Data Sources and Methodology of the 60-Industry Database of the Groningen Growth and Development Centre

Version: October 2005

Introduction

The 60-Industry Database of the Groningen Growth and Development Centre provides a comprehensive internationally comparable dataset on industrial performance at a detailed industry level for 26 OECD countries and Taiwan (see table 1). It enables the user to design indicators such as shares of output and employment by industry and to analyse trends in value added and labour productivity. Variables covered include current value added, value added deflators, persons engaged and hours worked for 56 industries for the period 1979-2003.[1]

The 60-Industry Database updates and extends previous work at the Groningen Growth and Development Centre (GGDC) and the National Institute on Economic and Social Research (NIESR) recently described in O'Mahony and van Ark (2003) (downloadable from http://www.ggdc.net/pub/EU_Productivity_and_Competitiveness.shtml). For most variables and countries, the OECD STructural ANalysis (STAN) database is taken as the point of departure, which in turn is largely based on recent national accounts of individual OECD member states.[2] In some cases, own countries' national accounts data is directly used for reasons of greater detail or more up-to-date figures. The STAN data and those from the national accounts are complemented, updated and backdated as well as further disaggregated by the use of information from industry surveys and historical sources for individual countries to provide a comprehensive data set for the period 1979-2003. In general the method employed was to use national accounts aggregates as control totals and the other data to divide these totals into sub-industries.[3] In a limited number of cases, STAN and national accounts also include working hours per person employed (or per job), but for many countries alternative sources, such as labour force surveys or other labour market statistics, had to be employed to obtain these estimates.

[1] Series on the number of employees and labour compensation are not available for all countries.
[2] See http://www.oecd.org/document/15/0,2340,en_2649_33703_1895503_119656_1_1_1,00.html
[3] Appromixately 50 per cent of individual cells in the 60-Industry Database is directly derived from OECD STAN, whereas the other 50 per cent represents extensions.

Country Coverage

Table 1 provides a list of countries covered in the database. The database covers 26 countries, of which 26 are OECD members as well as Taiwan. All 15 'OLD' EU countries[4] are covered, and aggregate estimates for the EU as a whole are also provided.

Table 1 List of countries in databases

Belgium	BE	United Kingdom	UK
Denmark	DK	United States	US
Germany*	DE	Canada	CA
Greece	GR	Australia	AU
Spain	ES	Norway	NO
France	FR	Switzerland	CH
Ireland	IR	Japan	JP
Italy	IT	South Korea	SK
Luxembourg	LU	Taiwan	TA
Netherlands	NL	Czech Republic	CZ
Austria	AT	Hungary	HU
Portugal	PT	Poland	PO
Finland	FI	Slovak Republic	SL
Sweden	SW		

*separate series are provided for West Germany for 1979-1991
and for unified Germany from 1991-2003

Industry coverage

Data are provided for 56 industries in total.[5] The industries are classified according to the International Standard Industrial Classification (ISIC) revision 3. This classification is very close to the European NACE rev 1 classification system. Table 2 provides a listing of the industries. The industry division is more detailed than in STAN which allows a focus on industries which are characterised by high ICT-investment shares and/or ICT-goods production. For example, additional entries were included for five of the six ICT producing industries (semiconductors, ISIC 321; communication equipment, 322, fiber optics, 313; radio and TV equipment, 323; and instruments, 331). Separate entries are also included for professional and "other" business services (ISIC 741-743 and 749) and agriculture and forestry (01 and 02).

[4] 'OLD' EU-15 refers to membership of the European Union until 30 April 2004.

[5] Despite the coverage of 56 industries, we have called the database "60-Industry database" as we aim to expand the industry detail a little further in the near future. For example, in due course, "Mining and quarrying" will be divided into "Mining and quarrying of energy producing materials" and "Mining and quarrying except energy producing materials", Food products, beverages and tobacco" will be divided into "Food products and beverages" and "Tobacco products", "Chemicals and chemical products" will be divided into "Chemicals excluding pharmaceuticals" and "Pharmaceuticals", "Basic metals" will be divided into "Iron and steel" and "Non-ferrous metals", "Manufacturing nec; recycling" will be divided into "Manufacturing nec" and "Recycling", "Electricity, gas and water supply" will be divided into "Electricity" and "Gas and water supply".

Table 2 Industries in the 60-Industry database

	Industry Name	ISIC rev 3
	TOTAL ALL INDUSTRIES	**01-99**
1	Agriculture	01
2	Forestry	02
3	Fishing	05
4	Mining and quarrying	10-14
5	Food, drink & tobacco	15-16
6	Textiles	17
7	Clothing	18
8	Leather and footwear	19
9	Wood & products of wood and cork	20
10	Pulp, paper & paper products	21
11	Printing & publishing	22
12	Mineral oil refining, coke & nuclear fuel	23
13	Chemicals	24
14	Rubber & plastics	25
15	Non-metallic mineral products	26
16	Basic metals	27
17	Fabricated metal products	28
18	Mechanical engineering	29
19	Office machinery	30
20	*Insulated wire*	*313*
21	*Other electrical machinery and aparatus nec*	*31-313*
22	*Electronic valves and tubes*	*321*
23	*Telecommunication equipment*	*322*
24	*Radio and television receivers*	*323*
25	*Scientific instruments*	*331*
26	*Other instruments*	*33-331*
27	Motor vehicles	34
28	*Building and repairing of ships and boats*	*351*
29	*Aircraft and spacecraft*	*353*
30	*Railroad equipment and transport equipment nec*	*352+359*
31	Furniture, miscellaneous manufacturing; recycling	36-37
32	Electricity, gas and water supply	40-41
33	Construction	45
34	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	50
35	Wholesale trade and commission trade, except of motor vehicles and motorcycles	51
36	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	52
37	Hotels & catering	55
38	Inland transport	60
39	Water transport	61
40	Air transport	62
41	Supporting and auxiliary transport activities; activities of travel agencies	63
42	Communications	64
43	Financial intermediation, except insurance and pension funding	65
44	Insurance and pension funding, except compulsory social security	66
45	Activities auxiliary to financial intermediation	67
46	Real estate activities	70
47	Renting of machinery and equipment	71
48	Computer and related activities	72
49	Research and development	73
50	*Legal, technical and advertising*	*741-3*
51	*Other business activities, nec*	*749*
52	Public administration and defence; compulsory social security	75
53	Education	80
54	Health and social work	85
55	Other community, social and personal services	90-93
56	Private households with employed persons	95

Variables

The following variables are covered:

Value added is current gross value added measured at producer prices or at basic prices, depending on the valuation used in the national accounts. It represents the contribution of each industry to total GDP. Value added is expressed in national currencies, and the measures are therefore not corrected for differences in purchasing power of the currencies. Value added in Euro-countries before the introduction of the euro are converted to euros with the 1999 official fixed euro conversion rates (see table 3).

Deflator is the change in the value added deflator. It can be combined with current value added to derive quantity indices of real value added at industry level.

Persons engaged comprises number of workers engaged in production, including employees as well as self-employed, working proprietors and unpaid family workers.

Hours refers to average annual hours worked per employee or per person engaged.

Value added volume indices (1995 =100) (for more details see below in text on "aggregation"),

Value added per person employed and **value added per hour worked** are volume indices (1995 =100) and are imputed from the variables reported above.

Table 3: 1999 Euro conversion rates (national currency/Euro)

Austria	13.76	Ireland	0.79
Belgium	40.34	Italy	1936.27
Finland	5.95	Luxemburg	40.34
France	6.56	Netherlands	2.20
Germany	1.96	Portugal	200.48
Greece	340.75	Spain	166.39

Sources

The construction of the 60-Industry Database was done in a two-step procedure to ensure national accounts compatibility. For most countries the basic point of departure is the STAN database. STAN provides data which is generally based on the latest official national accounts data of individual OECD member states. Due to changes in industrial classifications and the introduction of the new 1993 System of National Accounts (SNA) and the 1995 European System of Accounts (ESA), official series often lack industry detail, especially for the period before the 1990s. In some cases, STAN goes beyond the official published data and provides data at more detailed levels based on additional sources, but important gaps still remain.

For a number of countries data were directly obtained from the national accounts instead of from the OECD STAN database. For some countries there is no or only limited OECD STAN

data (i.e. Ireland, Taiwan), whereas for a few other countries national accounts data was more up-to-date than the OECD STAN data.

As a second step the STAN or national accounts data were complemented with information from detailed industry and services statistics and additional (historical) national accounts data for individual countries. Use was made of international statistics such as the *OECD Structural Statistics for Industry and Services*, the *OECD Services Statistics on Value Added and Employment, Eurostat Structural Business Statistics* and the *Eurostat Labour Force Survey*. Furthermore national sources have been used including national accounts and economic census and survey material. The spreadsheets associated with each of the country files provide a full account of the sources used for each country, each year and each variable.

Filling procedure

In case of missing data there are basically two procedures for estimating value added, employment and compensation data by industry: (1) applying shares from additional data to higher level aggregates or (2) applying higher-level growth rates to more detailed levels. The first is most useful when for a particular sub-sector there is no data available for any year. In that case, the share of the sub-sector in some higher level aggregate is derived from additional secondary data sources and applied to the aggregate in the basic source. In case data is available in the basic source for some years, secondary data shares are used for missing years provided they correspond closely to the basic source. If not, growth rates from secondary data are applied to the original basic data for missing years. To maintain national accounts compatibility a normalisation procedure is used so that subsectors add to the corresponding higher-level industry aggregates provided in the basic source. If there is a summation discrepancy, the subsectors absorb the residual. Each sub-sector does so in proportion to its weight in the parent industry. This procedure ensures that output and employment measures are national accounts compatible and, importantly, have the same economy-wide coverage.[6]

For series on hours worked per person and deflators for value added, gaps in the basic source (STAN or national accounts data) are filled in a simpler way. If no additional data could be found, higher level aggregates have been used. For example, in case no separate figure on hours worked per person for sub-sector 351 could be found, the figure for 35 (when available) is used instead. The same was done for deflators where necessary.

[6] Often additional data is taken from surveys. Sampling coverage and definitions in survey data can differ within and across countries.

The source descriptions in each of the country files provide a detailed account of the filling procedures used for each country, year and variable.

Alternative deflators for ICT production

At present there are only a few countries that have an adequate system in place for measuring prices of computers and semiconductors which take into account the fast increase in quality of these goods. To achieve international comparability, harmonised US deflators are applied for two ICT-producing manufacturing industries (ISIC 30 and 321) in all countries. US value added deflators are corrected for differences in overall inflation between each country and the US. Inflation is measured as the change in the deflator of all industries, excluding the ICT-producing manufacturing industries.[7] Since the ICT-producing industries are not separately distinguished in the US National Income and Product Accounts, these deflators were constructed using price changes in output and intermediate inputs of ICT goods as follows:

1) Value of shipments deflators for manufacturing industries and gross output deflators for non-manufacturing industries were obtained from the BEA data sets on "Gross Output by Detailed Industry" and "Shipments of Manufacturing Industries". A Törnqvist index was applied to obtain gross output (value of shipments) deflators for each of the industries.
2) The Input/Output (I/O) tables from the Bureau of Labor Statistics (BLS)[8] for 1977-2003 are then used to calculate an intermediate input deflator for each ISIC industry. For each of the 190 I/O industries a gross output deflator series was calculated.[9] These deflators are used to calculate an intermediate input price index for each industry. For industry *i* this is done in the following way:

$$\Delta \ln P_i^M = \sum_j \bar{s}_{i,j}^M \Delta \ln P_{i,j}^M \quad (1)$$

In equation (1) $P_{i,j}^M$ is the price of the j^{th} intermediate input used in industry *i*. The price change for this input is weighted by the average share of input *j* in total intermediate inputs in current prices of the industry over the two periods:

$$\bar{s}_{i,j}^M = \frac{1}{2}\left(\frac{P_{i,j}^{M,t} M_{i,j}^t}{\sum_j P_{i,j}^{M,t} M_{i,j}^t} + \frac{P_{i,j}^{M,t-1} M_{i,j}^{t-1}}{\sum_j P_{i,j}^{M,t-1} M_{i,j}^{t-1}} \right). \quad (2)$$

[7] This procedure is based on Schreyer (2000, 2002).
[8] Specifically from the Office of Occupational Statistics and Employment Projections (www.bls.gov/emp). These tables are used since they are available for each year in the sample and because the industry detail is much greater (190 versus 96 industries).
[9] If there is a many-to-one correspondence, a Törnqvist index is used to aggregate to the level of aggregation of the I/O table.

3) Using the deflators for gross output and intermediate inputs, gross output and intermediate inputs at constant prices was calculated. These are combined to calculate real value added growth :

$$\Delta \ln V_i = \frac{1}{\bar{s}_i^V}\left(\Delta \ln Q_i - \bar{s}_i^M \Delta \ln M_i\right) \qquad (3)$$

Here, V_i, Q_i and M_i are the quantity indices of value added, gross output and intermediate inputs respectively. Furthermore:

$$s_i^V = 1 - s_i^M = \frac{P_i^t Q_i^t - P_i^{M,t} Q_i^{M,t}}{P_i^t Q_i^t} \quad (4)$$

s_i^V is the share of current value added in current gross output. The average over the two periods is taken and used in equation (3). Finally, the value added deflator is derived as the difference between the growth rate of current and real value added. While this procedure does not exactly replicate the BEA procedure, it serves as a good approximation since the aggregate deflators are close to the original value added deflator from the national accounts.[10]

Aggregation

Some countries at present still use fixed-weight (Laspeyres) indices to calculate aggregate value added at constant prices. This can lead to serious substitution bias if the structure of the economy is changing over time. For example, when fixed weights are used the price decline for computers will be overstated because of the relatively large weight in the base year compared to successive years (Landefeld and Grimm, 2000). To correct for this problem, chain-weighted indices like chained Laspeyres or chained Fisher, or Törnqvist indexes are needed.

In (re)calculating real value added aggregates for industry groups and for the aggregate economy chain-weighted (Törnqvist) deflators for value added were used. For industry *i* this is done in the following way. Let P_i^V the deflator for value added in industry *i* and P^V the aggregate deflator, then the change of the deflator in period t is given by:

[10] Differences occur for a number of reasons: the BEA uses the detailed source material of the I/O tables, which is more disaggregated than the 190-industry table used here. Furthermore a single deflator was used for all inputs from a certain commodity category, while the BEA distinguishes between domestically produced and imported goods. Also, the price deflators used here correspond to the value of shipments of an *industry* not of the *commodities* that are used as inputs. Finally, the BEA uses Fisher aggregation instead of Törnqvist.

$$\Delta \ln P^{V,t} = \sum_i \bar{s}_i^{V,t} \Delta \ln P_i^{V,t} \qquad (5)$$

The price change in industry *i* is weighted by the average share of industry *i* in total value added over the two periods defined as:

$$\bar{s}_i^{V,t} = \frac{1}{2}\left(\frac{P_i^{V,t} V_i^t}{\sum_i P_i^{V,t} V_i^t} + \frac{P_i^{V,t-1} V_i^{t-1}}{\sum_i P_i^{V,t-1} V_i^{t-1}} \right). \qquad (6)$$

Finally, real aggregate value added growth is calculated as the growth rate of aggregate current value added[11] minus the growth rate of the deflator as follows:

$$\Delta \ln V^t = \Delta \ln V^t P^{V,t} - \Delta \ln P^{V,t} \qquad (7)$$

It should also be stressed that the use of Törnqvist aggregation diverts from simply adding up value added estimates at constant prices from the lowest industry level to higher aggregates. In particular when underlying industries show strongly different growth rates (such as ICT-producing industries relative to non-ICT industries), a summed result may deviate from a Törnqvist weighted result. Hence intermediate aggregates, such as a series for say total manufacturing, can only be obtained by again applying the Törnqvist aggregation procedure.

European Union Aggregation

In order to arrive at totals for the European Union, current price measures in national currencies are converted into euro's using sector specific purchasing power parities (PPPs).[12] The PPPs for all countries in this database stem from the Multilateral 1997 Industry-of-origin PPP set, which has been described in more detail in Timmer, Ypma and van Ark (2006, forthcoming)[13]. For agriculture (see van Ark, Ypma and Rao, 2006, forthcoming), manufacturing (O'Mahony and van Ark, 2003, chapter 7), trade (Timmer and Ypma, 2006, forthcoming) and transport and communication (Ypma, Timmer and van Ark, 2006, forthcoming) different methods for calculating PPPs have been used. These publications provide extensive analysis of the development of the Unit Value Ratios. The industry PPPs were all converted to a 1997 basis, and then applied to obtain value added shares by industry

[11] Aggregate current value added is the sum of industry current value added .

[12] See van Ark and Timmer (2003) for a discussion of approaches to obtain sector-specific PPPs. See also the discussion on the Productivity and Unit Labour Cost Database in O'Mahony and B. van Ark (2003), chapter 7.

and by country in the total EU aggregate for the whole period 1979-2003. These were then used to Törnqvist aggregate the deflators for individual industries to an EU aggregate.

German unification

In order to deal with German unification in 1990 two data sets have been constructed: one for Western Germany (covering the period 1979-1991) and one for unified Germany (covering the period 1991-2003). Growth rates for Western Germany are linked to 1991 data for unified Germany.

US reclassifications

As the US Standard Industrial Classification (SIC) differs considerably from the ISIC rev 3 (and NACE Rev.1) classification, several major adjustments were made to the US National Income and Product Accounts at detailed industry levels. The main reclassifications were carried out for motor vehicle trade and repairs (ISIC 50), retail trade (ISIC 52), hotels and restaurants (ISIC 55), post and telecommunications (ISIC 64), all industries in business services (ISIC 71-74) and Other community, social and personal services (ISIC 90-93). See the sources and methods description in the US file for more details. Price deflators for these industries have been developed in a similar way as for ICT producing industries using a double deflation procedure (see above).

Starting in 2002 the BEA switched from SIC based data to NAICS based data. In the summer of 2004, BEA published industry estimates for the period 1998-2003 based on the NAICS classification.[14] However, at present there is no straightforward reconciliation possible between SIC and NAICS at the level of the 60-Industry database. For the time being we use the 1998-2003 NAICS numbers as starting point and link the trend of the SIC series for the previous period in 1998. A more detailed description of differences in the US data between the present and previous version of the 60 Industry Database is available from the Groningen Growth and Development Centre upon request (ggdc@eco.rug.nl).

[13] Timmer, M.P., G.Ypma and B. van Ark (2006), Industry-of-origin PPPs for the Total Economy: A New Benchmark for 1997, GGDC Research Memorandum GD-82 (forthcoming)

[14] Per December 15, 2005, BEA released series back to 1987 , but these have not yet been incorporated.

References

Ark, B. van and M.P. Timmer (2003), *Measuring Productivity Levels. A Reader*, OECD (draft).

Ark, B. van, G. Ypma and D.S. Prasada Rao (2005), 'Agricultural Purchasing Power Parities and International Comparisons of Agricultural Output and Productivity', GGDC Research Memorandum GD-84 (forthcoming), University of Queensland/Groningen Growth of Development Centre

Eurostat Labour Force Survey, various issues, Luxembourg

O'Mahony, M. and B. van Ark (2003), *EU Productivity and Competitiveness: An Industry Perspective Can Europe Resume the Catching-up Process?*, Office for Official Publications of the European Communities, Luxembourg.

OECD (1996, 1997 & 2001), Services Statistics on Value Added and Employment, Paris

OECD (2002), Purchasing Power Parities and Real Expenditures, Paris.

OECD (2003), Structural Statistics for Industry and Services 1993-2000, Paris

OECD (2005), OECD STAN Database 2005

Schreyer, P. (2000), 'The Contribution of Information and Communication Technology to Output Growth: A Study of the G7 Countries', *STI Working Papers 2000/2*, OECD, Paris.

Schreyer, P. (2002), 'Computer Price Indices and International Growth and Productivity Comparisons', *Review of Income and Wealth*, 48 (1), pp. 15-31, March.

Timmer, M.P., G.Ypma and B. van Ark (2006), 'Industry-of-origin PPPs for the Total Economy: A New Benchmark for 1997', GGDC Research Memorandum GD-82 (forthcoming)

Timmer, M.P., G.Ypma and B. van Ark (2006), 'Productivity Levels in Distributive Trades: A New ICOP Dataset', GGDC Research Memorandum GD-83 (forthcoming)

Ypma, G., M.P. Timmer and B. van Ark (2006), 'Productivity levels in Transport, Storage and Communication: A new ICOP Dataset', GGDC Research Memorandum GD-85 (forthcoming)