eastern Scotland and South Western Scotland.

Regarding Finland, the supply and use tables are at the NUTS-3 level. The regions (*n*=21) can be aggregated to six NUTS-2 regions.[20] The resulting six regional supply and use tables are subsequently used as a prior.

Scottish supply and use tables are available annually for the period 1998–2009. Hence, the methodology described in the previous paragraph can be applied for nearly every year in the database. On the other hand, the Welsh use table is available for 2007 only, whilst the Finnish tables are for 2002. The Italian regional tables cover the years 2001 and 2006. This section describes the methodology used to extrapolate the results based on prior information for a given year to the full set of years in the database.

In order to create a time series of regional supply and use tables we first put all the tables in coefficients (percentages of row and column totals). These coefficients are being used in the regionalization method described below. We also calculated the change in these coefficients on the national level for all years between 2000 and 2010. We used these national changes in the coefficients to change the regional coefficients to obtain a time series of regional supply and use tables expressed in coefficients. Multiplying these coefficients with the regional totals for the years 2000-2010 gave us the regional supply and use tables for these years.

4. THE CONSTRUCTION OF INTERREGIONAL SUPPLY AND USE TABLES FOR THE YEARS 2000-2010

Taking the regionalized supply and use tables and the PBL regional trade data as a prior, the interregional supply and use tables are estimated for the years 2000-2010 using a constrained non-linear minimization procedure. The objective function in the non-linear optimization minimized the quadratic distances between the coefficients of the regional SAM matrix in relation to the coefficients of the national SAM matrix given regional data on value-added, fixed capital formation and household demand. A quadratic minimization function was preferred over a logarithmic function (often used in entropy minimization, e.g., Thissen and Lofgren, 1998) because of its mathematical properties that made it possible to solve the problem using conic programming which reduces computation time dramatically and made it possible to solve such a huge mathematical optimization problem (mosek linear programming package).The optimization was constrained in such a way that

[20] See the table in the annex.

all cells of the regional supply and use tables add up over the regions to the national cells presented in the WIOD supply and use tables. Moreover, by imposing the constraint that all exports of a region should be equal or smaller than production it is not possible to have regional re-exports.

The size of the constrained non-linear minimization problem is such that the procedure is best solved in two consecutive independent steps. In a first step, domestic trade and international trade from regions to countries was determined. In this step also the regional supply and use tables were determined. The second step involved subdividing the international trade from European regions to countries into trade between regions. Throughout this process, all normal consistency rules were applied, so that the amount of products exported from one region to another (destination) region or country would equal the amount imported into that destination region or country from that particular region of origin. Re-exports were not allowed on the regional level. The methodology for constructing the updated data set is outlined in more detail in the next sections.

4.1 INTERREGIONAL SOCIAL ACCOUNTING MATRIX (SAM)

For the sake of convenience, the supply and use data were first entered into a social accounting matrix (SAM) framework (SNA, 1993; 2008). Using such a complete national accounts framework in matrix format, has the advantage that all consistency checks can be performed immediately. Thus, the imported amount of products into region B from region A is, per definition, exactly the same as the exported amount of this product from region A to region B, as this amount is recorded in only one position in the matrix. In general, valuations in a SAM are in nominal terms. Its rows and columns list institutional agents or actors. The matrix shows the flow of goods between actors, from row to column, balanced by an opposite flow of money from column to row. The SAM framework also illustrates the methodology used. When additional information is used to update or improve the SAM matrix overall consistency enforces changes in different parts of the SAM matrix. Thus, a change in regional coefficients derived from regional supply and use tables induces changes in coefficients in other regions to keep the requirement that national coefficients do not change, and affect trade to guarantee that regional use matches the amount of goods supplied in the region. In all these cases we minimize the change in the structure of the matrix based on information available and More recent information was used from national and regional accounts to impose requirements that had to be met while minimizing any structural distance to the elements of the matrix on which no new information was available. Changes in regional demand or production have a direct impact on regional trade here, because what is exported must be produced and what is imported should represent demand. This minimization of the structural distance is applied by keeping the changes in

the relative numbers of the matrix to a minimum. The consistency of the system of national and regional accounts in a SAM framework, therefore, provides a large amount of information on regional trade developments.

Figure 2: A stylised interregional Social Accounting Matrix

SAM		Region 1: Use		Region 2: Use		Region 1: Supply		Region 2: Supply		Region 1: Use	Region 2: Use	total
		sector 1	sector 2	sector 1	sector 2	product 1	product 2	product 1	product 2	hhd - Investment	hhd - investment	
Region1	sector 1					3	7					10
	sector 2					2	6					8
Region 2	sector 1							5	2			7
	sector 2							3	6			9
Region 1	product 1	1	1	1						1	1	5
	product 2	2	3		2					5	1	13
Region 2	product 1	1		2	1						4	8
	product 2		1							1	6	8
Region 1	value added	6	3							transfer	transfer	
Region 2	value added			4	6					transfer	transfer	19
total		10	8	7	9	5	13	8	8		19	x

A stylized version of the SAM is presented in Figure 2, which distinguishes two regions. The SAM consists of a framework in which the use table is combined with the transposed supply table. The two regions distinguished in the SAM have two sectors, 1 and 2, producing two types of products. A product in the SAM, thus, has three characteristics: type of good, the region and the sector of production. The use of this product by different sectors in different regions is presented in the first four columns of the SAM. The production of these products is presented on the top 4 rows of the SAM. Total production in these sectors is provided in the last column on the right. In the bottom two rows one finds the total value-added which is an aggregation of both labour and capital income.

Table 4: Non-Industrial actors and other accounts distinguished in the SAM

S16	Final consumption expenditure by households and non-profit organisations
S17	Final consumption expenditure by government
S18	Net capital formation
S19	Inventory adjustment
S20	Trade and transport margins

S21	Taxes less subsidies on products
S22	Exports (outside Europe)
S23	Imports (outside Europe)

International trade takes place at the product level and is divided over different types of use. The use of the final demand categories is presented in the last two columns of the SAM. The use by the different producing sectors is presented in the first two columns. Thus, Region 1 exports a total of 3 units of product 2 to Region 2, 2 of which are used by Sector 2 and 1 unit is used by the final demand categories. The final demand for goods per region does not equal the total value-added earned in that same region. Only when interregional savings would be specified in the transfer part of the SAM these would be equal. In the SAMs that were applied, the transfer part was not specified because information on these flows was lacking and also not needed for the updating procedure. In the stylised presentation of the SAM, therefore, all value-added rows and all final demand columns were added to obtain equality between total value-added and total final demand. All other rows add up to the same respective columns representing the bookkeeping rules that for every actor total expenditure should equal total income.

The regional SAMs used for constructing EUREGIO distinguish the 14 industry categories presented in Table 3, and the non-Industrial actors and other accounts presented in Table 4. Products are classified according to the 2-digit Classification of Products by Activity (CPA, 1996) presented in Table 5. There is a total of 62 goods and services in CPA 2002. Yet, products with numbers 96, 97 and 99 (goods produced by households for own use, services produced by households for own use and services provided by extra-territorial organizations and bodies) are not included in the supply and use system of accounts, reducing the total number of products analyzed in this study to 59.

4.2 INTRANATIONAL TRADE AND EXPORTS OF REGIONS

First, constrained non-linear optimization is used to determine the trade of regions with regions within the same country (intranational trade), and the total exports and imports (international trade) of these region. The destination of the exports and the origin of the imports are determined in the

second step discussed in the next subsection. Please note that these two steps are completely independent because of the given bi-country trade matrix which is imposed as a constrained on the non-linear optimization problem.

Equation [43] shows the constrained non-linear quadratic minimization problem to be solved to estimate the intranational trade and associated regional coefficients of the SAM matrix. The function describes how new information is used to find updated matrices, given the growth in production and demand indicated in the national and regional accounts. In general, the change in the structure of the demand, supply and regional trade pattern was minimized, given new information on for instance regional value-added and international trade. The regionalization is based on region-specific coefficients. This procedure is very different from the commonly applied commodity balance method (Isard, 1953). The complete minimisation problem can be described in a simplified form as follows.

$$Min\ Z_c = \sum_{i \in c}\left[\left(\hat{a}_i^c - a_i^c\right)^2 + \left(\hat{a}_i^r - a_i^r\right)^2\right] + \sum_{i' \in c}\left[\left(\hat{a}_{i'}^c - a_{i'}^c\right)^2 + \left(\hat{a}_{i'}^r - a_{i'}^r\right)^2\right] \\ + \text{ other constraints} \qquad [25]$$

Where the index i stands for the region, a_i^c represents the elements of the regional SAM matrix divided by its column total, a_i^r represents the elements of the regional SAM matrix divided by its row total. The index i' stands fort he regions with additional information and $_{i'}$ is the number of NUTS2 regions that are part of the region i'. In equation [43] we therefore describe the minimization only by presenting the relative errors. In the actual minimization we use the methodology developed in Thissen et al. (2013) and explained in Appendix B to minimize the weighted average of both the absolute and relative error.

For matters of convenience, the summation over the elements of the SAM matrix were left out. The SAM matrix is aggregated for all the imports from regions and countries except the regions from the own country. Thus in this first step of the procedure only the domestic trade and non-trade coefficients of the regional SAM matrices are determined. The international trade is determined in the second step of the procedure. Since national trade is predetermined this first step in the procedure was done for all countries separately.

<u>The constraints on the objective function</u>

The objective function was constrained to generate outcomes conform the regional and national accounts in WIOD. Moreover, a non-negativity condition of trade flows was added to the procedure. Below, all constraints used are discussed along with the information they contain.

1. All elements summed over all regions in a country add up to the same elements in the national WIOD-based SAM. This constraint guarantees that the regional SAMs are completely compatible with the WIOD database.
2. All products sold by an economic agent are received and paid for by another economic agent. This bookkeeping rule was adhered to by imposing the equality of all row and column totals of the SAM for all activities (classified according to industry) and products.
3. Information was available on regional value-added for the distinguished industries. Value-added of the sectors in the regions was therefore fixed.
4. Information was available on regional total household demand and regional total gross fixed capital formation. These items have therefore been fixed also in the procedure.
5. Finally, a 'no re-export' constraint was applied to ensure that production would always exceed exports, for every region and product. Please note that this constraint should be imposed on the product level and not on the industry level.

Solving the minimization problem under these constraints resulted in the update of the regional SAM, including domestic trade from the year 2000 to the year 2010.

4.3 INTERNATIONAL TRADE BETWEEN REGIONS

The first step of the procedure, described above, yields estimates of the total international trade as an export or import column of the regional supply and use tables). These international trade flows were subsequently divided into regions of destination and regions of origin, resulting in a full regional origin–destination matrix. No additional information was available on these trade patterns, except the international trade between countries from the WIOD database and the regional trade from the PBL/RT data. Constrained non-linear quadratic optimization was used to combine this information with existing total exports and imports for the different regions to determine the final panel data on trade between NUTS2 regions for the 2000–2010 period. The following objective function was used in this second step in the estimation of the supply and use tables:

$$
\begin{aligned}
Min\ Z' &= \sum_{i,j} \left(\hat{}_{s,i,j} - \ _{s,i,j} \right)^2 \\
&s.t. \\
&1)\quad Ex_{s,c,d} = \sum_{i \in c, j \in d} \hat{}_{s,i,j} \\
&2)\quad Im_{s,d,c} = \sum_{i \in c, j \in d} \hat{}_{s,i,j} \\
&3)\quad Ex_{s,i \in c} = \sum_{j \notin c} \hat{}_{s,i,j} \\
&4)\quad Im_{s,j \in c} = \sum_{i \notin c} \hat{}_{s,i,j}
\end{aligned}
\qquad [26]
$$

in which $Ex_{c,d}$ denotes exports of country c destined for country d and $Im_{c,d}$ denotes imports in country d from country c (both directly taken from the WIOD tables). The estimated regional trade $\hat{}_{s,i,j}$ on the sector level (s) is completely consistent with the WIOD tables because of constraints 1) and 2). The priors of exports and imports $_{s,i,j}$ were determined by the regional trade pattern of exports (imports) from the PBL regional trade database for the year 2000. The regional exports $Ex_{s,i}$ and imports $Im_{s,i}$ were taken from the first step in the procedure.

In this second estimation that determines the international trade flows between regions, contrary to the first step, only absolute quadratic errors are minimized. The reason for only using absolute errors is related to the numerical properties for the problem at hand. The international bi-regional trade flows are often very small numbers. This is aggravated by using estimated trade flows from the PBL trade database that are usually even smaller with very few zero's. A relative error would put a strong emphasis on these small numbers. However, a trade flow of a value of, for example, 10 eurocents is far less reliable than a trade flow of 10 million euro. As a consequence using relative errors in the estimation of these regional international trade flows would result in erroneous results[21].

5. REGIONAL IO TABLES

After the estimation of regional supply and use tables it is a relatively easy step to create regional IO tables. Moreover, consistency of these IO tables is guaranteed because the supply and use tables are fully consistent with the trade linked WIOD database and have equal row and column totals. The preferred method D (Eurostat, 2008) was followed to determine the IO tables. Although there is a substantive amount of literature on the choice of methods to determine IO tables from supply and use tables, choices were here partly driven by the available data. Our supply and use tables are non-square, i.e. the number of products is not equal to the number of sectors. As a consequence the majority of methods that involve the inverse of either the supply or the use table cannot be used in our case. Moreover, the method used was also used in the construction of the WIOD IO tables, strengthening the comparability of both tables.

[21] Please note that intranational trade flows are also much larger than international trade flows.

6. CONCLUSION

In this paper we showed how the EUREGIO database of consistent European interregional IO tables at the NUTS2 regional division level over the years 2000-2010 is constructed. The longstanding need for consistent interregional input-output (IO) tables mentioned in the introduction (e.g., Isard, 1953; Hewings and Jensen, 1987) resulted already in an extensive use of the EUREGIO database in studies coverings several subfields in regional economics.

The first use of the database is based on the analysis of spatial revealed competition, identifying regional economic competition structures and their policy implications (Thissen et al. 2013, Van Oort et al., 2017). This research was followed by a dynamic analysis of spatial competition making use of a spatial growth decomposition method (Thissen et al. 2016a). This research uses the EUREGIO database in a spatial econometric estimation of a regional growth model. More recently the EUREGIO database starts being used in different econometric analysis such a panel data analysis to determine the economic effects of Brexit (Baltagi et al., 2017).

The EUREGIO database is especially equipped to be used in structural modelling that addresses regional trade issues. It can be directly used in multiregional Input-Output analysis such as done in Los et al. (2017) or value chain analysis using a hypothetical extraction directly on the database (Chen ta al., 2018). These studies both used the EUREGIO database to analyse the short run effects of the Brexit on European regions (Los et al., 2017), and value chain analysis using hypothetical extraction to determine the exposure of European regional economies to Brexit.

Extending these demand driven IO models with a supply side in supply and use modelling was done in Koks et al. (2016) addressing the regional spillover effects of flooding. In CGE models prices are introduced as equilibrating mechanism to equal regional supply and demand. The EUREGIO database was also central in the European Commission's regional CGE model (Mercenier et al. 2016) where the use of the data in this model is discussed in Thissen et al (2014).

These studies shed new light on recent policy-relevant phenomena, and by using the EUREGIO database they are able to explicitly take into account both differences in regional specialization and multiregional supplier-user linkages.

7. REFERENCES

Acemoglu, D. and P. Restrepo (2017), "The race between man and machine: Implications of technology for growth, factor shares and employment". Working Paper, MIT.

Andrew, R. M., and G. P. Peters (2013), "A multi-region Input-Output table based on the Global Trade Analysis Project Database (GTAP-MRIO)". *Economic Systems Research* 25(1): 99-121.

Autor, D., Dorn and G. Hanson (2013), "The China Syndrome: Local labor market effects of import competition in the United States". *American Economic Review* 103(6): 2121–2168.

Autor, D. , D. Dorn, G. Hanson and K. Majlesi (2017), "A Note on the effect of rising trade exposure on the 2016 presidential election". Working Paper, MIT.

Bachtler, J., C. Mendez & F. Wishlade (2013), *EU cohesion policy and European integration. The dynamics of EU budget and regional policy reform*. Alderhot: Ashgate.

Baldwin, R. E. (2006), "Globalisation: The great unbundling(s)", in *Globalisation Challenges for Europe*. Helsinki: Office of the Prime Minister of Finland.

Becker, S., T. Fetzer and D. Novy (2017), "Who voted for Brexit? A comprehensive district-level analysis". *Economic Policy* (forthcoming).

Chen, W., B. Los, P. McCann, R. Ortega-Argilés, F. van Oort and M. Thissen (2017), "The continental divide? Economic exposure to Brexit in regions and countries on both sides of the Channel". *Papers in Regional Science*.

Cherubini, L. and B. Los (2012), "Regional employment patterns in a globalizing world: A tale of four Italies", Paper presented at the 20th International Input-Output Conference, Bratislava, Slovak Republic.

Combes, P.P., G. Duranton, L. Gobillon and S, Roux (2010), "Estimating agglomeration economies with history, geology, and worker effects". In: E, Glaeser (ed.), *Agglomeration economics*. Chicago: The University of Chicago Press, pp. 15-66.

Derudder B., & Witlox F. (2005), An Appraisal of the Use of Airline Data in Assessing the World City Network: a Research Note on Data. Urban Studies, 42(13), pp. 2371-2388.

Diodato D. and M. Thissen (2011), Towards a New Economic Geography based Estimate of Trade Elasticity and Transport Costs. Working Paper.

Diodato D. and A. Weterings (2012), The Resilience of Dutch Regions to Economic shocks. Measuring the relevance of interactions among firms and workers. Utrecht: Utrecht University, Paper in Evolutionary Economic Geography.

Dixit A.K., and J.E. Stiglitz (1977), Monopolistic competition and optimum product diversity, American Economic Review 67: 297-308.

Eurostat (2008), Eurostat Manual of Supply, Use and Input-Output Tables. Eurostat Methodologies and Working Papers, Luxembourg: office for official publications of the European Communities.

Eurostat (2009), International trade in services. Data for years from 2000 to 2004, downloaded in 2009.

Eurostat (2009b), National use and supply tables. Data from year 2000 for Austria, Belgium, Czech Republic, Germany, Denmark, Estonia, Spain, Finland, France, Hungary, Ireland, Italy, Lithuania, Luxemburg, Malta, the Netherlands, Norway, Poland, Portugal, Sweden, Slovenia, Slovakia, United Kingdom. Data from year 1998 for Greece and Latvia. Downloaded in 2009.

Eurostat (2009c), 'High-technology and knowledge based services aggregations based on NACE Rev. 2', http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/htec_esms_an3.pdf. Brussels: Eurostat.

Dietzenbacher, E., Guilhoto, J., and Imori, D. (2012), "The role of Brazilian regions in the global value chain", Paper presented at the 59th Annual North American Meetings of the Regional Science Association International. Ottawa, Canada

Dietzenbacher E, B. Los, R. Steher, M.P. Timmer and G.J. de Vries GJ. (2013), "The construction of World Input-Output tables in the WIOD project". *Economic Systems Research* 25: 71–98.

Feenstra, R.C., R.E. Lipsey, H. Deng, A.C. Ma and H. Mo (2005), "World Trade Flows: 1962–2000". National Bureau of Economic Research.

Fingleton, B., H. Garretsen & R. Martin (2012), "Recessionary shocks and regional employment: evidence on the resilience of UK regions". *Journal of Economic Geography* 15: 907-934.

Foray, D. (2015), *Smart specialization. Opportunities and challenges for regional innovation policy*. London: Routledge.

Groot, S.P.T., J.H. Mohlmann, J.H. Garretsen & H.L.F. de Groot (2011), "The crisis sensitivity of European countries and regions: stylized facts and spatial heterogeneity". *Cambridge Journal of Regions, Economy and Society* 4: 437-456.

Hewings, G.J.D. and R.C. Jensen (1987), "Regional, interregional and multiregional input-output analysis". In G. Duranton, J.V. Henderson and W. C. Strange (eds.), *Handbook of Regional and Urban Economics* Volume 1, pp. 295-355.

Isard W. (1953), "Regional commodity balances and interregional commodity flows". *American Economic Review* 43: 167–80.

Koks E.E. and M. Thissen (2016): A Multiregional Impact Assessment Model for disaster analysis, Economic Systems Research: 429-449, DOI:10.1080/09535314.2016.1232701

Kronenberg T., (2009). Construction of Regional Input-Output Tables Using Nonsurvey Methods. The Role of Cross-Hauling. International Regional Science Review 32: 40-64

Krugman P.R. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review* 70: 950-959.

Krugman P.R. (1991), Geography and trade. Cambridge, Mass.: The MIT Press.

Lankhuizen M. and M. Thissen M. (2018), "The implications of re-exports for gravity equation estimation, NAFTA and Brexit". Paper presented at the 58th ERSA Congress "Places for People: Innovative, Inclusive and liveable Regions" 28- 31 August 2018, Cork, Ireland.

Lenzen, M., D. Moran, K. Kanemoto and A. Geschke (2013), "Building Eora: A global multi-region Input-Output database at high country and sector resolution". *Economic Systems Research* 25(1): 20-49.

Los, B., McCann, P., Springford, J., and Thissen, M., 2017, "The mismatch between local voting and the local economic consequences of Brexit". *Regional Studies* 5: 786-799.

Los, B., M. Lankhuizen and M. Thissen (2016), "New measures of regional competitiveness in a globalizing world", in: P. McCann, F. van Oort and J. Goddard (eds.), *The Empirical and Institutional Dimensions of Smart Specialisation*, London: Routledge: 105-126.

McCann, P. (2015), *The regional and urban policy of the European Union. Cohesion, results-orientation and smart specialization*. Cheltenham: Edward Elgar.

Meng, B. and N. Yamano (2017), "Compilation of a regionally extended intercountry Input-Output table and its application to global value chain analysis". *Journal of Economic Structures* 6, article 23.

Meng, B., Y. Zhang and S. Inomata (2013), "Compilation and applications of IDE-Jetro's International Input-Output Tables". *Economic Systems Research* 25(1): 122-142.

Mercenier, J., M.T. Álvarez-Martinez, A.Brandsma,F. Di Comite, O. Diukanova, A. Kancs, P. Lecca, M. Lopez-Cobo, P. Montfort, D. Persyn, A. Rillaers, M. Thissen, W. Torfs (2016), Rhomolo-v2 Model Description, A spatial computable general equilibrium model for EU regions and sectors, JRC technical reports, doi 10.2791/18446.

McKitrick, R. (1998), The econometric critique of computable general equilibrium modeling: The role of functional forms. *Economic Modelling* 15: 543-574.

MIDT (2010). Regional flights, business and first class. Data for november 2000.

Ministry of Infrastructure and the Environment (2007). Interregional freight. Data for years from 2000-2004.

OECD (2009), *Regions matter. Economic recovery, innovation and sustainable growth*. Paris: OECD.

Piattoni, S. and L. Polverari (2016) (eds.), *Handbook on cohesion policy in the EU*. Cheltenham: Edward Elgar.

Samuelson, P. A. (1954). The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments. *Economic Journal* 64: 264-289.

Simini, F., M.C. González, A. Maritan and A. Barabási (2012). "A universal model for mobility and migration patterns". *Nature* 484: 96–100.

Stone, R. (1963), "Input–output relationships 1954–1966". The Department of Applied Economics University of Cambridge, Chapman and Hall, Cambridge.

System of National Accounts (1993 and 2008). System of National Accounts (SNA). United Nation website: http://unstats.un.org/unsd/nationalaccount/sna.asp

Timmer, M.P., E. Dietzenbacher, B. Los, R. Stehrer and G.J. de Vries (2015), "An illustrated user guide to the World Input-Output Database: The case of global automotive production". *Review of International Economics* 23(3): 575-605.

Timmer, M.P., A.A. Erumban, B. Los, R. Stehrer and G.J. de Vries (2014), "Slicing up global value chains". *Journal of Economic Perspectives* 28: 99–118.

Thissen M and H. Löfgren (1998), "A new approach to SAM updating with an application to Egypt". *Environment and Planning A* 30: 1991–2003.

Thissen, M., F. van Oort, D. Diodato and A. Ruijs A. (2013a), *European regional competitiveness and Smart Specialization; Regional place-based development perspectives in international economic networks*. Cheltenham: Edward Elgar.

Thissen, M., D. Diodato D and F. van Oort (2013b), “Integrated regional Europe: European regional trade flows in 2000”. PBL Netherlands Environmental Assessment Agency, The Hague.

Thissen, M., D. Diodato D and F. van Oort F. (2013c), “European regional trade flows: An update for 2000–2010”. PBL Netherlands Environmental Assessment Agency, The Hague.

Thissen, M., F. van Oort and O. Ivanova (2016a), “Good growth, bad growth: A wake-up call of smart specialisation”. In: P. McCann, F. van Oort and J. Goddard (eds.), *The empirical and institutional dimensions of Smart Specialisation*. London: Routledge: 1-104.

Thissen, M., T. de Graaff and F. van Oort (2016b), “Competitive network positions in trade and structural economic growth: a geographical weighted regression analysis for European regions”. *Papers in Regional Science* 95: 159-180

Thissen M., Francesco Di Comite, d'Artis Kancs and Lesley Potters (2014), Modelling Inter-Regional Trade Flows; Data and Methodological Issues in RHOMOLO 1, WP 02/2014, ISBN: 978-92-79-44509-5, doi: 10.2776/871154, European Commission, Directorate-General for Regional Policy, Bruxelles

Tukker, A., A. e Koning, R. Wood, T. Hawkins, S. Lutter, J. Acosta, J.M.Rueda-Cantuche, M. Bouwmeester, J. Oosterhaven, T. Drosdowski and J. Kuenen (2013), “EXIOPOL - Development and Illustrative Analyses of a Detailed Global MR EE SUT/IOT”. *Economic Systems Research* 25(1): 50-70.

Tukker, A. and E. Dietzenbacher (2013), “Global multiregional Input-Output frameworks: An introduction and outlook”. *Economic Systems Research* 25(1): 1-19.

Van Oort, F. and M. Thissen (2017), “Regional economic competition and place-based policies”. In: R. Huggins and P. Thompson (eds.), *Handbook of regions and competitiveness. Contemporary theories and perspectives on economic development*. Cheltenham: Edward Elgar: 192-206

Wang, Y. , A. Geschke and M. Lenzen (2017), “Constructing a time series of nested multiregion Input-Output tables”. *International Regional Science Review* 40: 476-499.

APPENDIX A: COUNTRIES, REGIONS AND PRODUCT CATEGORIES IN THE DATA SET

This appendix gives an overview on the countries, regions and products covered by the EUREGIO dataset.

A.1 Countries

Table A.1 lists the EU-25 countries, with the exception of Cyprus and including Norway, for which regional information is available in EUREGIO. No (comparable) supply and use tables were available for Switzerland, Romania and Bulgaria at the start of the research on regional trade.

Table A.1 Countries in the dataset

L1	Austria	L11	Hungary	L21	Portugal
L2	Belgium	L12	Ireland	L22	Sweden
L3	Czech Republic	L13	Italy	L23	Slovenia
L4	Germany	L14	Lithuania	L24	Slovakia
L5	Denmark	L15	Luxembourg	L25	United Kingdom
L6	Estonia	L16	Latvia		
L7	Spain	L17	Malta		

The dataset also covers the trade of the distinguished European regions with the rest of the world. This rest of the world has been split up in main economic countries and groups of less important countries. These extra Europe trading partners are given in the Table 2 .

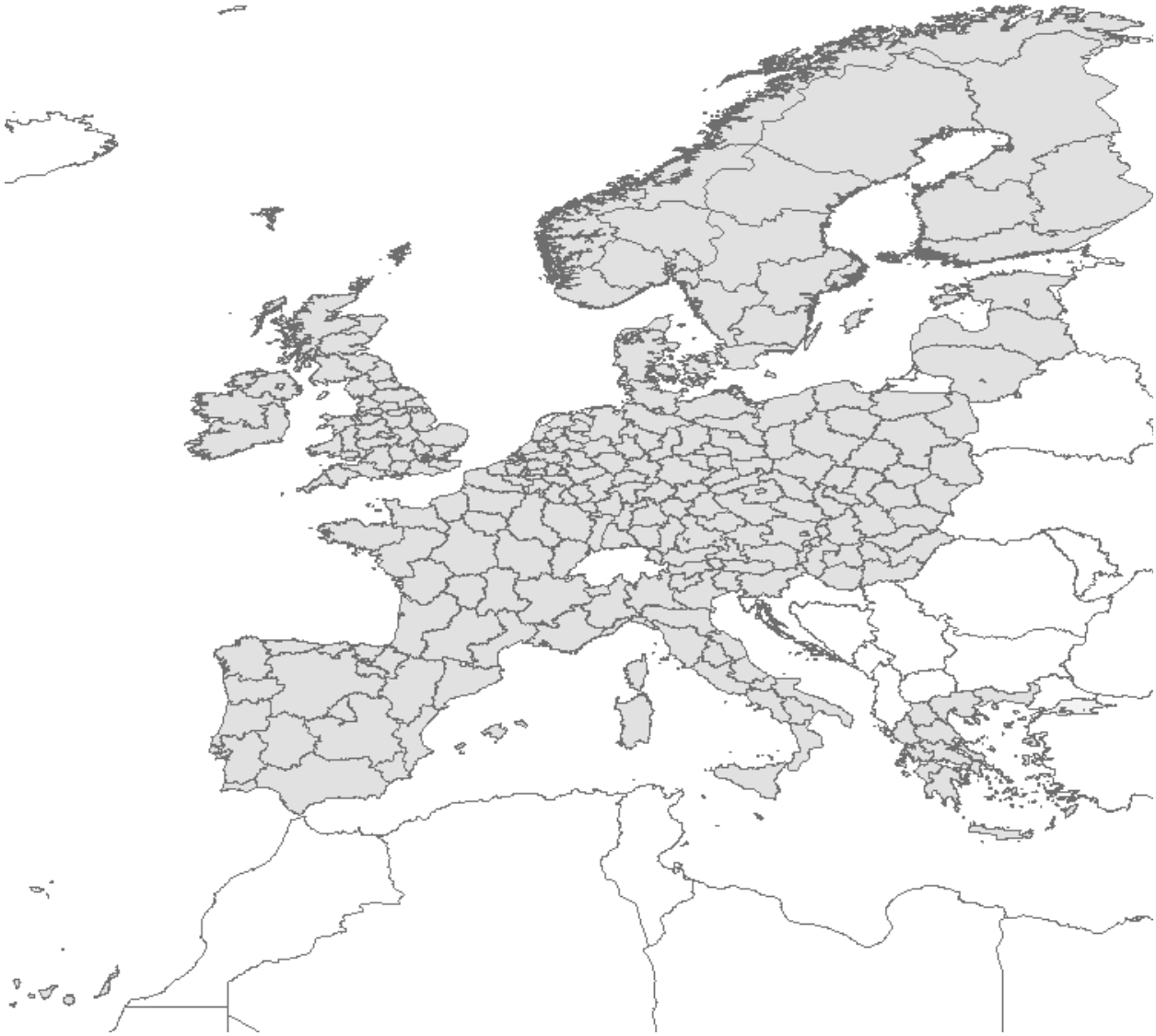
Table A.2 Extra-Europe trading partners

L26	Rest of Europe	L35	Cyprus
L27	Africa	L36	Canada
L28	Asia	L37	China
L29	Japan	L38	Hong Kong
L30	Middle and South America	L39	Korea
L31	Australia and Oceania	L40	Singapore

A.2 Regions

The EUREGIO dataset follows the NUTS2 regional classification for the EU. The classification distinguishes 256 European Nuts2 regions. Details are given in Table A3 and Figure A1.

Figure A.1: Overview Map of European regions covered in EUREGIO



The overview map shows that the data cover the largest part of Europe. The missing countries are Switzerland, Romania and Bulgaria.

Table A.3 Regions in the NUTS2 classification

Code	Name	Code	Name	Code	Name
AT11	Burgenland	FI1A	Pohjois-Suomi	PL51	Dolnoslaskie
AT12	Niederosterreich	FI20	Aland	PL52	Opolskie
AT13	Wien	FR10	Ile de France	PL61	Kujawsko-Pomorskie
AT21	Karnten	FR21	Champagne-Ardenne	PL62	Warminsko-Mazurskie
AT22	Steiermark	FR22	Picardie	PL63	Pomorskie
AT31	Oberosterreich	FR23	Haute-Normandie	PT11	Norte
AT32	Salzburg	FR24	Centre	PT15	Algarve
AT33	Tirol	FR25	Basse-Normandie	PT16	Centro (PT)
AT34	Vorarlberg	FR26	Bourgogne	PT17	Lisboa
BE10	Region de Bruxelles	FR30	Nord - Pas-de-Calais	PT18	Alentejo
BE21	Prov. Antwerpen	FR41	Lorraine	PT20	Região Autónoma dos Açores
BE22	Prov. Limburg (B)	FR42	Alsace	PT30	Região Autónoma da Madeira
BE23	Prov. Oost-Vlaanderen	FR43	Franche-Comte	SE11	Stockholm
BE24	Prov. Vlaams Brabant	FR51	Pays de la Loire	SE12	ostra Mellansverige
BE25	Prov. West-Vlaanderen	FR52	Bretagne	SE21	Sydsverige
BE31	Prov. Brabant Wallon	FR53	Poitou-Charentes	SE22	Norra Mellansverige
BE32	Prov. Hainaut	FR61	Aquitaine	SE23	Mellersta Norrland
BE33	Prov. Liege	FR62	Midi-Pyrenees	SE31	ovre Norrland
BE34	Prov. Luxembourg (B)	FR63	Limousin	SE32	Småland med oarna
BE35	Prov. Namur	FR71	Rhone-Alpes	SE33	Västsverige
CZ01	Praha	FR72	Auvergne	SK01	Bratislavský kraj
CZ02	Stredni Cechy	FR81	Languedoc-Roussillon	SK02	Zapadne Slovensko
CZ03	Jihozapad	FR82	Provence-Alpes-Cote d Azur	SK03	Stredne Slovensko
CZ04	Severozapad	FR83	Corse	SK04	Východne Slovensko
CZ05	Severovychod	GR11	Anatoliki Makedonia Thraki	UKC1	Tees Valley and Durham
CZ06	Jihovychod	GR12	Kentriki Makedonia	UKC2	Northumberland Tyne and Wear
CZ07	Stredni Morava	GR13	Dytiki Makedonia	UKD1	Cumbria
CZ08	Moravskoslezko	GR14	Thessalia	UKD2	Cheshire
DE11	Stuttgart	GR21	Ipeiros	UKD3	Greater Manchester
DE12	Karlsruhe	GR22	Ionia Nisia	UKD4	Lancashire
DE13	Freiburg	GR23	Dytiki Ellada	UKD5	Merseyside
DE14	Tubingen	GR24	Sterea Ellada	UKE1	East Riding, North Lincolnshire
DE21	Oberbayern	GR25	Peloponnisos	UKE2	North Yorkshire
DE22	Niederbayern	GR30	Attiki	UKE3	South Yorkshire
DE23	Oberpfalz	GR41	Voreio Aigaio	UKE4	West Yorkshire
DE24	Oberfranken	GR42	Notio Aigaio	UKF1	Derbyshire and Nottinghamshire
DE25	Mittelfranken	GR43	Kriti	UKF2	Leicestershire Rutland
DE26	Unterfranken	HU10	Kozep-Magyarorszag	UKF3	Lincolnshire
DE27	Schwaben	HU21	Kozep-Dunantul	UKG1	Herefordshire Worcestershire
DE30	Berlin	HU22	Nyugat-Dunantul	UKG2	Shropshire and Staffordshire
DE41	Brandenburg - Nordost	HU23	Del-Dunantul	UKG3	West Midlands
DE42	Brandenburg - Südwest	HU31	eszak-Magyarorszag	UKH1	East Anglia
DE50	Bremen	HU32	eszak-Alfold	UKH2	Bedfordshire Hertfordshire
DE60	Hamburg	HU33	Del-Alfold	UKH3	Essex
DE71	Darmstadt	IE01	Border Midlands	UKI1	Inner London

DE72	Giessen	*IE02*	Southern and Eastern	*UKI2*	Outer London
DE73	Kassel	*ITC1*	Piemonte	*UKJ1*	Berkshire Bucks Oxfordshire
DE80	Mecklenburg-Vorpommern	*ITC2*	Valle dAosta Vallee dAoste	*UKJ2*	Surrey East and West Sussex
DE91	Braunschweig	*ITC3*	Liguria	*UKJ3*	Hampshire and Isle of Wight
DE92	Hannover	*ITC4*	Lombardia	*UKJ4*	Kent
DE93	Luneburg	*ITD1*	Bolzano-Bozen	*UKK1*	Gloucestershire Wiltshire
DE94	Weser-Ems	*ITD2*	Provincia Autonoma Trento	*UKK2*	Dorset and Somerset
DEA1	Dusseldorf	*ITD3*	Veneto	*UKK3*	Cornwall and Isles of Scilly
DEA2	Koln	*ITD4*	Friuli-Venezia Giulia	*UKK4*	Devon
DEA3	Munster	*ITD5*	Emilia-Romagna	*UKL1*	West Wales and The Valleys
DEA4	Detmold	*ITE1*	Toscana	*UKL2*	East Wales
DEA5	Arnsberg	*ITE2*	Umbria	*UKM2*	North Eastern Scotland
DEB1	Koblenz	*ITE3*	Marche	*UKM3*	Eastern Scotland
DEB2	Trier	*ITE4*	Lazio	*UKM5*	South Western Scotland
DEB3	Rheinhessen-Pfalz	*ITF1*	Abruzzo	*UKM6*	Highlands and Islands
DEE0	Saarland	*ITF2*	Molise	*UKN0*	Northern Ireland
DED1	Chemnitz	*ITF3*	Campania	*BG31*	Severozapaden
DED2	Dresden	*ITF4*	Puglia	*BG32*	Severen tsentralen
DED3	Leipzig	*ITF5*	Basilicata	*BG33*	Severoiztochen
DEF0	Schleswig-Holstein	*ITF6*	Calabria	*BG34*	Yugoiztochen
DEG0	Thüringen	*ITG1*	Sicilia	*BG41*	Yugozapaden
DK01	Hovedstadsreg	*ITG2*	Sardegna	*BG42*	Yuzhen tsentralen
DK02	Ost for Storebælt	*LT00*	Lietuva	*CY00*	Kypros/Kıbrıs
DK03	Syddanmark	*LU00*	Luxembourg (Grand-D)	*SI01*	Vzhodna Slovenija
DK04	Midtjylland	*LV00*	Latvija	*SI02*	Zahodna Slovenija
DK05	Nordjylland	*MT00*	Malta		
EE00	Eesti	*NL11*	Groningen		
ES11	Galicia	*NL12*	Friesland		
ES12	Principado de Asturias	*NL13*	Drenthe		
ES13	Cantabria	*NL21*	Overijssel		
ES21	Pais Vasco	*NL22*	Gelderland		
ES22	Foral de Navarra	*NL23*	Flevoland		
ES23	La Rioja	*NL31*	Utrecht		
ES24	Aragon	*NL32*	Noord-Holland	*JPN*	Japan
ES30	Comunidad de Madrid	*NL33*	Zuid-Holland	*BRA*	Middle and South America
ES41	Castilla y Leon	*NL34*	Zeeland	*AUS*	Austrialia and Oceania
ES42	Castilla-la Mancha	*NL41*	Noord-Brabant	*MEX*	Northern America
ES43	Extremadura	*NL42*	Limburg (NL)	*RUS*	Russia
ES51	Cataluna	*PL11*	Lódzkie	*BGR*	Bulgaria
ES52	Comunidad Valenciana	*PL12*	Mazowieckie	*ROU*	Roumania
ES53	Illes Balears	*PL21*	Malopolskie	*IND*	India
ES61	Andalucia	*PL22*	Slaskie	*IDN*	Indonesia
ES62	Region de Murcia	*PL31*	Lubelskie	*CAN*	Canada
ES63	Ceuta (ES)	*PL32*	Podkarpackie	*CHN*	China
ES64	Melilla (ES)	*PL33*	Swietokrzyskie	*KOR*	Korea
ES70	Canarias (ES)	*PL34*	Podlaskie	*TUR*	Turkey

FI13	Ita-Suomi	*PL41*	Wielkopolskie	*USA*	United States
FI18	Etela-Suomi	*PL42*	Zachodniopomorskie	*TWN*	Taiwan
FI19	Lansi-Suomi	*PL43*	Lubuskie	*ZROW*	Rest of the World

A.3 Product categories

Trade between European regions is detailed at the product level. Export and imports flows are divided according to the 2-digit *Classification of Products by Activity* (CPA 1996), which is the system used by Eurostat to publish comparable national accounts of European countries. Table 4 summarizes the product classification.

Table A.4 The 2-digit Classification of Products by Activity (CPA, 1996)

AA01	Products of agriculture, hunting and related services
AA02	Products of forestry, logging and related services
BA05	Fish and other fishing products; services incidental of fishing
CA10	Coal and lignite; peat
CA11	Crude petroleum and natural gas; services incidental to oil and gas extraction excluding surveying
CA12	Uranium and thorium ores
CB13	Metal ores
CB14	Other mining and quarrying products
DA15	Food products and beverages
DA16	Tobacco products
DB17	Textiles
DB18	Wearing apparel; furs
DC19	Leather and leather products
DD20	Wood and products of wood and cork (except furniture); articles of straw and plaiting materials

DE21	Pulp, paper and paper products
DE22	Printed matter and recorded media
DF23	Coke, refined petroleum products and nuclear fuels
DG24	Chemicals, chemical products and man-made fibres
DH25	Rubber and plastic products
DI26	Other non-metallic mineral products
DJ27	Basic metals
DJ28	Fabricated metal products, except machinery and equipment
DK29	Machinery and equipment n.e.c.
DK30	Office machinery and computers
DL31	Electrical machinery and apparatus n.e.c.
DL32	Radio, television and communication equipment and apparatus
DL33	Medical, precision and optical instruments, watches and clocks
DM34	Motor vehicles, trailers and semi-trailers
DM35	Other transport equipment
DN36	Furniture; other manufactured goods n.e.c.
DN37	Secondary raw materials
EA40	Electrical energy, gas, steam and hot water
EA41	Collected and purified water, distribution services of water
FA45	Construction work
FA50	Trade, maintenance and repair services of motor vehicles and motorcycles; retail sale of automotive fuel
GA51	Wholesale trade and commission trade services, except of motor vehicles and motorcycles
GA52	Retail trade services, except of motor vehicles and motorcycles; repair services of personal and household goods
HA55	Hotel and restaurant services

IA60	Land transport; transport via pipeline services
IA61	Water transport services
IA62	Air transport services
IA63	Supporting and auxiliary transport services; travel agency services
IA64	Post and telecommunication services
JA65	Financial intermediation services, except insurance and pension funding services
JA66	Insurance and pension funding services, except compulsory social security services
JA67	Services auxiliary to financial intermediation
KA70	Real estate services
KA71	Renting services of machinery and equipment without operator and of personal and household goods
KA72	Computer and related services
KA73	Research and development services
KA74	Other business services
LA75	Public administration and defence services; compulsory social security services
MA80	Education services
NA85	Health and social work services
OA90	Sewage and refuse disposal services, sanitation and similar services
OA91	Membership organisation services n.e.c.
OA92	Recreational, cultural and sporting services
OA93	Other services
PA95	Private households with employed persons

APPENDIX B: CONSTRAINT NONLINEAR MINIMIZATION OF WEIGHTED ABSOLUTE AND RELATIVE ERRORS

In constraint nonlinear minimization often a choice is made between relative or absolute errors. This choice is ad hoc and may generate erroneous results by putting too much emphasis on the small numbers (in case of relative errors) or large numbers (in case of absolute errors). The most important part of the supply and use framework are the technical coefficients. These coefficients are crucial in, for instance, Input-output analysis. Hence, it makes sense to minimize the quadratic error of these coefficients. However, the strong weight that is given to small numbers and the combination of intranational trade coefficients and the regional input output coefficients makes that the results of the minimization were biased: Absolute values are just as important as relative values.

We found it therefore often more useful to have a balanced approach where both absolute and relative errors are being minimized. In order to have both relative and absolute errors in the constraint nonlinear minimization we use the methodology introduced in Thissen et al. (2013). Using this methodology the minimized error are related to an observed value (or a derived prior) for the variable to be estimated. In this case the relative errors determined by dividing the deviation of the estimated value from the observed value of a variable by the observed value of this variable. The absolute errors instead are divided by the average value of the observed variable. In this way both absolute as well as relative errors are equally weighted in the minimization problem.

In mathematical terms we can describe this as:

$$
\begin{aligned}
Min\ Z^{"} &= \widehat{re} + \widehat{ae} \\
&s.t. \\
\widehat{re} &= \sum_i \frac{1}{b_i}\left(\widehat{b_i} - b_i\right)^2 \\
\widehat{ae} &= \sum_i \frac{1}{\overline{b}}\left(\widehat{b_i} - b_i\right)^2 \\
&constraints
\end{aligned}
\qquad [27]
$$

Where a quadratic absolute $\widehat{ae}$ and a quadratic relative error $\widehat{re}$ of an estimated variable $\widehat{b_i}$ relative to a prior value b_i and an average value $\overline{b}$ are both minimized simultaneously.