

# **TFP Analysis DO.File: A first assessment**

**Chiara Angeloni and Nicola Benatti**  
CompNet

**CompNet Workshop - WS2 Session**  
**Frankfurt, 11 December 2012**

# Background

- **Economic research based on firm level data faces two major constraints: data availability and data confidentiality.**
- **In this respect, CompNet is expected to have a great potential thanks to its network linkages with NCBs.**
- **In June 2012, CompNet has started a country-specific assessment of data availability across its members, with respect to three typologies of firm-level data:**
  - **Trade Data**
  - **Balance Data**
  - **Ownership Structure**
  - + **Possibility of matching data across these categories**
- **The results of this survey were summarized into a table (the WS2 “coloured table”) which highlighted the effective data availability for all the assessed countries, together with the well-known constraints in terms of confidentiality.**
- **The analysis shed light on a great data potential that was not completely exploited.**
- **The DO.file exercises have been proposed as solution to address these confidentiality and availability issues and to better exploit the used data.**

- 1. TFP analysis DO.File**
- 2. Variable Definitions, Indicators and Country-Time Coverage**
- 3. Some initial results**
- 4. The steps ahead**

# 1. TFP analysis DO.File

- **The DO.file contains the STATA commands for the creation of country-level indicators for 11 relevant sectors, for all the (country-specific) available years. It is structured in three sessions:**
  - 1) Performance indicators**
  - 2) TFP**
  - 3) TFP Decomposition (Olley-Pakes Method)**
- **In order to guarantee the highest harmonization, the DO.file, the country specific deflators and the definition of the needed variables have been managed at a centralised level.**
- **The files have been sent to 18 country-teams. 13 country-teams had actively participated in the exercise, delivering the requested indicators (ES, FR, HU, BE, DE, SK, PL, SI, CZ, RO, EE, PT, IT). Additional results were provided also on the EFIGE sample of firms.**
- **Country-specific results are being pooled together and will be available to all CompNet members for research purposes (via a Dropbox Folder).**

## **2. Variable Definitions, Indicators and Country-Time Coverage**

## 2.1 Variable and Sector Definition

Needed Variables	First best	Second best
<b>Number of employees</b>	Full time employment	Total employment at the end of the exercise
<b>Total assets (Capital)</b>	Tangible assets (intangible assets excluded)	Tangible assets + intangible assets (balance sheet value of total assets)
<b>Material costs</b>	Intermediate inputs (even if not separated in energy and other non-storable supplies or raw materials)	
<b>Cost of Employees</b>	Wages + employer's contributions to obligatory social security funds	Any proxy used as numerator in ULC formula
<b>Added value</b>	Production - intermediate inputs  It should be used at basic prices terms, if available.	Total sales (turnover) - intermediate inputs

- For this first exercise all the variables come from **firm balance sheet data**.
- The analysis includes the following sectors:
  1. Mining
  2. Food
  3. Textile
  4. Chemicals
  5. Basic metal
  6. Elect. products
  7. Auto
  8. Electricity
  9. Construction
  10. Telecommunications
  11. Financial activities
- The sectors are base on **NACE rev.2** industry classification.

## 2.2 Performance Indicators

Indicators	Description	Var_list		
		Level	Log	First difference of log Var_list
Number of Employees			lnl	d_lnl
Real value added (RVA)	Value Added/Deflator	rva	lnrva	d_lnrva
Capital/Labour Ratio	K/L	k_l	lnk_l	
Real Value Added/Labour	RVA/L	rva_l	lnrva_l	d_lnrva_l
Capital Productivity	RVA/RK (deflated K)	capitalprod	lncapitalprod	
Wage Share	(W*L)/VA	wageshare		d_wageshare
Unit Labour Cost (ULC)	LC/RVA	ulc	lnulc	d_lnulc
Total Factor Productivity (TFP)	Levinsohn-Petrin TFP computed using VA	tfp		
<b>Olley-Pakes TFP Decomposition, with:</b>				
Value Added (VA)	OPD_RVA=AvTFP + Tot_SW_VA			
Employment (L)	OPD_L= AvTFP + Tot_SW_L			
DTFP	TFP- AvTFP			

The DO.file was aimed at delivering relevant descriptive statistics at a sector level for some performance indicators.

- **Descriptive Statistics:**
  - Mean
  - Standard deviation
  - 10, 25, 50, 75, 90 Percentiles
  - Skewness
- **Covariance with Size (ln L)**
  - Cov(ln L, TFP)
  - Cov(ln L, ln ULC)
  - Cov (ln L, Wage Share)
  - Cov(L, lnRVA\_L)

## 2.3 Country-Time Coverage

Country	Time Horizon	Threshold	Aver. Number of Firms
IT	1980-2012	[More than 50 and more than 20 employees]	
ES	2001-2010	More than 10 employees	2922
FR	1995-2010	No threshold	136525
HU	2003-2010	No threshold	3789
BE	1996-2011	Both with no threshold and with more than 20employees	All: 43728 Over 20: 3346
DE	1997-2010	Both with no threshold and with more than 20employees	All: 24652 Over 20: 9401
PT	2000-2009	[No threshold ]	
SK	2000-2011	More than 20 employees	1688
PL	2002-2011	More than 20 employees	9773
NL			
SI	1995-2011	Both with no threshold and with more than 20employees	All: 6719 Over 20: 1868
CZ	2002-2007	No threshold	2212
RO	2004-2011	Both with no threshold and with more than 20 employees	All: 43106 Over20: 8480
SE			
DK	Does not have the microdata readily available		
FI	Does not have the microdata		
AT			
EE	1995-2010	More than 20 employees	8067
<b>EFIGE DATA</b>	Countries: AT, FR, DE, HU, IT, ES, UK, for a small number of observations		

According to the confidentiality thresholds, two datasets have been created:

### 1) TFP\_DATASET\_FINAL

- Threshold: None
- Country-Coverage: FR, HU, BE, DE, PT, SI, CZ, RO, EE

### 2) TFP\_DATASET\_FINAL20 empl.

- Threshold: firms with more than 20 employees
- Country-Coverage: BE, DE, SK, PL, SI, RO, EE



# 3. Some initial results

**NB. This results are preliminary and tentative. They are aimed at showing the potential of the data and at collecting comments for the improvement of the DO.file.**

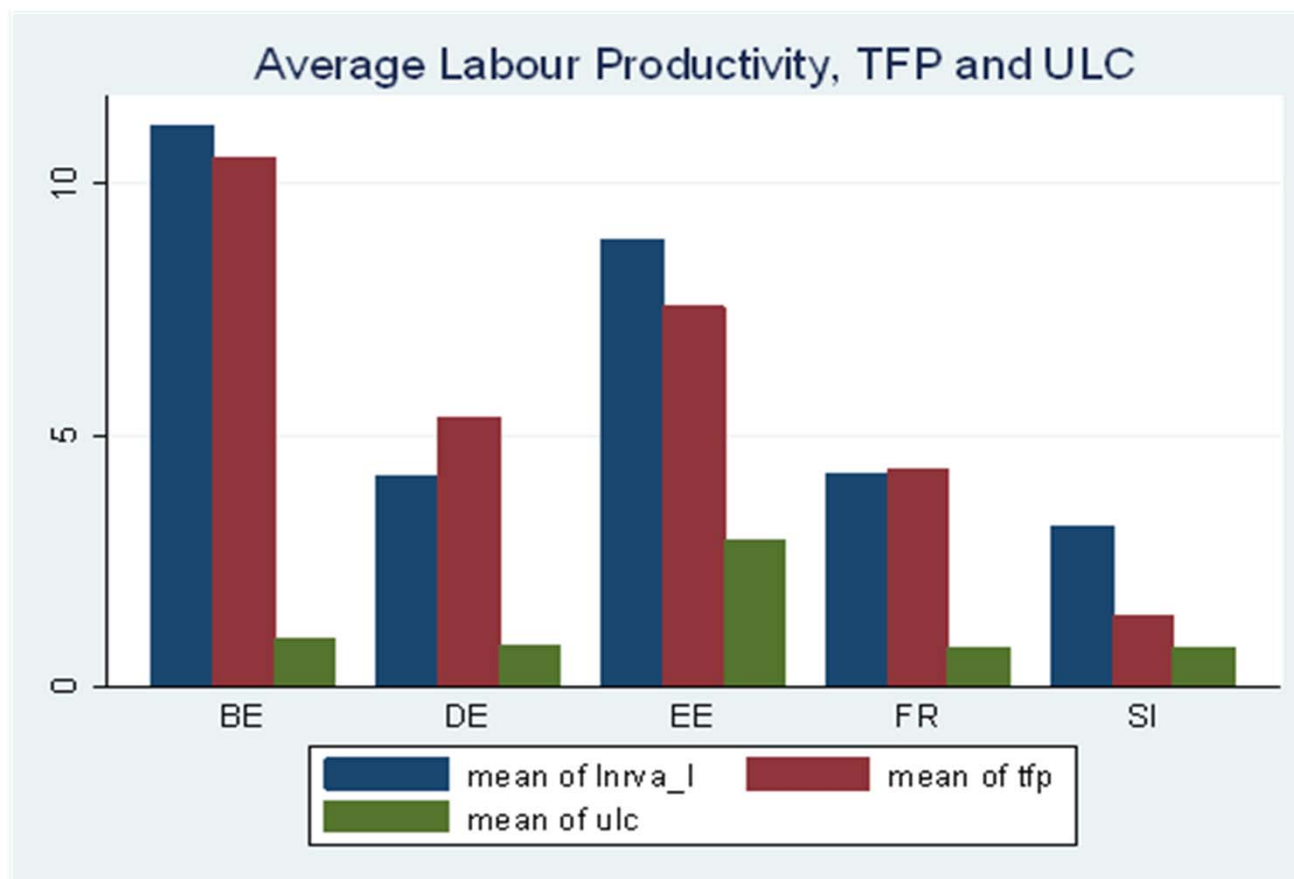
**The selection of the countries showed in the graphs is arbitrary and does not reflect the involvement of the respective country-teams in the exercise.**

		YEAR																
		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
BE	<i>Firms</i>		31656	34767	36366	37817	39186	40276	41365	42571	44102	45875	48168	50677	52678	54096	54418	45636
	<i>Av empl</i>		1754.747	1795.046	1718.823	1634.944	1606.47	1615.16	1453.352	1298.64	1267.487	1241.539	1233.461	1149.356	1136.834	1032.387	1014.376	1145.629
CZ	<i>Firms</i>								2227	2359	2237	2163	2156	2132				
	<i>Av empl</i>								1731.59	1632.146	1556.136	1540.435	1512.865	1509.67				
DE	<i>Firms</i>			23890	22669	23073	23073	23065	23332	23603	25434	27255	27744	27167	25883	25761	24934	21327
	<i>Av empl</i>			11128.62	9751.403	9001.875	9001.875	8287.909	10171.21	7634.098	5526.893	4977.079	4790.473	3908.991	3913.63	4025.925	3354.645	3550.335
EE	<i>Firms</i>	2116	2490	3293	3901	4363	4922	5385	5928	6480	7565	9435	11501	17905	12818	17976	12940	
	<i>Av empl</i>	466.7308	498.4824	594.7477	581.1689	427.6517	393.7314	407.0057	322.6516	333.9785	352.5923	372.4091	358.502	376.4156	354.9434	330.8015	250.2379	
FR	<i>Firms</i>	124175	121480	131040	129849	131353	134678	133507	138946	137168	141998	146809	149485	162203	137975	126454		
	<i>Av empl</i>	880.146	1287.302	1154.321	1110.901	1057.321	993.5917	1012.798	949.3477	894.4968	861.0513	803.3557	778.6444	735.4627	829.6802	813.8646		
HU	<i>Firms</i>									4785	4534	3970	3708	3369	3411	3234	3304	
	<i>Av empl</i>									1509.448	1669.126	1757.03	1721.592	1988.027	1838.282	2117.781	1688.936	
RO	<i>Firms</i>										33065	37249	41948	48294	53707	46877	43007	40706
	<i>Av empl</i>										1478.504	1283.228	1120.661	1011.517	927.28	897.9881	870.9338	915.522
SI	<i>Firms</i>	4371	4738	4736	5000	5102	5193	5369	5509	5890	6386	6882	7548	8562	9407	9609	9805	10117
	<i>Av empl</i>	1213.367	1165.138	1084.377	1034.626	992.4761	966.6862	972.5421	951.1552	924.9949	867.1204	823.0535	786.0673	768.9818	747.2864	676.4309	642.8824	584.2953

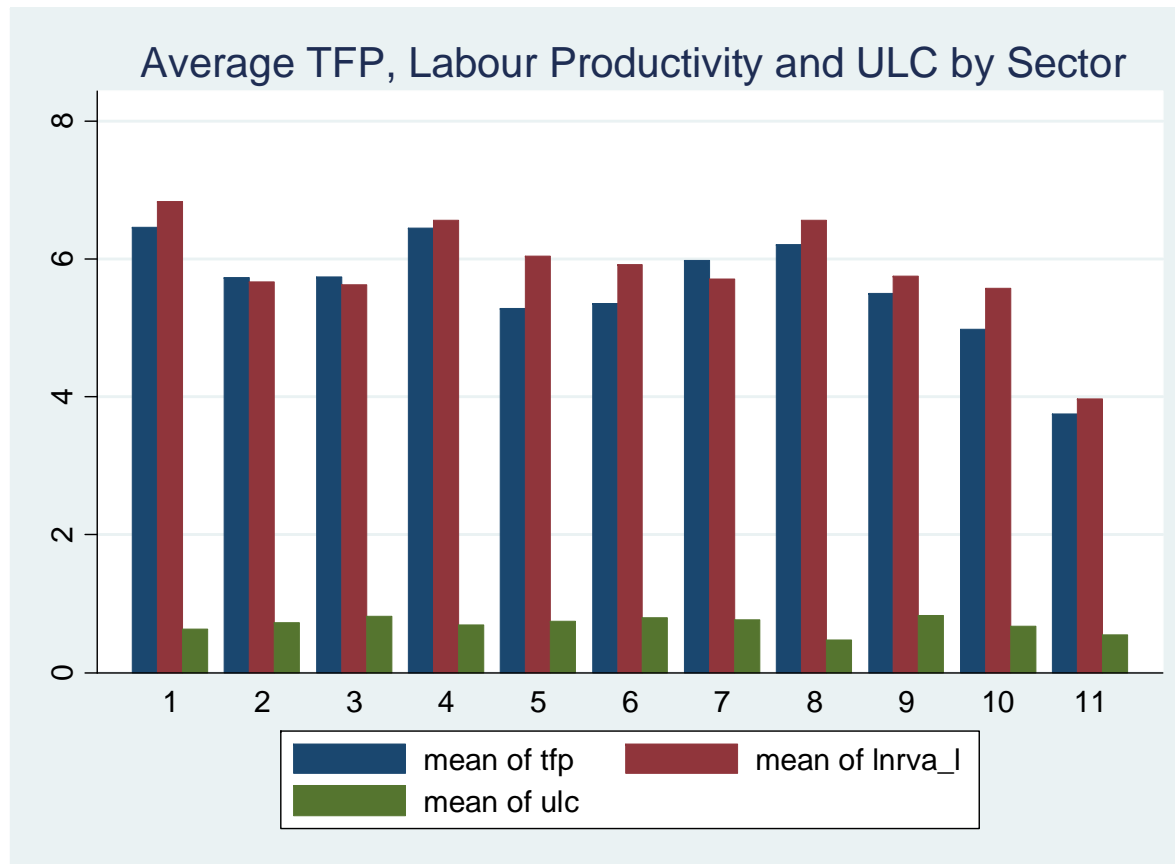
Note: computation with the unrestricted sample.

C O U N T R Y	SECTOR											
		Mining	Food	Textile	Chemicals	Basic Metals	Elect.Products	Auto	Electricity	Constructions	Telecommunications	Financial Services
BE	24	25	28	98	198	64	190	291	9	105		
DE	265	248	194	318	332	308	1013	172	81	252	172	
EE	6	33	20	45	42	65	36	27	7	45	5	
FR	19	40	29	80	115	79	98	100	15	237	2	
HU	44	67	65	83	208	172	306	193	79	901		
RO	94	28	43	48	118	49	289	190	13	27		
SI	54	33	38	65	113	69	137	71	11	32	51	

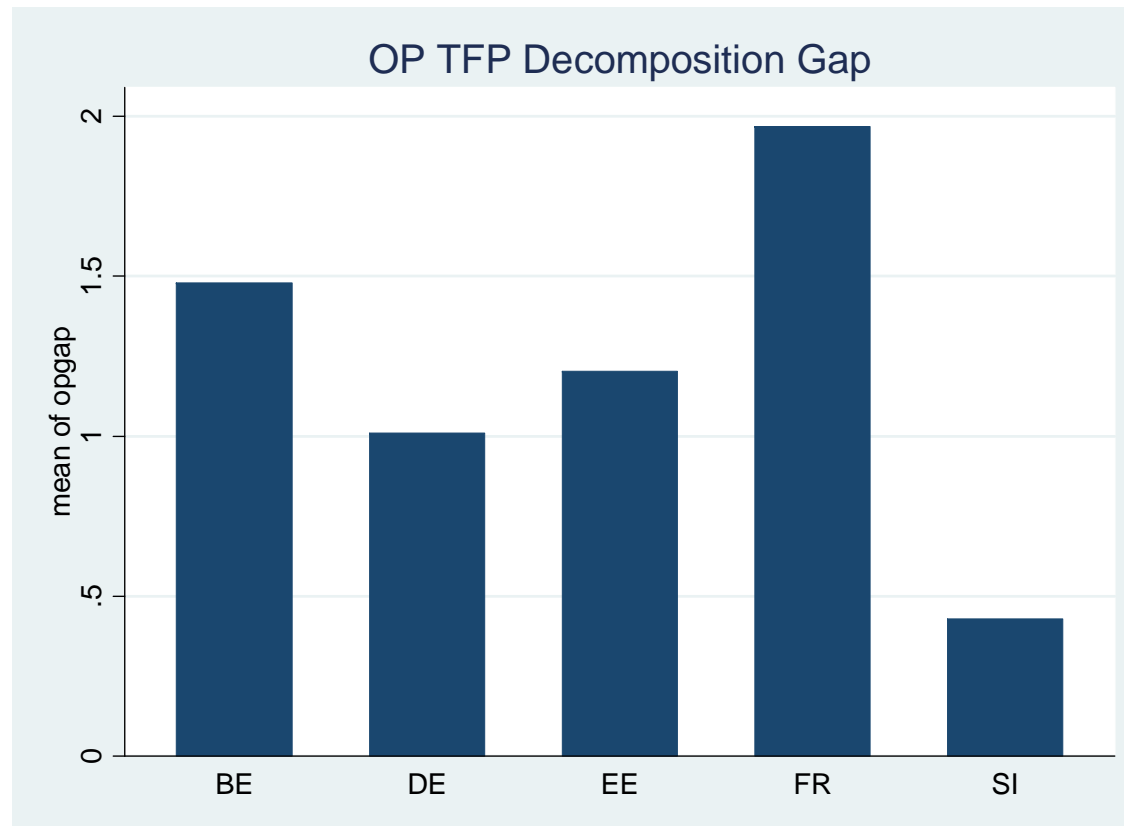
Note: computation with the unrestricted sample. Data refer to average number of employees per firm in 2009.



Note: computation with the unrestricted sample. Average of all sector for the years 2000-2009.

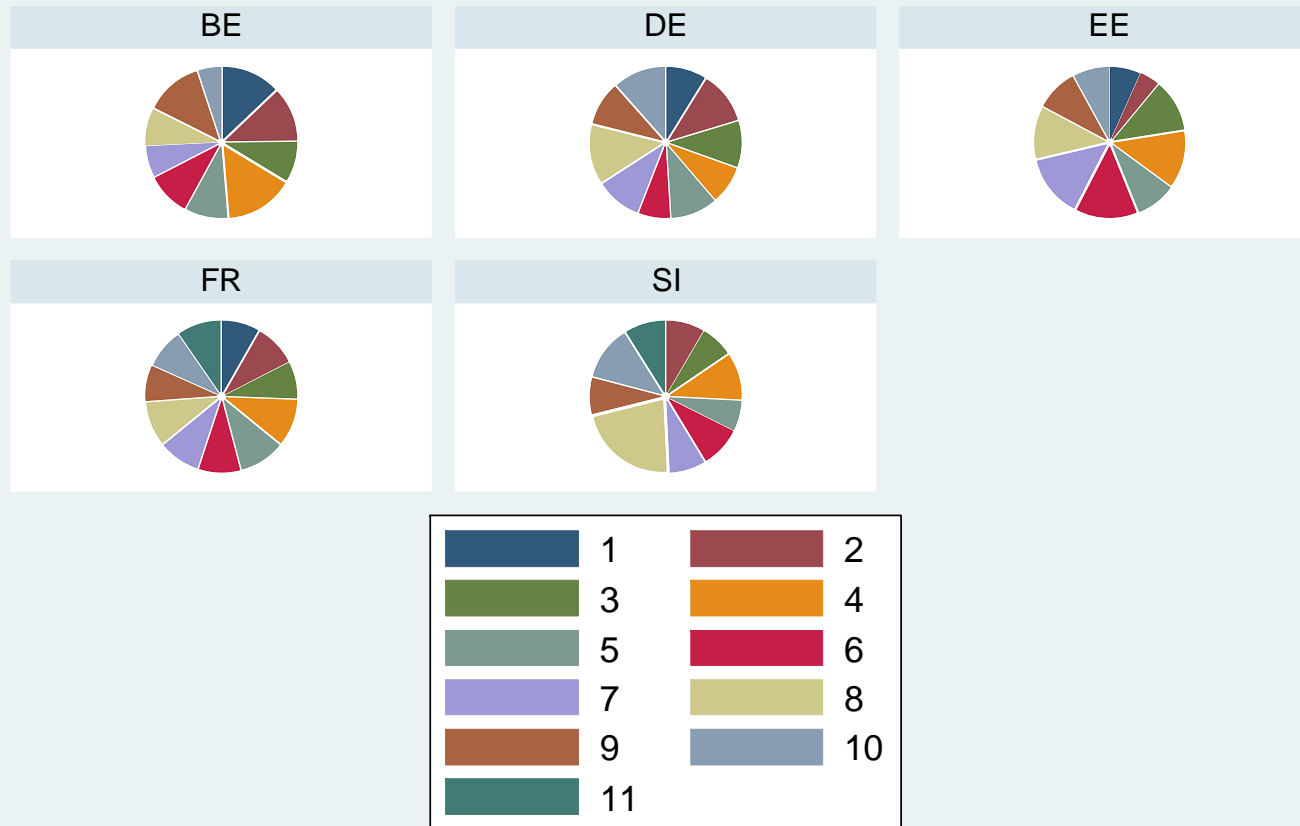


Note: computation with the unrestricted sample. Average of all countries for the years 2000-2009.



Note: computation with the unrestricted sample. Average of all sector for the years 2000-2009.

## TFP distribution over sector



Graphs by country

Note: computation with the unrestricted sample. Average of years 2000-2009.

# 4. The next steps



# 4.1 TFF Analysis DO.File-Second Round

- We have created an improved version of the existing DO.File, to be run before the end of the month.
- Main improvements:
  - 1) The performance indicators have been revised and corrected according to all the valuable comments received from WS2 CompNet Members;
  - 2) The indicators have been refined in order to guarantee synchronized methodologies with the NSO exercise (see later).
  - 3) The improved DO.file will be executed both with NACE rev.2 industry classification and the EUKLEMS classification. This would guarantee the possibility of comparison with other similar exercises (i.e. E. Bertelsman);
  - 4) Using the above mentioned industry classification will also improve the sector coverage of the exercise, insofar relatively limited
  - 5) The country coverage will be enlarged, too. Together with the 13 country-teams already involved, the analysis will be sent to the following additional countries: IE, CY, GR, LU, LT, MT, BG, LV, Turkey (in this regard, data availability has to be confirmed).

# 4.1 TFF Analysis DO.File-Second Round

- Focus on additional indicators and synchronized methodologies

## a) Inclusion of the distribution of capital productivity

## b) Olley-Pakes Decomposition:

### ➤ Labor Productivity

$$\text{Weight in OPD} \rightarrow \theta_{it} = l_{it} / \sum_i l_{it}$$

### ➤ Capital Productivity

$$\text{Weight in OPD} \rightarrow \theta_{it} = k_{it} / \sum_i k_{it}$$

### ➤ TFP

$$\text{Weight in OPD} \rightarrow 1) \theta_{it} = RVA_{it} / \sum_i RVA_{it}$$

$$2) xqm_{it} = 0.5 * ((m_{it} / RVA_{it})(m_{it-1} / RVA_{it-1}))$$

$$xqe_{it} = 0.5 * ((lc_{it} / RVA_{it})(lc_{it-1} / RVA_{it-1}))$$

$$imp_{it} = ((m_{it}^{xqm_{it}})(l_{it}^{xqe_{it}})(k_{it}^{(1-xqm_{it}-xqe_{it})}))$$

$$\theta_{it} = imp_{it} / \sum_i imp_{it}$$

## 4.2 Trade Data DO.File

- Based on the outcome of this first exercise, the second step should link the existing firm-level data with trade data within NCBs .
- The goal is to expand the existing micro-based datasets on competitiveness with indicators derived from the merged custom-firm-level data.
- In this respect, the exercise undertaken by the teams of FR, BE, HU represents an important starting point.
- In terms of data availability, in addition to FR, HU, BE, we would be able to add only PT, RO, SL (and possibly IT). Anyway, all the countries teams involved in the first exercise will be asked to run this second DO.file
- The relevant STATA codes of this second exercise will be prepared with the coordination of the French, Belgian and Hungarian teams and with the collaboration of C. Altomonte and E. Bartelsman.

# In Conclusion...

**Thanks to all of you for the great help  
and valuable collaboration!**

**[http://www.ecb.int/home/html/researcher\\_compnet.en.html](http://www.ecb.int/home/html/researcher_compnet.en.html)**

# **Appendix: Country-specific Notes**

Country	Notes On:		
	Sectors	Variable Definition	Other
ES	Sector 11 excluded	First best option variables	Data come from the Central Balance Sheet data sample (CBSD). In the sample, there is an over-representation of telecommunications and electricity sector and an under-representation of the construction and textile sectors. Moreover the CBSD is not representative of the Spanish population of firms; there is a clear bias to large firms.
FR	Sector 10 in 1995 excluded for confidentiality reason		
BE	Sector 11 excluded	<p>Capital stock: tangible fixed assets</p> <p>Number of employees : total employment in full time equivalent</p> <p>Costs of employees: total labor costs paid by the firm (wages, social security contributions, ...)</p> <p>Value added: added value as reported in the financial reports</p> <p>Materials: total inputs bought by the firm (coming from VAT declaration to include small firms in the sample)</p>	
PL	Sector 11 excluded	<p>Capital: total assets</p> <p>Intermediate goods: materials+energy</p> <p>NACE CONVERSION: NACE1-NACE2 conversion for the years 2002-2005</p>	
CZ	Sectors 1, 8, 10 and 11 excluded for TFP calculation		
RO	Sector 11 excluded	First best variables except for number of employees (second best)	

Country	Notes On:		
	Sectors	Variable Definition	Other
DE	Sector 11 excluded	<p>Total assets (capital): tangible assets (intangible assets excluded)</p> <p>Material costs: expenses for raw, auxiliary and process materials, for purchased goods + expenditure on external services</p> <p>Costs of employees: wages and salaries + social security and expenditure on pension schemes and other benefits</p> <p>Number of employees: mixed</p> <p>Wagebill: wages and salaries</p> <p>Gross value added: turnover+increase in finished and unfinished goods +other company-produced additions to plant and equipment+other operating income-material costs-other operating expenses-other taxes</p>	Since zeros may mean zero values (e.g. sales=0) or "not reported", they have been excluded from the analysis.
SK	Sector 11 excluded for TFP calculation	<p>All first best variables</p> <p>Material costs: Intermediate inputs, energy and other non-storable supplies</p> <p>NACE CONVERSION: Only in cases where there were just pre 2008 data (i.e. no information nace2), NACE1-NACE2 conversion has been used</p>	
EE			<p>Limitations for EE indicators:</p> <p>1) Mean-based RVA_L and ULC are very jumpy and probably unusable for EE. The number of firms within a cell is often small in EE; this means that one extraordinary transaction will affect means a lot. Medians are probably more useful for these indicators.</p> <p>3) Due to confidentiality, cases where there are 3 or less firms within a cell have been excluded.</p> <p>4) Data: Business Register data.</p>

Country	Notes On:		
	Sectors	Variable Definition	Other
SL		<p>Number of employees = average number of employees based on the number of work-hours in the period</p> <p>Total assests (k)= balance sheet value of total assets (including tangible and intengible asses, financial investments, operating receivables)</p> <p>Material costs (m)= intermediate inputs (ie. costs of merchandise, material, services, and other operating expenses, excluding labour costs and write-downs in value)</p> <p>Wagebill (w)= salaries and wages</p> <p>Costs of employees (l)= labour costs (including wagebill , retirement insurance costs, other social insurance costs, other labour costs)</p> <p>Added value (va)= gross operating returns minus intermediate inputs, where Intermediate inputs represent costs of merchandise, material, services, and other operating expenses, excluding labour costs and write-downs in value</p>	