FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Multinational Networks, Domestic, and Foreign Firms in Europe

Bruno Merlevede
Matthijs De Zwaan
Karolien Lenaerts
Victoria Purice

February 2015
2015/900

D/2015/7012/02

Multinational Networks, Domestic, and Foreign Firms in Europe

Bruno Merlevede*
Ghent University

Matthijs De Zwaan
Ghent University

Karolien Lenaerts
Ghent University

Victoria Purice
Ghent University

February 2015

Abstract

This paper introduces two datasets, *AUGAMA*, a panel of European firms for the period 1996-2011, and *EUMULNET*, a European Multinational Network data set. These datasets are constructed on the basis of the Amadeus database issued by Bureau Van Dijk Electronic Publishing. We document the process of building these data sets from the raw Amadeus data for 26 European countries. We show that the data sets adequately approximate the structure of the European economy across countries, regions, and industries as portrayed by data from Eurostat (Structural Business Statistics) and Cambridge Econometrics. As an illustration of possible application, we use the datasets to test a number of results from the theoretical literature regarding the productivity of multinational firms vis-a-vis domestic firms.

Keywords: multinationals, firm performance, total factor productivity, firm-level data

JEL classification: F23

*Corresponding author; Department of Economics, Ghent University; Bruno.Merlevede@UGent.be

1 Introduction

This paper in detail documents the build of a large pan-European firm-level data set, *AUGAMA* ('Augmented Amadeus'), with the aim to serve as a reference for future work. We extensively document our 'augmentation' that overcomes drawbacks -from an academic point of view- related to the way the data-provider, Bureau Van Dijk Electronic Publishing (BvDEP), issues the database. The advantage of our data is that it covers cross-country comparable firm-level data for 26 European countries in a single dataset. This allows for cross-country research at the firm level while maintaining representativeness that is comparable to other recent efforts (see for example CompNet (CompNet, 2014)). Our approach also improves the data with respect to exit and entry patterns. The paper further documents the build of a dataset of European multinational networks, *EUMULNET*. This dataset uses raw data from the Amadeus database by BVDEP as well. Both databases cover the following European countries: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Germany, Estonia, Spain, Finland, France, United Kingdom, Greece, Croatia, Hungary, Ireland, Italy, Latvia, Lithuania, the Netherlands, Norway, Poland, Portugal, Romania, Sweden, Slovenia, and Slovak Republic. We document representativeness across countries, industries, and regions. We then enrich the database with a measure of total factor productivity using the Wooldrige-Levinsohn-Petrin methodology (see Levinsohn and Petrin (2003), Wooldridge (2009), and Petrin and Levinsohn (2012)).

In the final part of the paper we present a set of empirical applications that makes use of the datasets to test a number of results from the theoretical literature regarding the productivity of multinational firms vis-a-vis domestic firms. Specifically, in line with (Markusen, 1995), we show that foreign firms are more productive than their domestic counterparts. We show that the TFP premium of foreign over domestic firms on average amounts to 48% in *AUGAMA*. Theoretical work further suggests that only the most productive domestic firms will engage in foreign direct investment (FDI) (see Helpman et al. (2004) for horizontal and Antràs and Helpman (2004) for vertical investment), because only the most productive firms are able to cover the costs associated with this investment. Using *EUMULNET* we show these predictions to hold for European multinationals. For a matched sample of foreign and domestic firms (following Alfaro and Chen (2012)) we find that foreign firms grow faster than domestic firms both pre and post crisis.

The paper is organised as follows. Section 2 documents the construction of the AUGAMA database. In Section 3 we consider overall representativeness of the data, while in section 4 the focus is on representativeness of the data in the industry and region dimension. In Section 5, we introduce the EUMULNET data. Section 6 discusses the estimation of total

factor productivity and in Section 7 uses estimated *TFP* in different applications comparing total factor productivity (growth) of domestic and foreign firms. Section 8 concludes this paper.

2 Database

2.1 The Amadeus database: basic data source

Raw data are taken from the Amadeus database provided by Bureau van Dijk Electronic Publishing (BvDEP). The Amadeus database is a pan-European database that comprises financial information on public and private companies. BvDEP gathers data from different local, i.e. country-specific, data providers and assembles them into a single database using a comparable format. Additionally, BvDEP assembles further information from firms' annual reports, media coverage, and other sources. It is not clear whether this is done for all firms and countries, but a bias towards large and listed firms seems likely. The Amadeus database is available in different flavours depending on the application of some thresholds for firms to be included. Our data originate from the *'full'* version where no thresholds are applied and all available firms are included in the database. The database contains both consolidated and unconsolidated accounts. We only consider firms that report unconsolidated accounts. This involves discarding less than 1% of firms (that only report consolidated accounts).

The Amadeus database provides the balance sheet and the profit and loss account of firms, information is not available at plant or establishment level. Financial information is aggregated up to a format that is comparable across European countries. Bearing in mind cross-country differences (in Europe) in terms of accounting formats, detailed items that potentially are available for specific countries are not included in Amadeus (e.g. the social balance sheet in Belgium). In addition to the financial information the database also provides us with information on the firm's main activity, its location, its date of incorporation, its ownership structure, and its affiliate structure (if the firm has any). The database further provides us with an unique firm ID that allows to link firms across different versions (*cf. infra*). A firm ID consists of 2-digit country initials followed by a number which is typically a VAT number or a registration number.

2.2 Augmenting Amadeus using multiple issues

BvDEP regularly updates the Amadeus database. Aside from the continuously updated online version, a physical DVD/BLURAY of the database is released monthly. We use the following issues of the DVDs to create our data set: 60, 72, 84, 96, 108, 119, 132, 144, 156,

168, 180, 192, 204 (September issue of years 1998–2010) and issues 210 and 220. By making use of a 'time series' of DVDs we overcome a number of issues that arise from the use of a single issue of the database. One may identify the following three limitations of a single issue of the database.

The first limitation is that a single issue only includes at most the last ten available years of financial information for an individual firm. We start from financial information that is available from the most recent issue of the database (i.e. issue 220). We then work our way back to earlier years using information from previous versions, starting with the second most recent version and so on.[1] For any given financial item and calendar year our rule is to prefer information from an issue as recent as possible. This procedure allows us to obtain a maximal time span of 18 years for an individual firm (i.e. years 1995-2012).[2]

Second, firms that go out of business are dropped fairly rapidly from the Amadeus database.[3] Our time series approach therefore allows for better tracking of exit and short-lived entry over the period covered. Firms report their date of incorporation which we use to determine entry and a firm's age. We define the year of exit as *'sample exit'*, i.e. the last year a firm reports basic financial information without showing up in later years in the database. We take potential changes in firm IDs into account by applying the ID changes listed on BvDEP's dedicated website to earlier versions of the database. Additionally, we check the help files of individual issues of Amadeus to control for systematic changes in firm IDs.[4] The updating of firm IDs from earlier versions avoids treating changes in firm IDs as exit.

Third, BvDEP updates individual ownership links between legal entities rather than the full ownership structure of firms. The ownership information in a specific issue of Amadeus therefore often consists of a number of ownership links. A single issue of the database only contains the most recent information on ownership links and therefore does not allow to track changes in ownership structure.[5] Our 'time series' approach remedies this limitation as it allows us to construct a time series of foreign ownership. We use ownership

[1]The most recent issue is version 220 in the current version of our data set. However, we did use version 228 to fill out missing financial information for the year 2012 for firms not yet reporting balance sheet and profit and loss account in the 220 issue.

[2]We have 18,732,383 observations available for 3,649,965 firms to estimate total factor productivity (TFP). The average time span is 6.1 years.

[3]In recent versions of the database, a larger set of firms exiting the market more than two or three years earlier seems to be available for some countries. This is not the case for earlier versions of the database.

[4]For example, in Belgium the firm ID is based on the VAT number. Recently, the administration added an additional zero in front of the existing 9-digit VAT numbers. For data retrieved from older issues we added the additional zero that was introduced in the official VAT number to the ID to have a comparable ID across different issues. For Romania we detected an even more drastic change from the Chamber of Commerce number to the VAT number as a basis for a firm's ID.

[5]More recent versions do contain some history on ownership links, but not all the way back to the late 1990s.

information to separate foreign firms from domestic firms (in Amadeus the variables are the following: *'shareholder ID', 'shareholder name', 'shareholder direct %', 'shareholder total %', and 'shareholder country'*). We focus on direct shareholder links to determine whether a firm is foreign-owned or not.[6] BvDEP updates individual ownership links rather than the entire ownership structure, therefore each ownership link in a given issue has a reference date which may differ even up to a couple of years. Because ownership link information is updated irregularly, there is not necessarily for each firm-owner-*year* combination information available. Further, the BvDEP ownership manual suggests that the date of an ownership link is not always updated when it is verified at a later point in time. We therefore assume that all reported ownership information is valid at the moment when the specific version of Amadeus is issued and assign all ownership information of a given version (that sums to 100% in the vast majority of cases) to the year of the issue.[7] In line with a common definition applied by e.g. the OECD and the IMF, we require that at least 10% of shares are owned by a single foreign investor for a firm to be considered foreign.[8] Foreign owners are owners with a known nationality that differs from the host country nationality. If the country of origin is not known (in some countries not all small firms report ownership) we consider these owners as domestic. We also keep track of the percentage of shares owned by foreign firms. This allows us to separate majority from minority foreign-owned firms or to apply a more stringent definition and only consider firms as foreign if more than 50% of shares is foreign-controlled.

2.3 Industry classification

The raw Amadeus data provide us with a primary 4-digit code in the European NACE classification of activities. NACE stands for *"Nomenclature Statistique des Activités économiques dans les Communautés Européennes"*. Our time series approach to the database implies that we have annual industry codes for firms. We deal with potential variation in industry codes

[6]For part of the firms an ultimate owner is also recorded, but this is often only the case for larger firms. Quite often ultimate owners are individuals. For example, Lakshmi Mittal or the Mittal family are sometimes recorded as ultimate owner of Mittal steel affiliates in Europe.

[7]In Merlevede et al. (2014) we experiment with a dataset at the firm-*owner*-year-level for Romania with the available information on ownership *links* from Amadeus. There, we fill out missing firm-owner-year-entries under restriction that the full ownership structure cannot exceed 100%. In case of time gaps between entries for the same owner-firm combination but with a different share-size we assume that changes show up immediately in the database. We then fill out the gaps with the older information. In the end, this more elaborated but very cumbersome procedure (the majority of owners have no ID number and need to be matched by names that tend to show lots of small variations across versions) does reveal only marginal differences with our current approach. We therefore apply the more straightforward procedure of assigning all ownership information (i.e. the ownership structure) to the year of the issue from which the information is retrieved.

[8]A firm of which only 5% of shares are owned by one or more foreign firms is considered a domestic firm. We observe 29,208 ownership changes from domestic to foreign or vice versa.

by creating different versions of our industry classification used in the data. If a firm reports the same code in the first and last year of information, we use that code throughout the entire period. For firms where only the fourth or the third (and fourth) digit are different between the first and last year's code we use the code of the last year. If a code is a clear outlier in the firm's time series, that code-year is ignored. For the remaining firms that show more 'bumpy' patterns (this is only a limited number of cases) we consider three alternatives. First we simply use the most recent code; second we use the most frequent code; and third we use the most recent code but allow for 'structural breaks'. A structural break is defined as a firm reporting two different codes with a single break and the less frequent code appearing in at least three years (versions) of the raw data. The first alternative serves as our basis, while the others are available for robustness tests. As indicated above, this issue comes up only in a limited number of cases.

Our firms are classified according to revision 1.1 of the NACE nomenclature. Revision 2 of the classification became the standard classification near the end of our sample period. Because most of the firm-level information and most of our other data (*cf. infra*) refer to NACE revision 1.1, we opted to convert the industry codes of young firms at the end of the sample for which we only observe a NACE revision 2 code to revision 1.1 codes (because they are only included in the later issues of the database). For older firms we have a revision 1.1 code from the earlier versions. To convert the codes we use a conversion table obtained from Eurostat. In the conversion table most old codes match in multiple new codes (or vice versa). We deal with this issue in two alternative ways. One way is deterministic in the sense that we start from the available one-to-one matches and exclude these from multiple matches where these codes are mentioned. Transforming this into an iterative procedure results in a large number of one-to-one matches. For the remaining many-to-many matches we obtain a single match by preferring manufacturing over agriculture and services. Our second approach randomly matches a revision 2 code with one of the possible revision 1.1 matches from the conversion table. Random matching is performed firm by firm, not industry by industry. The deterministic approach serves as our basis, while the randomisation is available as robustness check. Note that for the vast majority of firms we have an original NACE revision 1.1 code from the Amadeus database.

Although a 4-digit code is available, we mostly rely on 2-digit (or slightly more aggregated) industry classifications for practical implementation (*TFP* estimations for example). In Table 16 in the Appendix a list of the industries used is provided.

2.4 Location of firms: Region classification

In *AUGAMA* regions are defined using the EU's NUTS-classification ('Nomenclature of territorial units for statistics') which is a hierarchical system for dividing the economic territory of the EU. In Table 17 in the Appendix we list the main criteria and an example of the NUTS-structure. Firms are assigned to a region on the basis of their zip code, which is available in Amadeus. Eurostat provides tables mapping zip codes into NUTS regions for most countries. For other countries we rely on national data sources (the Eurostat website provides contact details). When a zip code is not available for a firm, we try to map the '*region*'-variable reported in Amadeus to the NUTS-classification. As with the industry classification above, our time series approach to the database implies that we have annual zip codes for firms. We deal with potential time variation by creating different versions of the regional classification. If a firm reports the same zip code in the first and last year of information, we use that code throughout the entire time period. When a code is clearly a one-time outlier in the firm's time series, that code-year is ignored and replaced. For the limited number of firms that change zip code, we consider three alternatives. First we simply use the most recent zip code to assign the firm to a region; second we use the most frequent zip code; and third we allow a 'structural break' for firms reporting two different zip codes with a single break that implies that the less frequent code appears in at least three years of the raw data. The first alternative serves as our basis, while the others are available for robustness tests. As with the industry classification above, the large majority of firms consistently reports the same zip code throughout the different versions.

We account for changes in the NUTS classification itself by means of conversion tables retrieved from the Eurostat website. In our data we use the 2006 vintage of the NUTS-classification. Where other data sources use a different version of the NUTS classification (e.g. the Cambridge Econometrics data described below), we reclassify the NUTS codes in the data. In most cases, codes change because of slight modifications of the area definition, such as border shifts. Since these shifts are small and unlikely to have important economic consequences, we rename the new code back to the old, and merge with our data. In other cases, regions are split or merged. Where a one-to-one correspondence between codes is straightforward to establish, we do so. Where it is not (as for example in the case where two old regions are split into more than two new regions), we distribute data values proportionally over those regions. In most cases, regions are redefined at the NUTS 3 level, and have no impact at the NUTS 2 distribution. Moreover, since such changes always involve bordering regions, regional differences are never very pronounced. Eurostat discusses changes in the NUTS classification in its publications on 'Regions in the European Union' (e.g. European Commission, 2011), and provides spreadsheets for changes between differences NUTS vintages

at its website.[9]

2.5 Deflation and currency

The data retrieved from Amadeus are downloaded in units of national currency. In order to make cross-country comparisons, these data are converted to Euro. Because our price deflators refer to national currency, we first deflate the data in national currency to obtain unit equivalents and then convert them to Euro using 2005 exchange rates.[10] By making use of the 2005 exchange rate we avoid that exchange rate movements would drive cross-country comparisons (see Gal, 2013). For Euro-zone countries earlier data in the old national currency are converted using the Euro-entry conversion rate. For countries adopting the Euro more recently (e.g. Slovenia) financial information dating before Euro adoption was converted to Euro by BvDEP using concurrent exchange rates. These data were converted by multiplying them with the ratio of the Euro-entry conversion rate and the concurrent exchange rate.

Our main data source for output deflators is the EU KLEMS database. These deflators have been incorporated and updated by Eurostat. We use EU KLEMS data up to the year 2005 and then continue with Eurostat data. For the last three years of our sample NACE revision (rev.) 1.1 price deflators are no longer available (nor are NACE rev. 2 price deflators for the earlier years). We therefore apply the percentage change of a corresponding NACE rev. 2 series to the later years of the NACE rev. 1.1 series (both revisions do report similar broad categories such as e.g. food processing). We define our capital deflator as the average of the following five NACE rev. 1.1 industries: machinery and equipment (29); office machinery and computing (30); electrical machinery and apparatus (31); motor vehicles, trailers, and semi-trailers (34); and other transport equipment (35) (see Javorcik, 2004). We calculate an intermediate input deflator as a weighted average of output deflators where country-time-industry-specific weights are based on intermediate input uses retrieved from input-output tables. We obtain most input-output tables from Eurostat.[11] For countries for which Eurostat does not yet provide input-output tables, we use input-output tables from the World Input-Output Database which are slightly more aggregated in terms of industries (Eurostat tables are at NACE 2-digit level) (Timmer, 2012).

Value added is double deflated, i.e. real value added is calculated as output deflated with an output deflator minus intermediate use deflated with the calculated intermediate input deflator.[12] Note that making use of industry-level deflators has some implications for our

[9] `http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/history_nuts`

[10] For comparison with Eurostat's Structural Business Statistics (SBS) database (*cf. infra*), data in national currency are converted to Euro with annual exchange rates rather than 2005 fixed exchange rates.

[11] For most countries we have IO-tables for 1995, 2000, and 2005.

[12] For Croatia we do not have detailed prices, nor IO-tables.

measure of total factor productivity (*cf. infra*).

2.6 Variable definitions

We define the following variables. Output Y is measured as *'operating revenue turnover'*, real output y is obtained by deflating Y by producer price indices of the appropriate NACE industry (*cf. supra*). Value added VA is defined as output minus intermediates M, i.e. operating revenue minus *'material costs'* (from the Amadeus database).[13] Real value added va is double deflated and defined as real output y minus real material costs m. The latter are defined as material costs deflated by the intermediate input deflator defined above. Labour L is the *'number of employees'* (end-of-period). Capital K is measured by *'tangible fixed assets'*, real capital k is obtained by deflating K by the capital deflator defined above. The age of a firm is calculated on the basis of its *'date of incorporation'*. We have information on the number of months accounts refer to. We use this information to convert flow variables (operating revenues, material costs, and thus value added) to twelve month equivalents as far as the number of months is not below 6 or above 24. Outside these boundaries variables are set to missing. End-of-period variables such as tangible fixed assets and the number of employees are unchanged. The number of non-12 months accounts is very small and generally below 0.5% for a country-industry-year cell. We define labour productivity as operating revenues divided by the number of employees and estimate a measure of total factor productivity (TFP). We prefer total factor productivity to labour productivity as the latter does not control for intermediate inputs usage and capital intensity differences across firms (Gal, 2013).

The strength of $AUGAMA$ is that it provides information for firms from many countries and industries for more than 15 years including the Great Recession and Euro crisis period. These are important features as the CompNet Task Force notes that *"firm-level analysis in Europe is hampered by a lack of sufficient and comparable data across countries"* (CompNet, 2014). There are however some caveats one needs to bear in mind when interpreting results obtained using $AUGAMA$ in empirical analysis. Regarding the measurement of real output, we have no data on output quantities but can only observe output expressed in terms of revenue. As indicated above, output is deflated with industry-level price deflators as we do not have firm-level price deflators. This implies that we are only able to consider $TFPR$ and not $TFPQ$ (total factor productivity in terms of revenue rather than quantity). According to Syverson (2011) this approach is satisfactory when differences in prices only reflect differences

[13] Amadeus does contain value added figures for some countries that are either obtained directly from the data-provider or are calculated using an accounting definition, but gauging from the manual, the definition differs across countries.

in product quality. When differences in prices also reflect differences in market power, measured efficiency/productivity of firms will also reflect market power (Syverson, 2011). On the input side we have information on the total number of employees, but not on the total number of hours worked nor on other employee characteristics, such as skill levels.[14] Estimated productivity levels should therefore be thought of as including labour quality and capacity utilisation (Gal, 2013). The Amadeus database provides data on the total stock of *'tangible fixed assets'*, but more detail is not available. Changes in (capital) capacity utilisation can thus not be accounted for. Note that these issues are not specific to our data set, but are faced by many micro-level data and studies (e.g. CompNet, 2014, or Gal, 2013).

2.7 Basic data cleaning

First, negative values of the number of employees, tangible fixed assets, operating revenue, sales, material costs, and value added (defined as the difference between turnover and material costs) are set to missing. We then calculate growth rates of the aforementioned variables and replace observations associated with growth rates below (above) the 1st (99th) percentile with missing values (cf. CompNet (2014)). Further cleaning, e.g. after *TFP* estimation, is done in function of specific applications of the data for specific research questions.

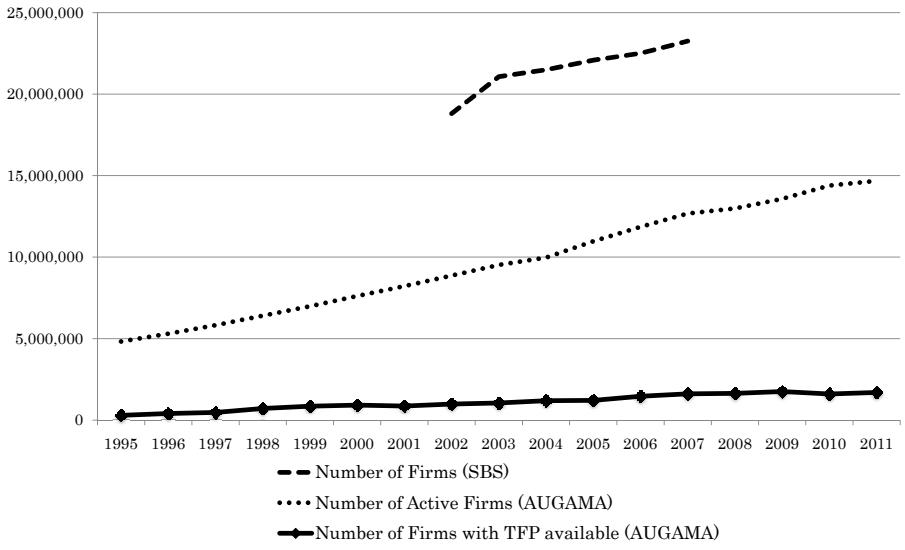
3 Representativeness

In this Section we discuss *AUGAMA* in terms of coverage and representativeness. Our main comparison base is the Structural Business Statistics database (SBS) provided by Eurostat.[15] SBS data in NACE revision 1.1 are available for the period 2002–2007 (2003–2007 for some countries). We use this period to infer the representativeness of our data set. We consider firms in the 'business economy' (mining, manufacturing, construction, and services, excluding financial services), i.e. NACE 2-digit codes 10 to 74, excluding 65 to 67. Table 16 in the Appendix lists all the 2-digit industries included in *AUGAMA*. In order to get a first broad

[14] Total wage costs, *'costs of employees'*, are reported in Amadeus, and in principle could be used as a quality adjusted labour input. The variable is filled out less frequently, however, and more importantly it is also prone to cross-country differences in the regulatory framework (e.g. social security contributions). Therefore it is likely to be a good reflection of actual labour costs, but cross-country comparisons in terms of 'quality' are not recommendable. Further, it is not always clear from the Amadeus manual whether the definition of *'costs of employees'* is similar across countries (e.g. whether management compensation is included or not). For this reason, we restrict ourselves to the number of employees as labour input.

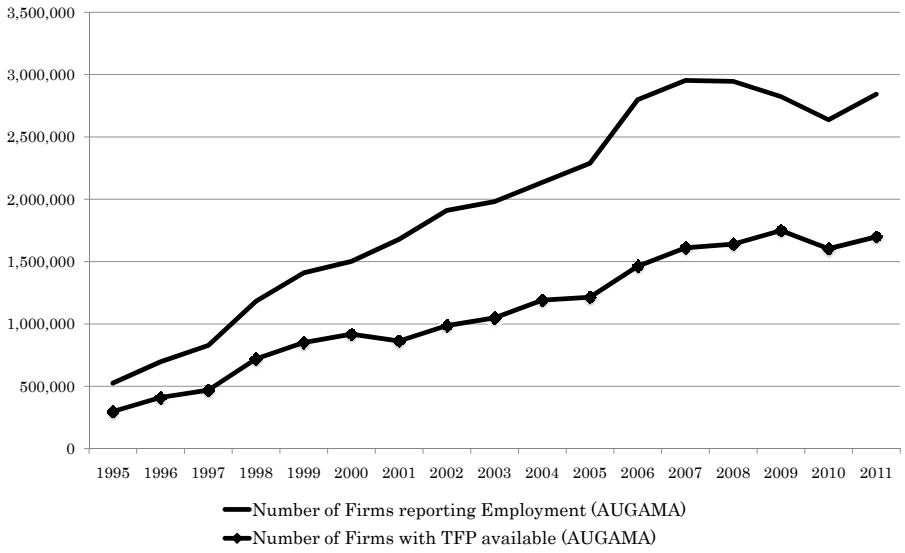
[15] We think of the SBS data as providing the population of firms. Do note that SBS data are retrieved from surveys with incomplete coverage of the population of firms for some countries, which might result in ratios above 100% (CompNet, 2014). Furthermore, it is not always clear whether SBS data consider only companies or a larger set of business entities which also includes sole proprietors.

Figure 1: Number of firms in Europe according to Eurostat SBS (Structural Business Statistics) and *AUGAMA* (1995–2011).



Notes: *AUGAMA*: all active firms and all firms with *WLP-TFP* available are displayed. Countries included are AT, BE, BG, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LT, LV, NL, NO, PL, PT, RO, SE, SI, and SK (HR is not available in SBS).

Figure 2: Number of firms reporting employment and number of firms with *WLP-TFP* in *AUGAMA* (1995–2011).



Countries included are AT, BE, BG, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LT, LV, NL, NO, PL, PT, RO, SE, SI, and SK (HR is not available in SBS).

overview Figure 1 plots the number of firms over time.[16] The SBS data count more than 20 million firms in Europe (some countries do not report numbers in 2002, hence the *'jump'* in 2003). We see a gradual increase in the number of active firms recorded in Amadeus from about five to fifteen million firms from 1995 tot 2011. In 2007 this accounts for 55% of SBS firms, up from 45% in 2003. These numbers are smaller when we only count firms that report the variables necessary to obtain a measure of total factor productivity based on the Wooldridge-Levinsohn-Petrin estimator, *WLP-TFP* (see Wooldridge, 2009 and Petrin and Levinsohn, 2012; $cf.\ infra$). From Figure 2 one can infer that the number of firms for which we are able to estimate TFP also steadily increases to more than 1.5 million observations from 2007 onwards (from 5% (2003) to 6.9% (2007) of the number of SBS firms). The number of firms that at least report employment is considerably higher (12.7% of SBS firms in 2007).

Table 1 shows further numbers illustrating representativeness in columns two to five. The entries in Table 1 are based on *AUGAMA* corrected for outliers following the procedure described above. The percentages shown are averaged over industry-time cells by country. A country-industry-time cell in this case is defined by the host country, a broad NACE category (SBS does not report finer detail, see Table 16), and the year of observation. Based on the information from *AUGAMA*, a firm is assigned to a cell. After assigning firms to cells we calculate cell aggregates (total number of firms, employees, total turnover, and total labour costs) and create the ratio with the corresponding aggregate from the SBS data. The numbers in Table 1 are obtained by averaging over industry-time cells by country. Note that calculations are based on all firms that report the indicated variable and that 'coverage' in this respect may differ not only between countries, but also within countries across variables. Table 1 reveals that coverage in terms of the number of firms ranges from a low of 5.2% for the Netherlands to a high of 86.9% in Estonia. On average over countries we observe a quarter of SBS firms. The coverage in terms of total employment, labour costs, and turnover is higher and indicates that *AUGAMA* typically includes larger firms. Averaged over countries, *AUGAMA* accounts for about 60% of employment and turnover and 53% of wage costs. The last four columns of Table 1 reveal that our data set is slightly biased towards manufacturing firms in comparison to what is reported in SBS statistics (count of firms), but the discrepancy falls within very reasonable margins. Table 2 shows the distribution of firm size according to both SBS data and *AUGAMA*. *AUGAMA* is generally biased towards larger firms (especially firms with between 20 and 249 employees). In most countries this bias increases when we focus on firms for which *WLP-TFP* is available, but not to a large extent.[17]

[16] Countries included are: AT, BE, BG, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LT, LV, NL, NO, PL, PT, RO, SE, SI, and SK. HR is not reported in SBS.

[17] For AT, BE, DE, and LV we record larger changes. In Belgium, for example, this is due to the fact that most smaller firms file an abridged account and are not obliged to report material costs.

Table 1: *AUGAMA* versus Eurostat Structural Business Statistics (SBS): Representativeness.

	SBS 2002–2007	AUGAMA as share of SBS 2002–2007				SBS 2007		AUGAMA 2007	
	# firms	# firms	# empl.	turnover	costs of empl.	share of firms in manufacturing	share of firms in services	share of firms in manufacturing	share of firms in services
AT	269,426	11.0%	46.5%	40.8%	34.6%	10.0%	90.0%	22.2%	77.8%
BE*	333,564	42.3%	85.0%	81.3%	87.7%	9.1%	90.9%	14.2%	85.8%
BG	221,116	17.6%	60.3%	83.5%	66.4%	11.8%	88.2%	15.2%	84.8%
CZ	871,067	11.4%	74.7%	76.0%	72.5%	16.9%	83.1%	19.7%	80.3%
DE	1,714,904	14.6%	32.4%	42.0%	36.1%	11.0%	89.0%	21.0%	79.0%
DK	198,369	18.4%	41.9%	36.7%	45.8%	8.7%	91.3%	14.2%	85.8%
EE	38,270	86.9%	98.6%	97.7%	62.1%	12.2%	87.8%	16.2%	83.8%
ES	2,499,620	36.9%	71.3%	75.1%	72.1%	7.9%	92.1%	17.2%	82.8%
FI	186,972	28.2%	49.1%	47.1%	39.9%	12.0%	88.0%	17.3%	82.7%
FR	2,158,887	23.6%	62.1%	63.9%	62.9%	10.2%	89.8%	14.0%	86.0%
GB	1,571,916	10.0%	80.0%	65.8%	67.1%	8.9%	91.1%	22.0%	78.0%
GR	694,183	12.7%	52.3%	50.4%		11.2%	88.8%	27.2%	72.8%
HU*	551,119	8.5%	35.2%	38.7%	35.7%	10.4%	89.6%	19.2%	80.8%
IE	87,175	12.6%	26.0%	31.2%	8.6%	4.7%	95.3%	18.8%	81.3%
IT	3,790,324	15.0%	55.8%	58.8%	57.8%	13.1%	86.9%	30.7%	69.3%
LT	88,187	22.8%	52.5%	61.9%		11.6%	88.4%	20.4%	79.6%
LV	60,581	18.7%	54.9%	57.1%	7.0%	11.1%	88.9%	16.5%	83.5%
NL**	497,613	5.2%	54.5%	42.9%	39.0%	8.6%	91.4%	18.4%	81.6%
NO	198,926	38.6%	72.5%	65.8%	72.5%	12.0%	88.0%	11.9%	88.1%
PL	1,452,512	6.5%	46.8%	50.3%	27.9%	13.3%	86.7%	30.9%	69.1%
PT	711,778	34.0%	30.9%	34.9%	33.6%	10.9%	89.1%	16.0%	84.0%
RO	389,286	67.6%	87.7%	36.1%	34.6%	12.6%	87.4%	15.8%	84.2%
SE	514,925	32.2%	64.5%	78.6%	79.2%	10.9%	89.1%	13.6%	86.4%
SI	91,065	24.2%	80.8%	80.2%	79.6%	17.8%	82.2%	30.6%	69.4%
SK	42,525	40.7%	78.4%	89.3%	89.2%	14.3%	85.7%	21.4%	78.6%

For BE and HU the period is 2003–2007; ** For NL some outlier firms in levels were removed after manual inspection. Firms are assigned to a country-year-industry cell, for which aggregates are calculated. The ratio of each *AUGAMA* cell with the corresponding SBS cell is determined. Table entries are then obtained by averaging over industry-year. Only firms that report the indicated variable are considered in the analysis. This results in differences in coverage across countries , but also across variables within countries.

Table 2: *AUGAMA* versus Eurostat Structural Business Statistics (SBS): Firm Size Distribution.

	SBS (2002–2007 average)				AUGAMA (firms with employment) (2002–2007 average)				AUGAMA (firms with TFP) (2002–2007 average)			
	1–19	20–49	50–249	250+	1–19	20–49	50–249	250+	1–19	20–49	50–249	250+
AT	94.3%	3.7%	1.7%	0.3%	52.1%	21.6%	20.4%	5.9%	17.8%	14.4%	42.8%	25.1%
BE	96.3%	2.5%	0.9%	0.2%	87.6%	8.1%	3.5%	0.7%	36.3%	30.2%	27.5%	6.0%
BG	94.9%	3.1%	1.7%	0.3%	75.3%	12.3%	10.0%	2.4%	60.4%	19.2%	16.6%	3.9%
CZ	97.5%	1.4%	0.8%	0.2%	69.1%	14.6%	13.0%	3.3%	65.8%	15.8%	14.7%	3.8%
DE	92.6%	4.4%	2.3%	0.5%	69.3%	15.6%	12.2%	3.0%	40.7%	19.5%	28.9%	10.9%
DK	93.6%	4.2%	1.9%	0.3%	80.2%	11.8%	6.6%	1.4%				
EE	91.0%	5.7%	2.9%	0.4%	86.1%	8.8%	4.5%	0.6%	85.8%	9.2%	4.4%	0.6%
ES	96.7%	2.3%	0.8%	0.1%	86.6%	9.5%	3.4%	0.5%	84.7%	10.8%	3.9%	0.6%
FI	96.2%	2.3%	1.2%	0.3%	88.9%	7.2%	3.2%	0.7%	88.3%	7.7%	3.3%	0.6%
FR	96.2%	2.5%	1.0%	0.2%	84.7%	9.7%	4.6%	1.1%	83.4%	10.4%	4.9%	1.2%
GB	94.6%	3.3%	1.7%	0.4%	40.3%	18.3%	31.0%	10.5%				
GR	98.5%	0.8%	0.3%	0.1%	68.1%	21.7%	9.1%	1.2%				
HU	96.0%	1.5%	0.7%	0.1%	74.1%	13.1%	10.0%	2.9%	73.4%	13.4%	10.2%	2.9%
IE	91.2%	5.1%	2.7%	0.5%	42.8%	22.4%	28.6%	6.2%				
IT	98.2%	1.3%	0.5%	0.1%	75.6%	15.2%	8.1%	1.1%	74.3%	16.0%	8.5%	1.1%
LT	93.2%	4.6%	2.6%	0.3%	52.1%	24.8%	19.8%	3.2%				
LV	91.8%	5.2%	2.7%	0.4%	64.2%	19.2%	14.4%	2.2%	47.1%	16.9%	27.7%	8.4%
NL	94.7%	3.4%	1.6%	0.3%	43.6%	19.1%	30.2%	7.0%	55.8%	16.6%	21.5%	6.1%
NO	99.0%	2.7%	1.2%	0.2%	87.6%	8.4%	3.3%	0.6%	85.8%	9.7%	3.8%	0.7%
PL	97.6%	1.3%	0.9%	0.2%	35.9%	21.7%	33.3%	9.2%	36.2%	22.9%	32.7%	8.2%
PT	97.1%	1.8%	0.8%	0.1%	90.7%	6.0%	2.8%	0.5%	89.2%	7.0%	3.3%	0.5%
RO	93.6%	3.7%	2.3%	0.5%	88.9%	6.4%	3.9%	0.8%	88.1%	6.8%	4.1%	0.9%
SE	97.2%	1.8%	0.8%	0.2%	92.1%	5.2%	2.2%	0.5%	91.4%	6.1%	2.2%	0.3%
SI	96.2%	2.1%	1.3%	0.3%	78.0%	10.8%	8.5%	2.7%	76.5%	11.6%	9.1%	2.8%
SK	89.0%	4.9%	4.7%	1.1%	66.2%	11.4%	17.3%	5.2%	63.1%	12.0%	19.1%	5.8%

Comparison of the share of firms that employ 1 to 19 employees, 20 to 49 employees, 50 to 249 employees or at least 250 employees in SBS and AUGAMA. The share of firms that fall within a specific size category is based on the average number of firms in period 2002–2007 for the SBS and AUGAMA data. In the case of AUGAMA, only firms for which the number of employees or *WLP-TFP* are available are considered.

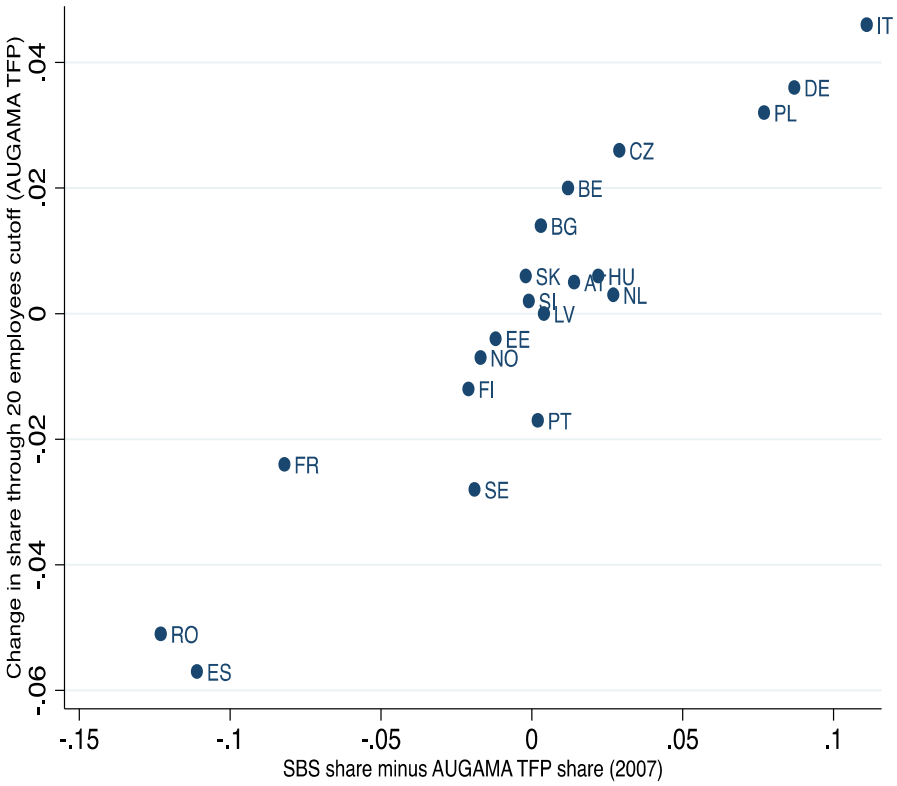
The dispersion across countries in representativeness in terms of the number of firms found, and to a lesser extent in terms of the number of employees, turnover and wage costs (see Table 1) is mainly due to differences in the coverage of small firms. When we focus on a sample of firms with on average more than 20 employees (as in e.g. Gal, 2013 and CompNet, 2014), the sample becomes more balanced in cross-country terms. Figure 3 plots the difference between a country's share in the number of firms according to SBS data and its share in the number of firms according to *AUGAMA* in 2007, against the resulting change in the share in the number of firms according to *AUGAMA* when a cut-off of at least 20 employees (on average by firm) is applied to the data. There is a clear relation between a country's share falling short of the SBS share (more to the right on the horizontal axis) and an increase in its share when applying the cut-off (more to the top on the vertical axis). This brings the cross-country distribution closer to the SBS data. Figure 4 shows the number of firms with more than 20 employees for SBS and the number of firms with more than 20 employees and *WLP-TFP* available for *AUGAMA*.[18] By 2007 our TFP-sample accounts for a quarter of the number of firms with more than 20 employees. Tables 18 and 19 in the Appendix list the annual number of observations in the TFP sample for all firms and foreign firms separately.

4 Representativeness of economic activity across industries and regions

In this Section we focus on representativeness of *AUGAMA* in terms of the distribution of economic activity across industries and across regions (within individual countries and within 'Europe'). Representativeness across industries is presented as the correlation of total industry activity recorded in *AUGAMA* with the total activity reported in SBS statistics. Industries are defined as 21 'broad' industries grouping sometimes several NACE 2-digit industries (see Table 16 in Appendix) as this is the level of aggregation reported in the SBS statistics. The period considered is generally 2002–2007, which is determined by the availability of SBS statistics in NACE revision 1.1. As indicators of activity we consider output, employment, wage costs, and material costs. Table 3 contains the pairwise correlations for all countries in the sample. The correlations are always positive and statistically significant. The majority of correlations is well above 0.75. This especially holds for output and employment. For wage costs and material costs correlations are generally somewhat lower, but still statistically

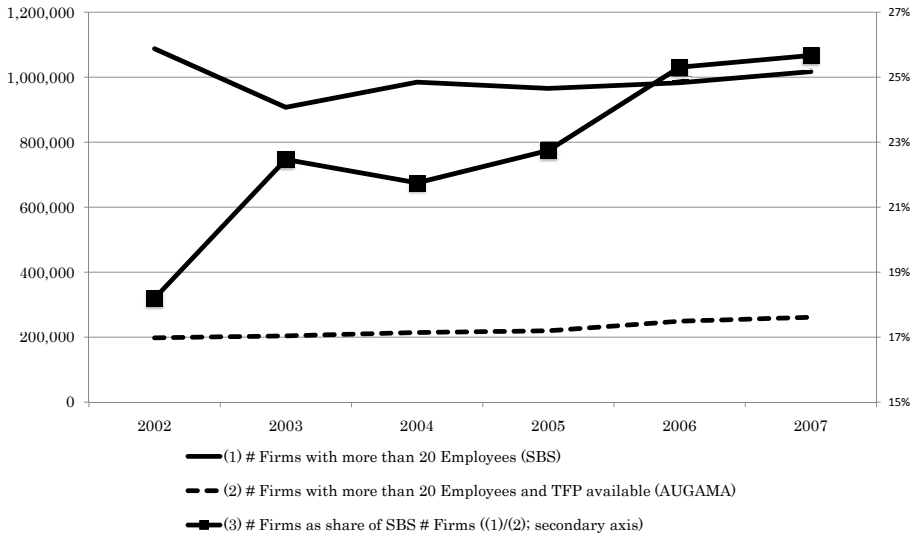
[18] DK, GB and IE are not included in the figure because one of the variables for TFP-calculation is missing for all firms. GR and LT are included but only have a very small sample of TFP firms, either limited in the time dimension (GR) or limited in coverage (LT).

Figure 3: The impact of a 20 employee cut-off on the distribution of firms across countries: Eurostat Structural Business Statistics (SBS) and *AUGAMA* in 2007.



Difference between a country's share in number of firms in SBS compared to *AUGAMA* plotted against the change in a country' share in *AUGAMA* when a 20 employee cut-off is applied.

Figure 4: Number of firms when applying a 20 employee cut-off in Eurostat Structural Business Statistics (SBS) and *AUGAMA* (2002–2007).



AUGAMA: only firms with *WLP-TFP* available are considered. Ratio of number of firms in *AUGAMA* and SBS displayed on secondary axis on the right.

significant. Table 3 clearly indicates that the distribution across industries found in *AUGAMA* is very well aligned with the distribution derived from SBS statistics.

We compare the regional distribution of economic activity found in *AUGAMA* with the regional distribution we derive from the Cambridge Econometrics Regional Database.[19] We consider both within-country and Europe-wide regional representativeness of economic activity. The Europe-wide regional distribution might be affected by differences in coverage across countries (see Table 1 above) while the within-country regional distribution is not. For country-by-country or country-specific analysis within-country distribution comparisons are relevant, whereas for Europe-wide regional analysis the Europe-wide distributions' comparison should be considered. Economic activity is measured by the total number of employees and the total amount of generated value added in a specific region. Because before we detected a tendency towards better representativeness of larger firms in countries with a lower overall coverage in terms of the total number of firms, we may expect the distribution of economic activity to be less affected as larger firms will account for the bulk of economic activity.

We consider two samples of industries: i) the business economy, i.e. NACE 2-digit codes 10 to 74 (see Appendix), and ii) 'broad' manufacturing, i.e. NACE 2-digit codes 10 to 42. In case of the former Cambridge Econometrics data also cover financial services (NACE 2-digit

[19]This database has been used in academic research by Becker et al. (2010, 2012), among others.

Table 3: Pairwise correlation coefficients of economic indicator totals for broad industries as calculated from *AUGAMA* and recorded in Eurostat SBS.

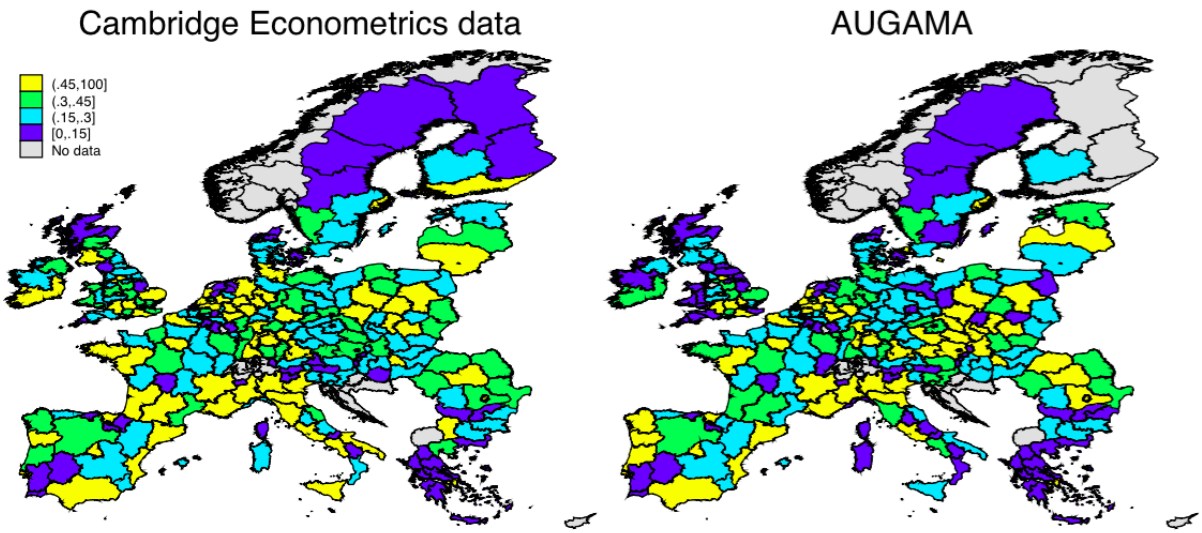
	Output	Employment	Wage costs	Material costs
AT	0.76	0.87	0.65	0.64
BE	0.97	0.97	0.79	0.98
BG	0.94	0.87	0.82	0.88
CZ	0.97	0.98	0.82	0.95
DE	0.92	0.91	0.77	0.93
DK	0.92	0.94	0.81	-
EE	0.96	0.91	0.64	0.82
ES	0.98	0.99	0.71	0.99
FI	0.77	0.88	0.74	0.55
FR	0.95	0.96	0.84	0.91
GB	0.97	0.94	0.69	-
GR	0.80	0.84	-	-
HR	-	-	-	-
HU	0.75	0.81	0.74	0.72
IE	0.81	0.90	0.26(a)	-
IT	0.86	0.96	0.82	0.67
LT	0.81	0.83	-	-
LV	0.97	0.92	0.41	0.54
NL	0.80	0.71	0.50	0.83
NO	0.94	0.98	0.63	0.93
PL	0.90	0.95	0.68	0.83
PT	0.75	0.51	0.50	0.65
RO	0.98	0.99	0.88	0.89
SE	0.97	0.98	0.85	0.65
SI	0.96	0.98	0.71	0.96
SK	0.95	0.89	0.88	0.90

Period 2002-2007, except for BE and HU 2003-2007. All correlations are statistically significant at 5% level except (a) where 0.26 is not significant (if we exclude the year 2002, we obtain a statistically significant correlation of 0.47).

codes 65 to 67) which we do not consider in our data (*cf. supra*). As far as the financial services sector's regional distribution is more or less in line with the distribution of other activity this should not have sizeable implications. Nevertheless, one should bear this in mind interpreting comparisons below. Therefore we also consider a more narrow definition labelled 'broad' manufacturing[20] which is the most detailed level provided in the Cambridge

[20]'Broad' manufacturing includes Mining (10-14) and Utilities (41-42) in addition to 'narrow' pure manu-

Figure 5: NUTS 2-digit regions' share in total European employment in 2005 according to the Cambridge Econometrics database and *AUGAMA*.



Countries included: AT, BE, BG, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LT, LV, NL, PL, PT, RO, SE, SI, and SK.

Econometrics database.

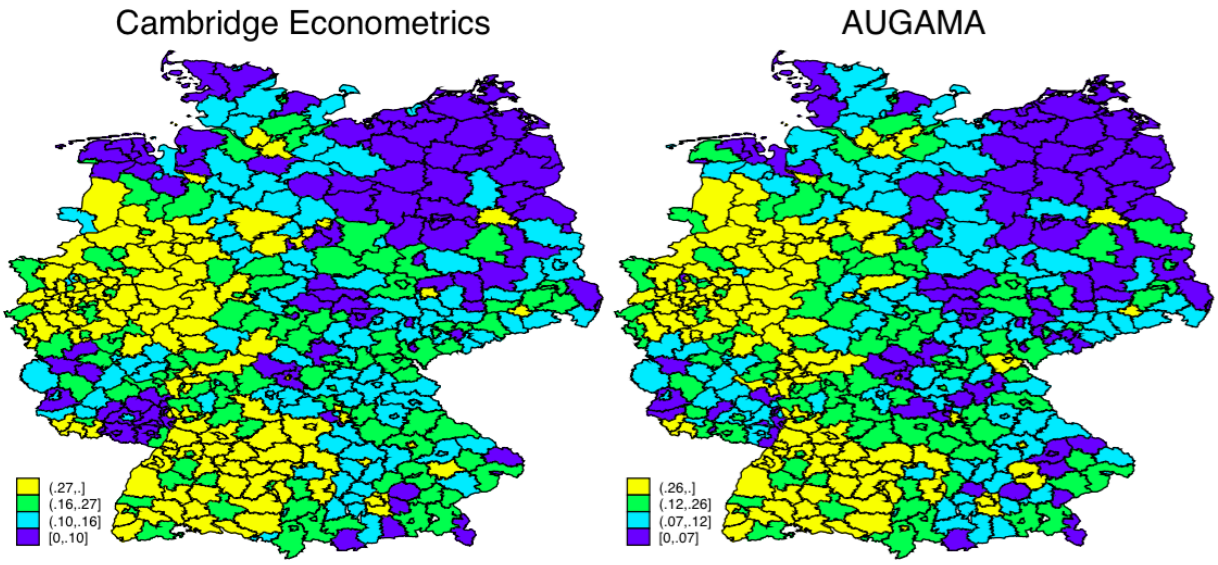
As a first indication we plot a map for the year 2005 in Figure 5 representing NUTS 2-digit regions' share in total *Europe-wide* employment according to the Cambridge Econometrics database (left panel) and *AUGAMA* (right panel). In Germany, the Netherlands, and the UK the shares of the NUTS 2-digit regions seem smaller for *AUGAMA* than for the Cambridge Econometrics data. In Figure 6 we plot similar maps for German NUTS 3-digit regions' share in total *German* broad manufacturing employment. The comparison of both panels suggests that *AUGAMA* very reasonably approximates the distribution of German economic activity for this sample.

In Tables 4 and 5 we quantify the information visualised in the maps by calculating the amount of economic activity (employment) that needs to switch region to align the distribution of regional economic activity obtained from *AUGAMA* with the distribution according to the Cambridge Econometrics data. In Table 4 numbers are obtained as the sum of absolute values of a region's share in total European employment according to the Cambridge Econometrics data minus its share according to *AUGAMA* divided by two.[21] Over time the number is quite stable. For the sample of *TFP* firms about 25% of employment

facturing (15-37).

[21]There are a few regions where *AUGAMA* records zero (no) activity. Deleting or retaining these regions from/in both databases prior to calculation does not affect conclusions.

Figure 6: NUTS 3-digit regions' share in total German employment in 2005 according to the Cambridge Econometrics database and *AUGAMA*.



needs to switch regions across Europe for the distributions obtained from the Cambridge Econometrics data and *AUGAMA* to be equal. This suggests that to a large extent the distribution of activity in *AUGAMA* follows the distribution of overall economic activity. For the 'broad manufacturing' industries numbers are generally smaller than for the business economy sample. Table 5 suggests that the bulk of this number refers to cross-country movement of employment. Table 5 lists the amount of activity that needs to switch regions *within* a country to match both distributions for the year 2005. For most countries this number is much smaller than the average in Table 4. Only for Bulgaria, the Netherlands, and Portugal the number is higher; only in Bulgaria this is the case for all samples considered.

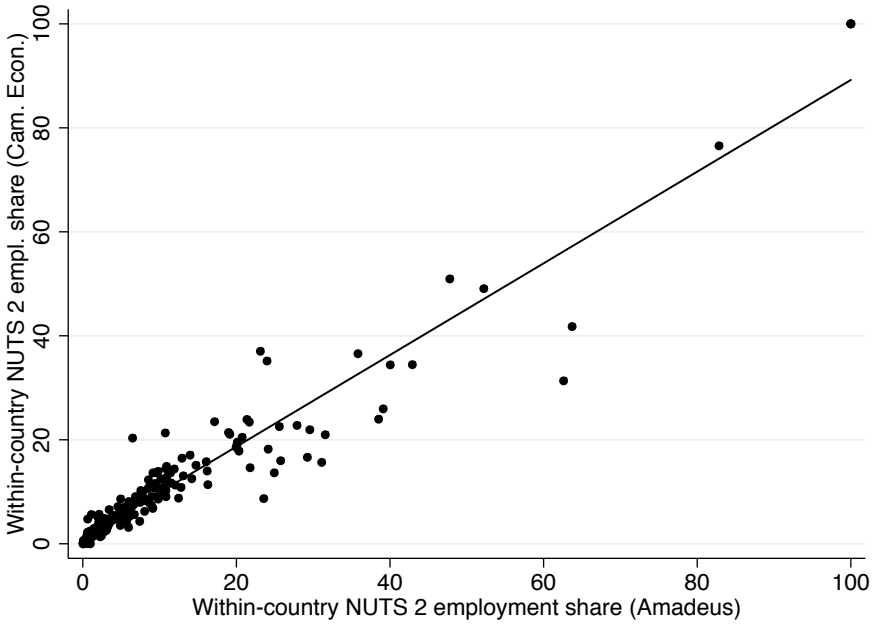
Figure 7 is a scatter plot of within-country shares in employment of NUTS 2-digit regions according to the *AUGAMA* database (horizontal axis) and the Cambridge Econometrics database (vertical axis) for the year 2005. The Figure suggests a high correlation between both measures. This is confirmed in Table 6 where pairwise correlations (also for 2005) are listed for NUTS 3-digit regions' share in within-country economic activity (measured either as employment or value added). Using all regions in all countries (i.e. row headed with 'Europe') the correlation varies between 0.71 and 0.87 according to the sample and activity measure. The correlation is always statistically significant. If we calculate the correlation for the 24 countries as a whole (in the NUTS classification these can be thought of as *'0-digit'* regions) we obtain a correlation coefficient of 0.96 for the sample used in the first column in Table 6 .

Table 4: Annual share of European employment that needs to switch region for the *AUGAMA* regional distribution to equal the Cambridge Econometrics distribution.

	All available		*TFP Sample*	
	Business economy	Broad manufacturing	Business economy	Broad manufacturing
2003	24.5	22.1	28.3	28.1
2004	23.6	21.5	28.6	27.4
2005	22.9	20.7	27.2	25.4
2006	20.4	18.6	24.1	22.5
2007	19.6	17.2	25.0	23.1
2008	20.5	17.5	25.4	23.3
2009	23.2	19.6	26.3	24.0
2010	21.6	18.6	27.9	25.7

Numbers are obtained as the sum of absolute differences between Cambridge Econometrics and *AUGAMA* region shares divided by two.

Figure 7: Within country share in employment of European NUTS2 regions according to the *AUGAMA* database and the Cambridge Econometrics database in 2005.



Countries included: AT, BE, BG, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LT, LV, NL, PL, PT, RO, SE, SI, and SK.

Table 5: Share of employment that needs to switch region within-country for the *AUGAMA* regional distribution to equal the Cambridge Econometrics distribution.

	All available		*TFP Sample*	
	Business economy	Broad manufacturing	Business economy	Broad manufacturing
AT	14.5	10.3	20.5	17.5
BE	12.0	9.8	15.9	12.2
BG	34.6	27.0	36.4	29.7
CZ	12.6	7.0	12.4	7.1
DE	20.1	19.9	23.2	26.2
EE	15.1	14.9	14.9	17.1
ES	11.9	8.0	10.8	8.0
FI	1.7	1.6	1.5	1.6
FR	17.5	23.3	13.6	23.0
HU	9.1	12.1	11.2	11.8
IT	20.2	17.0	20.0	17.1
LV	13.7	9.2	17.6	17.5
NL	6.7	15.6	26.9	30.8
PL	15.0	11.2	19.5	15.9
PT	31.4	24.5	25.6	24.5
RO	18.7	15.8	18.0	15.7
SE	13.2	8.2	7.0	8.7
SI	5.5	6.5	5.8	6.5
SK	14.4	9.8	14.9	10.4

Numbers are obtained as the sum of absolute differences between Cambridge Econometrics and *AUGAMA* region shares divided by two.

In the other rows of Table 6 we show the correlations over NUTS3 regions by country. For most countries these correlations are high and statistically significant, also for the countries with a larger number of regions.[22] Only in the cases of Bulgaria, Lithuania, and the Slovak Republic, the regional distribution of *AUGAMA* is less well aligned with the one from the Cambridge Econometrics database. For Bulgaria and the Slovak Republic this seems especially driven by the construction and services sectors as the correlation becomes positive and statistically significant if we only consider broad manufacturing. Though positive and significant, the correlation remains smaller than that for most other countries. Overall, we conclude that *AUGAMA* captures the regional distribution of economic activity fairly well. Bearing in mind some smaller caveats highlighted by the above numbers, the use of *AUGAMA* for regional firm-level analysis seems warranted.

5 *EUMULNET* - A European Multinational Network data set

In addition to our 'regular' panel of European firms, *AUGAMA*, we have also created a separate data set on European multinational networks based on Amadeus (*EUMULNET* henceforth). *EUMULNET* contains parent-affiliate combinations for which we have information both on the side of the parent and its affiliate from Amadeus. For every firm, Amadeus contains information on whether or not the firm has any affiliates. For firms with affiliates, Amadeus provides a list of affiliates with a limited amount of further information. The basis for *EUMULNET* is that affiliates available as separate entries in Amadeus are identified by their unique ID number. For these affiliates we are able to retrieve the full information set from their own entry in Amadeus rather than being limited to the information provided through the parent's entry in the database. *EUMULNET* is then the dataset of those parent-affiliate combinations for which both firms are listed as separate entries in Amadeus with full information. For affiliates not listed in Amadeus, we do have information on their existence and country of operation from the parent's entry in the database. There are two more potentially useful variables in the parent's entry in the database: operating revenue turnover of and employment at the affiliate.[23] However, this information from the parent's entry is not always filled out. Therefore this information is not very useful for further analysis

[22] The conclusions from Table 6 are unchanged when we recalculate correlations after excluding -by country- the most concentrated region that often looks like an 'outlier' compared to the other regions (typically the most concentrated region is also the home of the capital).

[23] These variables became available only in later versions of the database. Furthermore, with respect to the timing of the information it is also unclear what the calendar year is, since the variable refers to the *latest available* year.

Table 6: Pairwise correlations of within-country shares of NUTS3 regions in economic activity according to *AUGAMA* and Cambridge Econometrics

	Bus. Econ.		Broad manuf.		Bus. Econ.		Broad manuf.		Bus. Econ.		Broad manuf.	
	L	#reg	L	#reg	L	#reg	L	#reg	VA	#reg	VA	#reg
Europe	0.803*	1239	0.753*	1255	0.872*	975	0.832*	1049	0.723*	975	0.709*	1049
AT	0.985*	35	0.974*	35	0.964*	24	0.941*	34	0.986*	35	0.857*	35
BE	0.965*	44	0.895*	44	0.942*	39	0.860*	44	0.972*	44	0.874*	44
BG	0.154	28	0.496*	28	0.200	17	0.448*	28	0.346	28	0.533*	28
CZ	0.923*	14	0.737*	14	0.923*	14	0.744*	14	0.970*	14	0.864*	14
DE	0.871*	429	0.730*	429	0.858*	392	0.788*	425	0.758*	429	0.601*	429
DK	0.936*	11	0.935*	11	-		-		-		-	
EE	0.977*	5	0.945*	5	0.965*	5	0.884*	5	0.993*	5	0.984*	5
ES	0.973*	52	0.991*	52	0.979*	52	0.990*	52	0.968*	52	0.984*	52
FI	0.983*	6	0.985*	6	0.987*	6	0.983*	6	0.979*	6	0.972*	6
FR	0.907*	96	0.623*	96	0.903*	96	0.628*	96	0.913*	96	0.771*	96
GB	0.664*	120	0.402*	122	-		-		-		-	
GR	0.991*	46	0.993*	51	-		-		-		-	
HU	0.979*	20	0.902*	20	0.966*	20	0.871*	20	0.989*	20	0.908*	20
IE	0.993*	8	0.926*	8	-		-		-		-	
IT	0.954*	107	0.947*	107	0.955*	107	0.947*	107	0.956*	107	0.943*	107
LT	0.589	8	0.596	8	-		-		-		-	
LV	0.995*	6	0.991*	6	0.948	3	0.662	6	0.827*	6	-0.167	6
NL	0.983*	40	0.841*	40	0.886*	38	0.747*	39	0.858*	40	0.706*	40
PL	0.908*	66	0.790*	66	0.894*	66	0.751*	66	0.966*	66	0.776*	66
PT	0.950*	16	0.830*	25	0.967*	14	0.837*	25	0.982*	28	0.894*	28
RO	0.917*	41	0.868*	41	0.924*	41	0.871*	41	0.939*	41	0.863*	41
SE	0.968*	21	0.983*	21	0.993*	21	0.961*	21	0.989*	21	0.946*	21
SI	0.997*	11	0.996*	11	0.996*	11	0.995*	11	0.994*	11	0.986*	11
SK	-0.568*	8	0.757*	8	-0.543	8	0.752*	8	-0.156	8	-0.022	8

L stands for employment, VA for value added, #reg refers to the number of regions. The 'Business Economy' contains all NACE 2-digit sectors from 10 to 74, with the exception of financial services, 'broad manufacturing' contains all all NACE 2-digit sectors from 10 to 42. 'All unconsolidated' refers to the sample of firms that report the number of employees in unconsolidated accounts. 'TFP sample' refers to all firm-year observations for which we are able to obtain a *WLP-TFP* measure.

compared to the full entry information. We do retain a variable indicating whether the parent has extra-Europe affiliates or not and focus on parent-affiliate combinations where both firms are listed in Amadeus with full information. This also implies that the resulting data set is limited to the the European network of the parent should it also own non-European affiliates.

To create *EUMULNET* we use the following procedure. First, we extract parent-affiliate ID number combinations (plus the actual share owned by the parent in the affiliate) from every issue of the database.[24] This creates a time series of parent-affiliate links. We then restrict our attention to those combinations where the parent owns at least 50% of the affiliate at some point in time. We then replace '*direct*' parents that are found to be affiliates themselves with the '*ultimate*' parent as detected in Amadeus. In the resutling '*ultimate*' parent-affiliate-year data set we then fill out the *AUGAMA* information both on the parent and affiliate side. For earlier/later years when the link does not exist, we do fill out information for parent and/or affiliate from *AUGAMA* when available.[25] Our final data set forms a traditional panel data set in the affiliates-year dimension with full information on the parent side attached to each affiliate-year entry (as such duplicating parent-year information when the parent has multiple affiliates).

Table 7 lists the annual number of links of more than 50% between a parent and its affiliate that both have an Bureau Van Dijk ID number. From column (1) one can infer that the number of links we retrieve considerably increases over time, which is influenced by increased coverage over time. For about 4.1% of these links we are able to obtain a *TFP* measure (*cf. infra*) for both parent and affiliate. When we consider the evolution over time of this subset of links in column (2), we still observe an increase in links, but from 2002 onwards, and even more so from 2004 onwards, the number of links is fairly stable. From 2002 onwards between 10 and 17% of these links is between a parent and affiliate in a different European countries (see column (3)).

Table 8 focuses on the cross-country distribution for the year 2007. The first column lists the number of parents with a given nationality in the data set. This number is affected by cross-country differences in coverage, but only to a limited extent since we do not require any financial information to be provided by a firm to obtain these numbers. Most parents are found in the UK and the Netherlands, followed by Germany and France. The second column shows the number of affiliates owned by these parents (irrespective of the host country). Across countries, parents on average own between 1.4 and 2.9 affiliates. When we restrict the

[24] We limit ourselves to European ID numbers. For some affiliates there is a non-European ID number that refers to other Bureau Van Dijk products. This however applies to a very small number of firms. We also do not consider affiliates in Russia and Ukraine at this point.

[25] Occasionally a link is not reported in the year t issue of the database, while it is in the $t-1$ and $t+1$ issues, we then assume the link to exist in t as well.

Table 7: Total number of parent-affiliate links over time.

		parent-affiliate links with...	
	All links	... double TFP	... of which abroad
	(1)	(2)	(3)
1997	24,385	1,221	150
1998	49,795	3,385	455
1999	93,637	7,664	1,021
2000	144,031	13,314	1,444
2001	272,807	19,306	2,064
2002	469,312	23,674	2,983
2003	542,621	24,806	3,378
2004	726,771	29,280	3,899
2005	808,268	32,038	4,168
2006	835,149	34,895	4,719
2007	898,022	34,847	5,127
2008	1,016,984	34,099	5,537
2009	1,139,099	38,969	6,485
2010	1,125,608	38,275	6,504
2011	1,197,820	46,191	6,604

Only parent-affiliate links where the parent owns at least 50% of the affiliate at some point in time are considered. Column (1) shows all links that fulfil these requirements. Column (2) presents the number of these links for which *TFP* is available and column (3) shows links for which *TFP* is available with an affiliate abroad.

data to those parent-affiliate combinations for which *WLP-TFP* is available on both sides of the link we retain 34,847 observations in 2007, about 15% of these affiliates is located abroad.[26] For Denmark, the UK, Ireland and Latvia we are not able to compute a measure of total factor productivity because a necessary variable is missing for all firms.[27] For Greek firms we were able to obtain total factor productivity, but only for a limited number of firms for a limited number of years (late 1990s, early 2000s; not 2007). The last three columns of Table 8 focus on the number of affiliates located in the country indicated by the row heading. The correlation with the number of affiliates owned by parent firms from the country is fairly large (abstracting from the requirement of *WLP-TFP* availability), indicating that a lot of these affiliates are typically domestically-owned. The overall share of foreign-owned affiliates is 15% like the share of affiliates owned abroad before.

Finally, Table 9 considers the distribution of affiliates per parent for the sample without *WLP-TFP* restrictions (column (1) in Table 8) for the year 2007. The general conclusion from the Table is that a small number of parents owns a disproportionally large share of affiliates. Columns (1) and (2) reveal that 61% of parents owns a single affiliate, while another 19% owns two affiliates. In total 94% of parents owns five or less affiliates. Columns (3) and (4) reveal that parents owning affiliates in a foreign country are exceptional: 91.4% of parents does not engage in cross-border investment. 5.3% of parents owns a single foreign affiliate, 3.3% owns two or more foreign affiliates. In columns (5) to (8) we consider the number of affiliates rather than the number of parents. The 65% of parents with a single affiliate account for 26% of the total number of affiliates. 35% of affiliates is owned by parents that own more than five affiliates. The distribution of foreign affiliates looks fairly similar. Foreign affiliates typically belong to multi-affiliate parents. More than 70% of foreign affiliates are owned by parents that have at least two affiliates, 38.4% of foreign affiliates are owned by parents that have six or more affiliates.

6 Total Factor Productivity

6.1 Estimation framework

This Section is devoted to the estimation of total factor productivity (*TFP*). As input choices of firms are likely to be based on their productivity, the estimation of total factor productivity

[26] These numbers should be interpreted with care as they are partly driven by differences in both pure coverage and data quality (i.e. reporting variables necessary to obtain *WLP-TFP*) across countries. For example, for the Netherlands and Germany we only retain 0.4% and 3.5% of reported links because of poor reporting of financial information. This also accounts for the fact that a large share of the affiliates is located abroad, i.e. in a country with better reporting of financial information.

[27] UK firms do report value added in Amadeus but not material costs.

Table 8: Cross-country breakdown of parents and affiliates for the year 2007.

	# parents	# affiliates owned	# parent–affiliate links		# affiliates in country	# parent–affiliate in country	
			with double TFP...	... of which abroad		with double TFP...	... of which foreign-owned
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AT	5,783	14,063	319	187	14,261	272	140
BE	16,429	30,271	1,344	362	27,373	1,378	396
BG	3,931	12,803	866	1	13,011	903	38
CZ	463	648	164	27	1,363	318	181
DE	38,438	110,832	3,863	1,351	111,305	2,857	345
DK	25,168	44,179	-	-	41,185	-	-
EE	715	1,134	110	3	1,635	241	134
ES	20,945	47,127	7,659	465	49,564	7,998	804
FI	3,584	8,764	1,985	302	8,552	1,756	73
FR	35,244	87,867	5,776	739	86,972	5,915	878
GB	82,929	227,790	-	-	232,883	-	-
GR	1,113	2,195	-	-	2,408	-	-
HR	442	818	365	10	1,073	427	72
HU	506	789	206	36	1,655	354	184
IE	4,834	11,376	-	-	11,791	-	-
IT	8,848	24,335	5,859	840	22,406	5,498	479
LT	247	364	-	-	643	-	-
LV	146	225	3	1	563	7	5
NL	86,083	171,895	703	510	165,703	256	63
NO	17,333	36,105	1,713	111	37,347	1,740	138
PL	1,803	3,586	228	11	6,471	613	396
PT	2,862	6,161	953	53	6,962	1,152	252
RO	2,733	4,043	1,238	3	6,470	1,620	385
SE	25,745	50,453	1,446	84	46,162	1,478	116
SI	73	134	39	27	77	23	11
SK	33	65	8	4	187	41	37

Column (1) show the number of parents in each country, with columns (2)–(4) providing information about their affiliates and the parent-affiliate links (when *TFP* is available, affiliate located abroad or not). Columns (5)–(7) hold information on the number of affiliates in each country and the parent-affiliate links (when *TFP* is available, domestically-owned or foreign-owned).

Table 9: Distribution of the number of affiliates per parent for the year 2007 (without *TFP* restrictions).

	# parents with				total # affiliates owned			
	X affiliates		X foreign affiliates		all		foreign	
X	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	234,360	60.6%	20,573	5.32%	234,360	26.1%	20,573	26.2%
2	72,427	18.7%	5,252	1.36%	144,854	16.1%	10,504	13.4%
3	30,261	7.8%	2,456	0.64%	90,783	10.1%	7,368	9.4%
4	15,712	4.1%	1,385	0.36%	62,848	7.0%	5,540	7.0%
5	9,164	2.4%	891	0.23%	45,820	5.1%	4,455	5.7%
6	5,778	1.5%	591	0.15%	34,668	3.9%	3,546	4.5%
7	3,859	1.0%	409	0.11%	27,013	3.0%	2,863	3.6%
8	2,726	0.7%	319	0.08%	21,808	2.4%	2,552	3.2%
9	1,984	0.5%	239	0.06%	17,856	2.0%	2,151	2.7%
10	1,507	0.4%	157	0.04%	15,070	1.7%	1,570	2.0%
>10	8,652	2.2%	906	0.23%	202,942	22.6%	17,532	22.3%
0	-		353,252	91.41%	-		-	
Total	386,430		386,430		898,022		78,654	

Columns (1)–(4) show the number of parents who own a certain number of affiliates (domestic or abroad), columns (5)-(8) show the number of affiliates owned (by domestic or foreign parents, also considering the number of (other) affiliates this parent owns).

will be biased if the endogeneity of inputs is not addressed (Griliches and Mairesse, 1995). A number of alternative estimation procedures have been suggested in order to tackle this issue. The most popular alternatives are the semi-parametric approaches developed by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP), in which a proxy is introduced to handle the endogeneity bias. Olley and Pakes (1996) use investment as a proxy. Levinsohn and Petrin (2003) argue that investment is lumpy and does not react smoothly to productivity shocks and propose to use material inputs as a proxy instead. In a more recent contribution, Ackerberg et al. (2008) (ACF) present an alternative semi-parametric procedure that deals with potential collinearity issues in Olley and Pakes (1996) and Levinsohn and Petrin (2003). Wooldridge (2009) shows a method to implement OP/LP in a GMM framework with several advantages over ACF: i) estimators are more efficient; ii) the first stage of the algorithm contains identifying information for the parameters on the variable inputs, and iii) fully robust standard errors are easy to obtain. In short, Wooldridge (2009) derives two equations with the same dependent variable (output) and fixed and variable inputs as explanatory variables. The difference between both equations is the approximation of unobserved productivity which provides a different set of instruments for identification of the production function parameters. We use the implementation of Petrin and Levinsohn (2012) of this methodology (referred to as *WLP-TFP* henceforth).

6.2 Estimation and coefficients

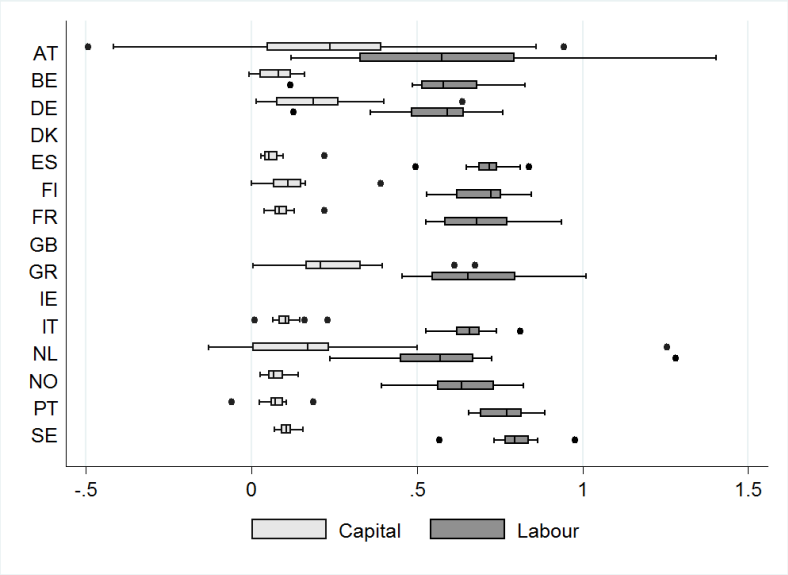
The production function to be estimated is given in its logarithmic form in (1) with ω_{it} the unobserved productivity shock known to the firm but not to the researcher and va double deflated value added (*cf. supra*). The sum of the constant term, β_0, and ω_{it} captures Hicks-neutral TFP. ϵ_{it} is a standard i.i.d. error term incorporating unanticipated shocks and measurement error. As indicated above, we use the GMM-approach advocated by Wooldridge (2009) as implemented by Petrin and Levinsohn (2012). The trade-off we face is between allowing β_l and β_k to vary maximally across countries and industries and retaining enough data points to estimate β_l and β_k.

$$\ln va_{it} = \beta_0 + \beta_l \ln l_{it} + \beta_k \ln k_{it} + \omega_{it} + \epsilon_{it} \tag{1}$$

We first estimate equation (1) by country-industry pair using all available years. Industries are defined as 21 '*broad*' NACE aggregates capturing one or more NACE 2-digit categories (listed in Table 16). Figures 8, 9, and 10 show box plots[28] of the coefficients by country and

[28]In a box plot (see Tukey, 1977), the vertical line within the box indicates the median, while the edges of the box represent the 25th and 75th percentiles. The whiskers of a box indicate the upper and lower adjacent

Figure 8: Boxplot of capital and labour elasticities for the old EU15+ countries.



WLP-TFP estimations by country-industry pair; countries indicated on vertical axis, (values of) capital and labour elasticities on horizontal axis.

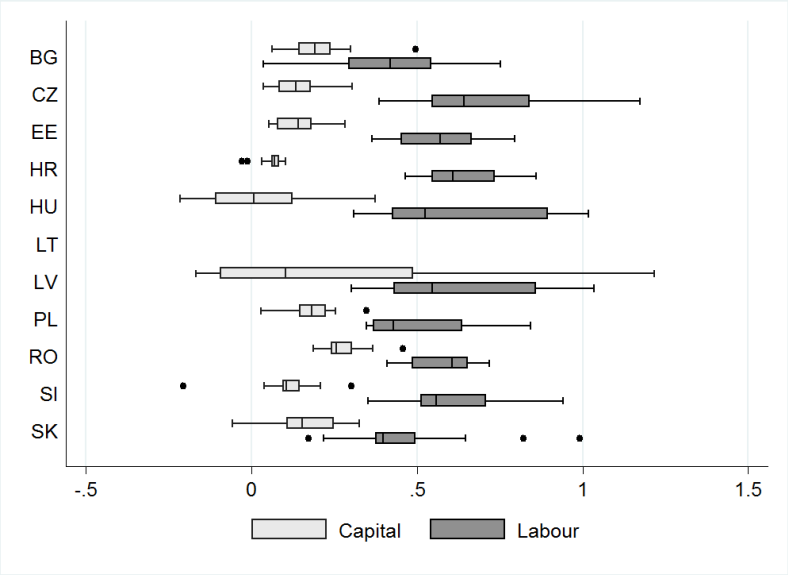
by industry respectively. As one can infer from the box plots in Figures 8, 9, and 10, for multiple countries we obtain capital and labour coefficients that fall outside the unit interval. Moreover, for many country-industries, the capital coefficient is not significant at conventional levels, even when it falls within the unit interval (this is also the case in CompNet (CompNet, 2014)). Gal (2013) deals with this issue by not calculating *TFP* for firms in industries where either the capital or labour coefficient falls outside the unit interval.

We proceed by estimating production functions by industry, but aggregating over countries. When estimating equation (1), we restrict β_l and β_k to be the same across countries but allow β_0 to be country-specific (capturing for example country-specific technology levels or management skills). We realise that this is a strong assumption[29], but we prefer to do so because this results in sensible estimates for capital and labour coefficients as shown in Figure 11. This allows us to obtain *TFP* for the largest possible set of firms. Furthermore, specifically for multinational (foreign) firms (12% of observations in the dataset) a European production function might be as relevant as the 'local' production function. Our analysis

values. These are calculated as follows. Let $x_{[25]}$ and $x_{[75]}$ be the 25th and 75th percentiles of for an ordered variable x. Define U as $x_{[75]} + 1.5(x_{[75]} - x_{[25]})$, the upper adjacent value is then x_i such that $x_i \leq U$ and $x_{i+1} > U$. Define L as $x_{[25]} - 1.5(x_{[75]} - x_{[25]})$, the lower adjacent value is then x_i such that $x_i \geq L$ and $x_{i+1} < L$. Values falling outside of this range are indicated by dots.

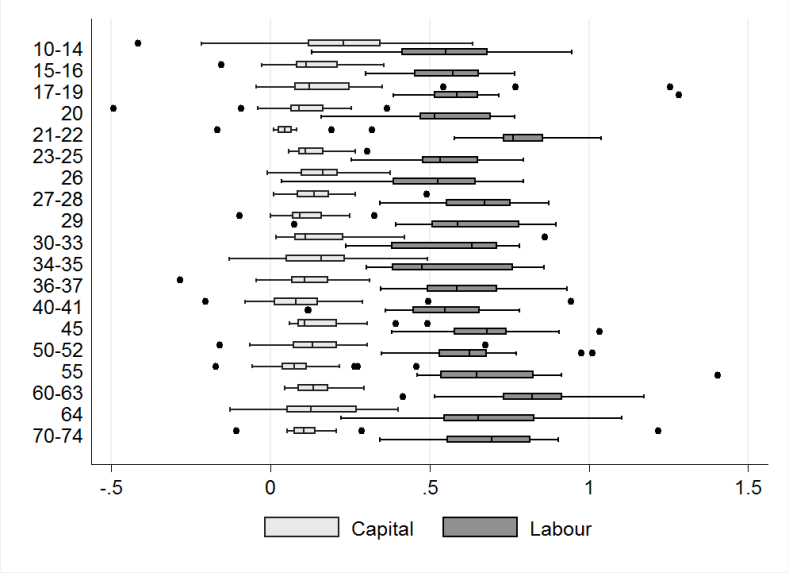
[29]Differences in labour market institutions do exist for example.

Figure 9: Boxplot of capital and labour elasticities for the CEEC10+ countries.



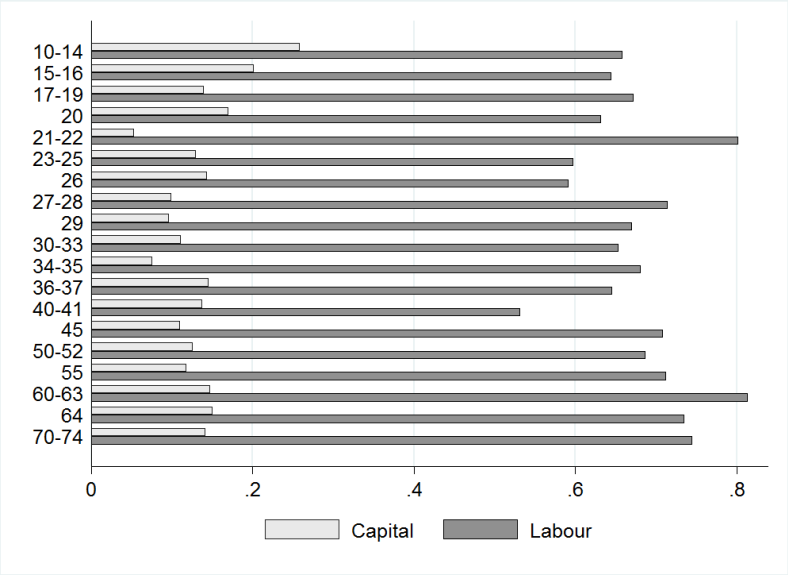
WLP-TFP estimations by country-industry pair; countries indicated on vertical axis, (values of) capital and labour elasticities on horizontal axis.

Figure 10: Boxplot of capital and labour elasticities for 21 broad NACE categories.



WLP-TFP estimations by country-industry pair; industries indicated on vertical axis, (values of) capital and labour elasticities on horizontal axis.

Figure 11: Capital and labour elasticities for Europe-wide production functions by 21 broad NACE categories.



WLP-TFP estimations by industry (aggregated over countries) including country dummies; industries on vertical axis, (values of) capital and labour elasticities on horizontal axis.

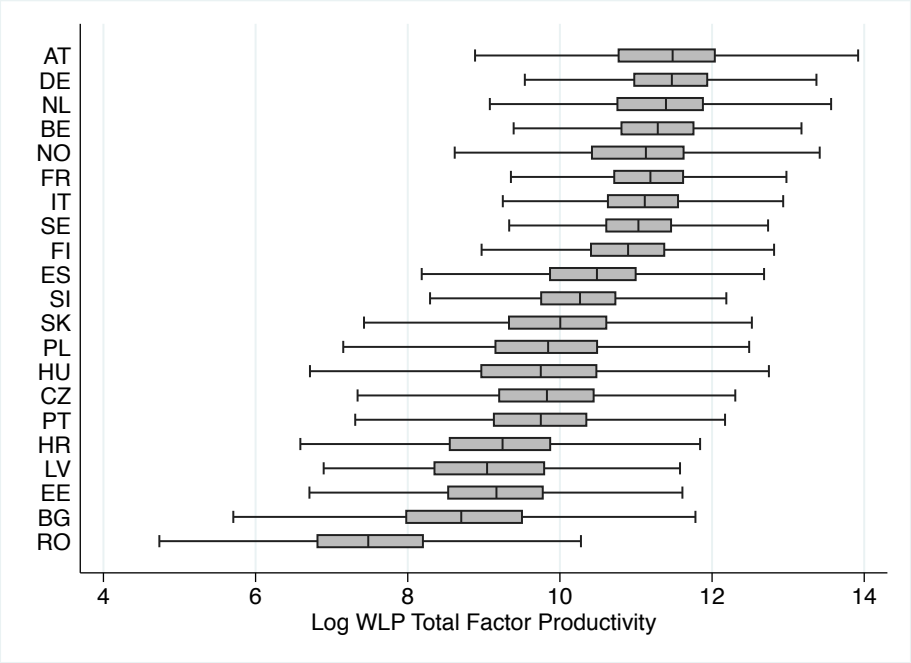
in the next Section is therefore based on *TFP*-values obtained using the estimation results visualised in Figure 11. In Figure 12 we present a box plot of log *WLP-TFP* by country. The Figure is based on the sample of manufacturing firms with more than 20 employees. The period considered is 2003-2010. Countries are ranked on the basis of the 75th percentile of the *TFP*-distribution. The ranking is in line with what one would expect. The old EU-15 countries are generally more productive than the new members from Eastern Europe. Among the old EU-15 Spain and especially Portugal are among the countries with less productive firms.

7 Total Factor Productivity and Foreign Ownership

In this Section we analyse *TFP* differences between multinationals, foreign, and domestic firms. In order to take a first look at the productivity levels of the foreign and domestic firms across Europe, we plot the distributions of their *TFP*-levels in Figure 13.[30] We find that the distribution for foreign firms is clearly to the right of that for domestic firms. Figures 14 and

[30]The period considered is 2003-2007, i.e. we exclude both the earlier years where coverage is more unbalanced across countries and the later years to eliminate potential crisis effects. Only firms with on average at least 20 employees are considered. This leaves us with 1,345,454 observations that are used in the Figure. 166,969, i.e. 12,4%, of observations refer to foreign firms.

Figure 12: Boxplot log WLP Total Factor Productivity by country.



Sample of manufacturing firms with more than 20 employees; 2003-2010; countries ordered by the 75th percentile of the *TFP*-distribution.

15 display the productivity distributions for domestic and foreign firms in manufacturing and services respectively. In both cases, the productivity distribution for foreign firms is to the right of that for domestic firms. For firms in services industries, the distance between both distributions seems larger (*cf. infra*).

When we split Europe in three macro-regions (East, North and South[31]), we can draw a similar conclusion from the first three panels of Figure 16: foreign firms appear to be more productive than domestic firms in all three regions. The last two panels in Figure 16 respectively illustrate the productivity distributions for the domestic firms in the three macro-regions and for the foreign firms in these three regions. For domestic firms we find a clear ranking with firms in the North outperforming firms in the South and firms in the South outperforming firms in the East. With respect to foreign firms, however, the distribution for South is closer to the distribution for North. Foreign firms in the East do seem to be considerably less productive on average.

In order to get further insight into the magnitude of foreign firms' premium in terms of TFP (and several other performance indicators), we perform an empirical exercise along the

[31]East is BG, CZ, EE, HR, HU, (LT,) LV, PL, RO, SI, SK; North is AT, BE, DE, (DK,) FI, FR, (GB,) NL, SE; and South is ES, (GR, IE,) IT, PT. For countries between brackets TFP is not available for the period considered. North contains 425,539 observations; South 516,432; and East 373,783.

Figure 13: *WLP-TFP* distributions for domestic and foreign firms in Europe.

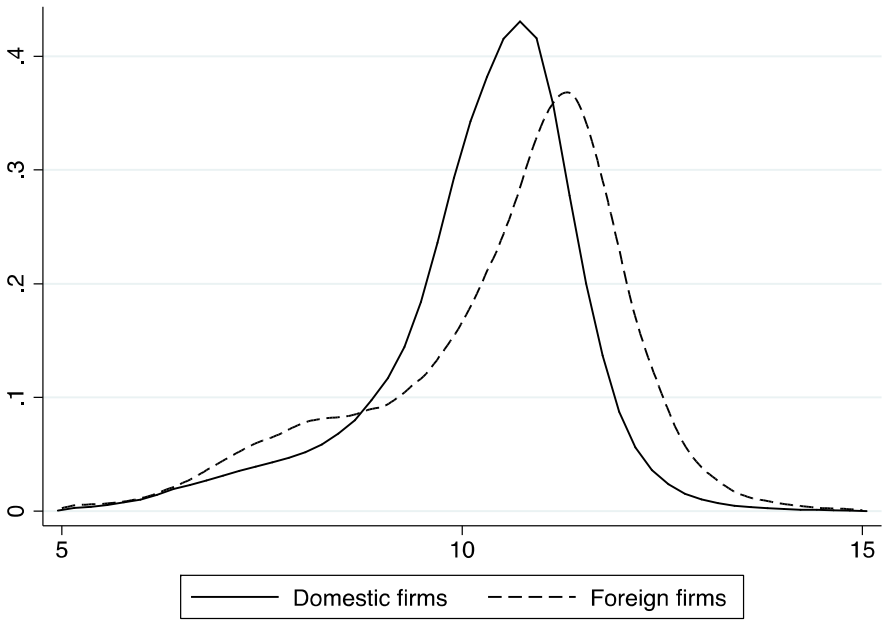


Figure 14: *WLP-TFP* distributions for domestic and foreign firms in Europe in the manufacturing industries.

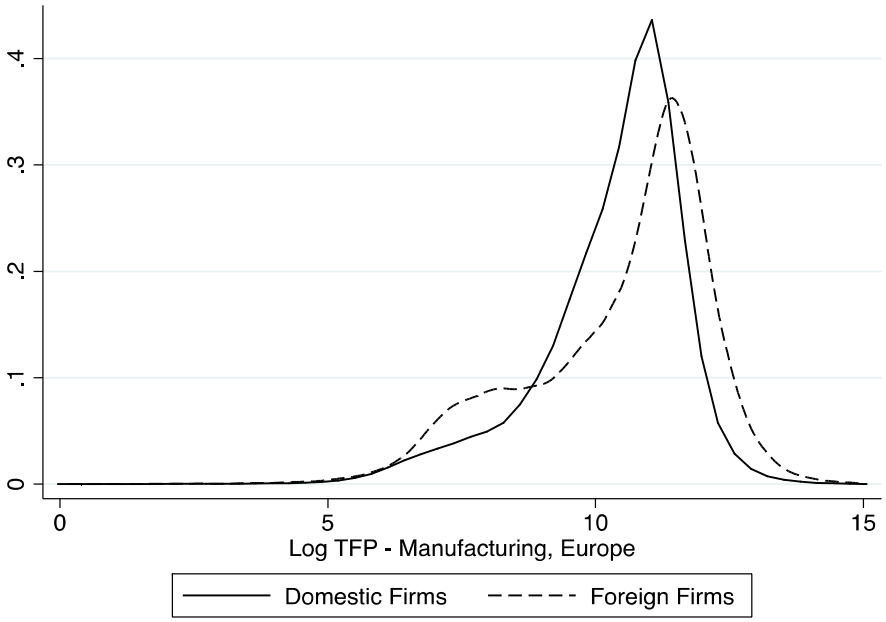
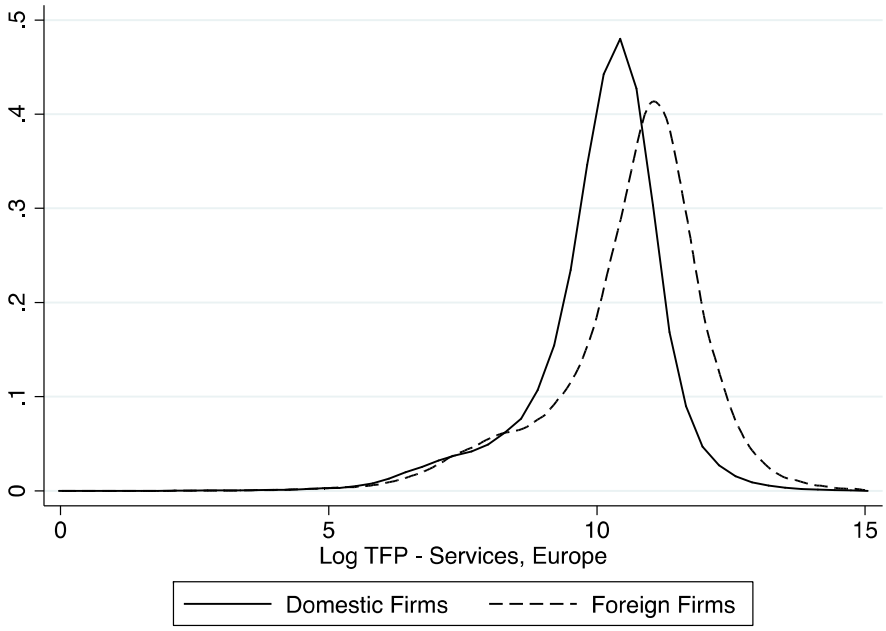


Figure 15: *WLP-TFP* distributions for domestic and foreign firms in Europe in the services industries.



lines of Bernard and Jensen (1999). The analysis is fairly straightforward and consists of retrieving the foreign premium from estimating a regression of the following form:

$$lnX_{ijrt} = \alpha + \beta \text{Foreign}_{ijrt} + \delta L_{ijrt-1} + \gamma_t + \gamma_j + \gamma_r + \epsilon_{ijrt} \tag{2}$$

In equation (2), we regress the level of the performance indicator (X) on a dummy for foreign ownership (*Foreign*), the lagged size of the firm (L, measured as the natural log of the number of employees), and a set of time t, industry j, and region r dummies.[32]

In Table 10 we consider the full sample of foreign and domestic firms for which we are able to obtain *WLP-TFP*, but trim the data for extreme values of *WLP-TFP* by removing values below (above) the first (99th) percentile in each country-industry-size-year cell (to preserve the sample distribution in these dimensions).[33] Table 10 contains the estimated values for β in equation (2). The first column presents premiums for *WLP-TFP* for different subsamples as indicated by the row headings, the third column presents premiums for value added per worker as a comparison check. Columns (2) and (4) contain the number of observations used in the estimation. Premiums are always significant at the 1% level. Controlling for size, time,

[32]Industries are defined as the '*broad*' NACE aggregates. We use NUTS2 region dummies (*cf. supra*).

[33]Size classes are defined as micro firms with less than 10 employees; small firms with between 10 and 50 employees; medium-sized firms with between 50 and 250 employees; and large firms with more than 250 employees.

Figure 16: *WLP-TFP* distributions for domestic and foreign firms in three Macro-regions (North, South and East) and *WLP-TFP* distributions for domestic and foreign firms separtely in these regions compared.

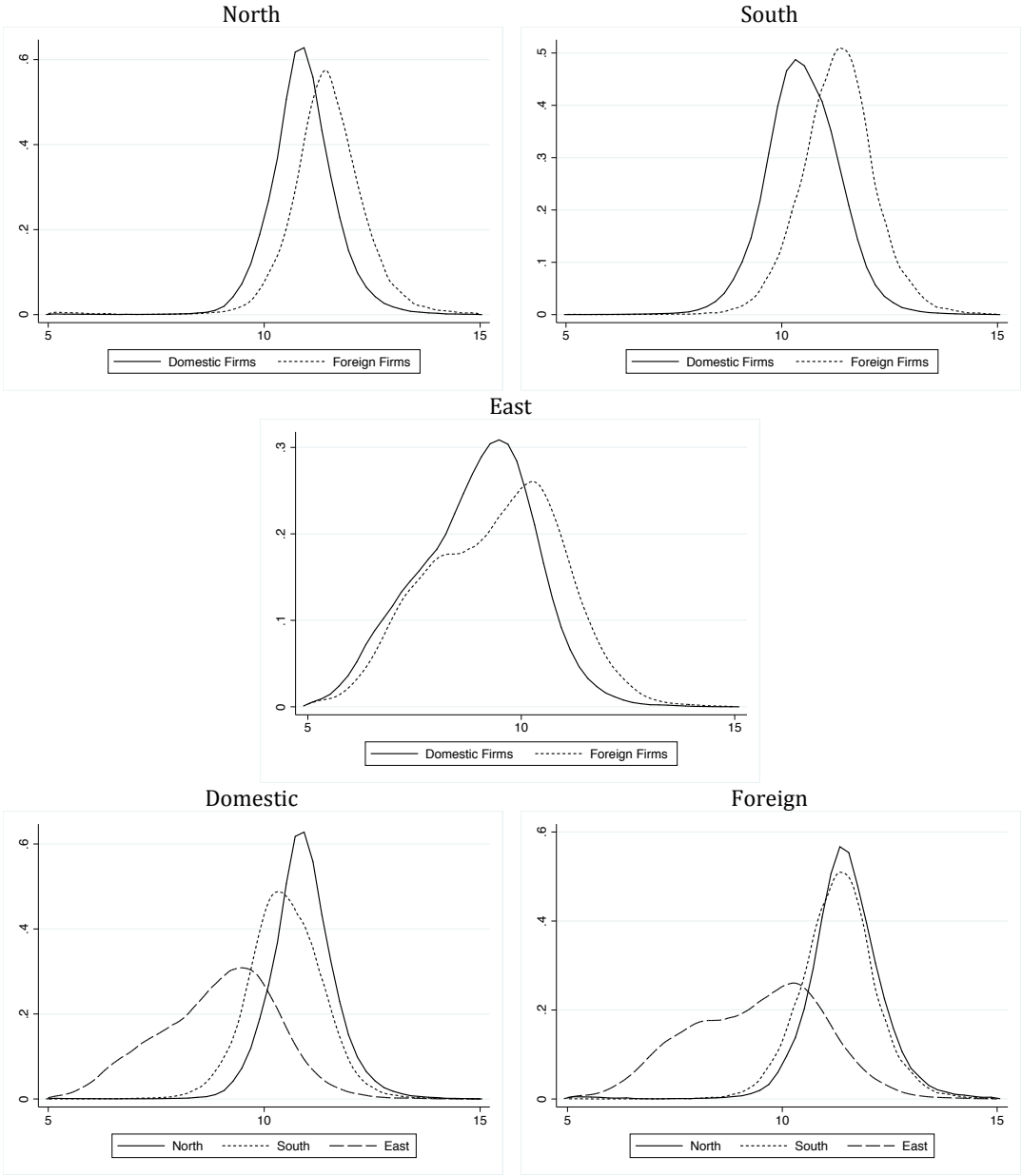


Table 10: TFP premium of foreign over domestic firms based on the EU-wide sample and different subsamples.

	ln WLP-TFP	# obs	ln VA pw	# obs
	(1)	(2)	(3)	(4)
trimmed	0.480	13,023,107	0.503	13,026,194
non-trimmed	0.555	13,238,694	0.573	13,238,694
manufacturing	0.324	3,020,906	0.345	3,021,943
services	0.557	7,919,363	0.584	7,920,568
before 2003	0.490	3,507,385	0.518	3,507,991
2003–2007	0.483	4,714,522	0.511	4,714,983
after 2007	0.469	4,801,200	0.484	4,803,220
majority foreign-owned firms	0.510	13,064,783	0.538	13,069,327
minority foreign-owned firms	0.335	13,064,783	0.365	13,069,327
micro firms (L$\leq$10)	0.532	8,505,507	0.566	8,488,588
small firms (10<L$\leq$50)	0.441	3,462,642	0.458	3,467,654
medium firms (50< $L \leq$250)	0.335	871,972	0.367	879,794
large firms (L>250)	0.329	182,986	0.368	190,158

Premiums for *WLP-TFP* and value added per worker (VA pw). Subsamples are obtained by considering manufacturing and services industries separately, by splitting up the sample period in three shorter periods, by distinguishing between majority and minority foreign-owned firms and by considering four size classes of firms. The foreign premium is statistically significant at the 1% level in all cases in columns (1) and (3). In the trimmed sample, values of *WLP-TFP* below (above) the 1st (99th) percentile in each country-industry-size-year cell are removed.

industry, and region, we find that foreign firms' level of *WLP-TFP* is on average 48% higher in Europe. This number is confirmed for value added per worker in column (3) where we find a 50% premium. For the non-trimmed sample these premiums are about 7 percentage points higher. When we consider manufacturing and services industries separately we find, in line with Figures 14 and 15 above, that the premium is considerably larger for services industries. The premium seems fairly stable over time with potentially a slight tendency to decrease, but given changes in sample constellation (*cf.* Tables 18 and 19 in the Appendix), one should not read too much in this decrease. As indicated above the criterion to classify a firm as foreign is a single foreign owner controlling at least 10% of shares. When we split foreign firms in a group which is majority foreign-owned (more than 50%) and a group which is minority foreign-owned (more than 10% of the shares, but less than 50%)[34], we find that both groups outperform domestic firms, but that majority foreign-owned firms also outperform minority foreign-owned firms. Majority foreign-owned firms are 51% more productive than domestic firms, whereas minority foreign-owned firms are 33% more productive. Finally, we consider four size categories inspired by the EU's definition of micro (employing less than 10 employees), small (between 10 and 50 employees), medium (between 50 and 250 employees), and large (more than 250 employees) firms. The productivity premium decreases by size class. It is well over 50% for micro firms, about 45% for small firms, and about 33% for medium and large firms.

[34]Of the 313,677 foreign firms (after trimming) 51,523 firms are minority foreign-owned, while 262,154 firms are majority foreign-owned.

Table 11: TFP premium of foreign over domestic firms. Cross-country differences for different subsamples.

	All firms	≥ 20 employees				All firms	≥ 20 employees		
	1995–2011	1995–2011	2003–2007	2003–2007 manuf.		1995–2011	1995–2011	2003–2007	2003–2007 manuf.
country	(1)	(2)	(3)	(4)	country	(1)	(2)	(3)	(4)
AT	0.140***	0.140***	0.130**	0.131	**BG**	0.439***	0.502***	0.555***	0.419***
	11,860	10,837	5,048	1,882		173,065	78,099	19,612	6,304
BE	0.259***	0.309***	0.313***	0.266***	**CZ**	0.541***	0.522***	0.536***	0.411***
	127,543	84,281	27,157	10,430		280,183	103,678	49,411	18,995
DE	0.334***	0.283***	0.299***	0.168***	**EE**	0.585***	0.435***	0.395***	0.258***
	111,977	85,763	34,642	12,652		170,610	26,134	9,880	3,392
ES	0.602***	0.443***	0.447***	0.260***	**HR**	0.489***	0.376***	0.395***	0.247***
	3,291,491	516,590	196,458	64,397		322,076	38,612	18,675	5,903
FI	0.461***	0.316***	0.318***	0.214***	**HU**	0.566***	0.517***	0.548***	0.471***
	314,284	47,284	17,990	6,711		56,296	31,345	5,212	2,465
FR	0.382***	0.242***	0.261***	0.093***	**LV**	0.727***	0.313***	0.328***	0.02
	2,668,045	486,644	163,455	55,449		1,743	444	196	103
IT	0.396***	0.354***	0.364***	0.234***	**PL**	0.659***	0.651***	0.674***	0.469***
	1,611,475	461,499	154,451	85,339		137,341	95,899	39,437	14,471
NL	0.323***	0.191***	0.190***	-0.055	**RO**	0.479***	0.358***	0.396***	0.279***
	12,343	7,565	3,367	1,200		2,018,655	255,552	89,272	37,437
PT	0.624***	0.415***	0.388***	0.322***	**SI**	0.381***	0.304***	0.303***	0.273***
	650,243	68,845	14,607	6,357		99,798	20,373	9,566	4,849
SE	0.334***	0.202***	0.168***	0.114***	**SK**	0.632***	0.594***	0.627***	0.518***
	515,775	48,124	19,665	9,756		58,930	21,751	9,625	4,146
NO	0.422***	0.224***	0.268***	0.124***					
	380,418	57,258	21,204	6,225					

Numbers in rows below estimated coefficients refer to the number of firm-year observations used in estimation. ***/**/* denotes significance at 1/5/10 percent. Columns (1) and (2) present results for the full sample period (1995-2011), columns (3) and (4) for the period 2003-2007. Column (1) shows results for all firms, in columns (2)-(4) only firms with at least 20 employees are considered. In column (4) only manufacturing firms with at least 20 employees are considered.

Table 12: TFP premium of foreign over domestic firms. Cross-industry differences for different subsamples.

	All firms	≥ 20 employees			All firms	≥ 20 employees	
	1995–2001 (1)	1995–2011 (2)	2003–2007 (3)		1995–2001 (1)	1995–2011 (2)	2003–2007 (3)
industry	**Manufacturing**			industry	**Mining and quarrying**		
15t16	0.353***	0.332***	0.341***	**10t14**	0.267***	0.265***	0.298***
	430,871	129,738	44,564		54,111	17,311	5,707
17t19	0.436***	0.331***	0.347***				
	322,683	130,767	43,618	industry	**Electricity, gas and water supply**		
20	0.158***	0.097***	0.127***	**40t41**	0.596***	0.518***	0.541***
	187,631	43,839	15,907		48,194	26,159	9,573
21t22	0.336***	0.271***	0.279***				
	295,332	77,745	27,193	industry	**Construction**		
23t25	0.348***	0.276***	0.273***	**45**	0.457***	0.321***	0.368***
	251,737	112,947	39,221		1,962,901	333,339	121,122
26	0.324***	0.290***	0.325***				
	161,834	58,257	20,345	industry	**Services**		
27t28	0.262***	0.207***	0.213***	**50t52**	0.555***	0.475***	0.509***
	579,383	185,578	66,012		4,614,026	597,650	215,078
29	0.277***	0.209***	0.192***	**55**	0.174***	0.174***	0.195***
	274,235	109,596	38,412		914,574	110,682	40,097
30t33	0.277***	0.189***	0.204***	**60t63**	0.597***	0.536***	0.501***
	215,967	79,367	28,219		661,077	145,195	51,872
34t35	0.331***	0.278***	0.306***	**64**	0.608***	0.454***	0.464***
	86,758	39,677	13,934		40,382	9,769	3,714
36t37	0.278***	0.198***	0.206***	**70t74**	0.553***	0.530***	0.528***
	232,107	64,456	22,807		1,689,304	278,834	101,535

Numbers in rows below estimated coefficients refer to the number of firm-year observations used in estimation. ***/**/* denotes significance at 1/5/10 percent. Columns (1) and (2) report results based on the full sample period (1995-2011), results in column (3) refer to the period 2003-2007. Column (1) show results for all firms, in columns (2) and (3) only firms that employ at least 20 employees are considered.

Tables 11 and 12 report foreign *WLP-TFP* premiums by country and by industry respectively.[35] For each country in Table 11 we consider four different samples of firms: *i*) all firms; *ii*) firms with more than 20 employees (which improves the representativeness of our data); *iii*) firms with more than 20 employees for the period 2003-2007 (resulting in a stable number of firms, and excluding crisis effects); and *iv*) firms in manufacturing with more than 20 employees for the period 2003–2007. We examine the fourth sample of firms because the *WLP-TFP* estimation algorithm is probably more tailored towards manufacturing firms.[36] In Table 12, foreign *WLP-TFP* premiums are evaluated by industry (for 21 broad industries) for samples *i*) to *iii*). In both Tables, numbers in rows below the coefficients refer to the number of firm-year observations used in the estimation. In Table 11 we detect the largest premiums for countries in Eastern Europe (on the right-hand side in the Table). For the old EU15 Members States, the largest premiums are found for Portugal, Spain, and Italy. For all countries *WLP-TFP* premiums decrease as the sample becomes more restrictive. The larger premiums are still typically found in Eastern Europe, though for a country such as Slovenia the premium is close to that of the Western European countries. In Table 12, we find that premiums in manufacturing industries range from 12.7% to 34.7% for the most restrictive sample. Except for mining (industries 10-14) and hotels and restaurants (industry 55), the premiums in all other industries are considerably higher than those in manufacturing.

[35] Yasar et al. (2007) and Castellani and Giovannetti (2010) examine TFP and exporter premia for Turkey and Italy. These studies find that foreign firms are more productive than domestic counterparts.

[36] Services do account for large parts of value added in all countries.

Table 13: Foreign ownership and *WLP-TFP* growth 2005-2011. Evidence from a matched sample.

	All (1)	All (2)	All (3)	Manuf. (4)	Services (5)	All (6)
Foreign	0.030***	0.012***		0.028***	0.033***	0.034***
	[0.002]	[0.003]		[0.003]	[0.003]	[0.002]
Foreign (50%)		0.030***	0.039***			
		[0.003]	[0.002]			
Foreign*crisis						-0.009**
						[0.004]
Crisis						-0.161***
						[0.004]
log empl.$_{t-1}$	0.022***	0.022***	0.022***	0.020***	0.024***	0.022***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Age	-0.000***	-0.001***	-0.001***	-0.000***	-0.001***	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Observations	375,122	375,122	375,122	128,385	230,715	375,122
R^2	0.063	0.063	0.063	0.046	0.077	0.063

Estimations contain year, industry, and country dummies. Standard errors in brackets; ***/**/* denotes significance at 1/5/10 percent.

Table 14: Productivity of *parents* with and without foreign affiliates. Evidence from *EU-MULNET*.

	All aff. (1)	$\leq$ 20 aff (2)	$\leq$ 5 aff (3)	Single (4)	$\leq$ 5 aff (5)	$\leq$ 20 aff, manuf. (8)
foreign aff.	0.281***	0.264***	0.238***	0.221***	0.124***	0.165***
	[0.006]	[0.006]	[0.006]	[0.009]	[0.012]	[0.008]
Lagged log empl.	0.061***	0.059***	0.065***	0.083***	0.064***	0.145***
	[0.001]	[0.001]	[0.001]	[0.002]	[0.001]	[0.002]
# foreign aff.					0.082***	
					[0.007]	
Observations	213,154	210,158	189,865	119,247	189,865	54,123
R^2	0.46	0.46	0.46	0.50	0.47	0.64

Estimations contain year, industry, and country dummies. Standard errors in brackets; ***/**/* denotes significance at 1/5/10 percent.

Table 15: Premium of foreign over domestic firms for other performance indicators.

	1995–2011		2003–2007	
performance indicator	premium (1)	# obs (2)	premium (3)	# obs (4)
ln WLP-TFP	0.480	13,023,107	0.266	358,463
ln Value Added per worker	0.503	13,026,194	0.300	361,217
ln Number of Employees*	0.691	11,426,678	0.401	352,576
ln Value Added	0.598	13,003,206	0.369	353,798
ln Capital per worker	0.231	12,992,328	0.298	360,920
ln Operating Revenue	0.565	13,003,535	0.425	353,858
ln Wage	0.326	12,961,377	0.192	361,195
ln Intangible Fixed Assets	0.262	6,247,780	0.298	265,298
ln Total Assets	0.757	12,993,792	0.475	354,003
ln Cash Flow	0.639	10,445,839	0.517	312,699
ln Profit/Loss before Tax	0.771	9,466,307	0.593	278,959
ln Profit/Loss after Tax	0.794	9,129,196	0.616	264,893
Implicit Tax Rate**	-2.457	12,838,221	-2.666	356,248

* specification does not contain lagged number of employees; ** denoted in percentage points. Samples are trimmed for extreme values of the respective indicators by removing values below (above) the 1st (99th) percentile in each country-industry-size-year cell. Columns (1) and (2) show results obtained from a sample covering all firms in all industries in the full sample period (1995-2011). Columns (3) and (4) show results for manufacturing firms with at least 20 employees in theperiod 2003-2007. The foreign premium is statistically significant at the 1% level in all cases in columns (1) and (3).

We further analyse differences in productivity dynamics between domestic and foreign firms over the period 2005–2011, i.e. three years before and three years after the crisis. To mitigate potential endogeneity of foreign ownership, we employ a matching technique to create a missing counterfactual for each foreign firm's performance. We therefore match each foreign firm with a domestic counterpart in the year 2005. We use Stata's *psmatch2* procedure and slightly modify it to make sure that firms are matched within the same 'broad' industry (and country). Observations are only retained if they belong to the common support. We focus on the year 2005 since this allows us to track a sample of decent size through pre- and post crisis years.[37] After obtaining the set of BVD ID numbers of the matched foreign and domestic firms for the year 2005, we add in the financial and other information for all

[37]We retain all firms from *AUGAMA* that are present in the data set at least from 2005 onwards (could be earlier) and at least until 2009 (could be later) and that report sufficient information to obtain *WLP-TFP*. On this subset of *AUGAMA* we run a probit for the year 2005 to explain foreign ownership. The explanatory variables are lagged *WLP-TFP*, the lagged number of employees, and age (see Alfaro and Chen (2012)). The balancing hypothesis is satisfied for all three variables.

available years (2005–2011). We use this sample to evaluate the effect of foreign ownership by comparing the growth performance of matched firms (pre- and during the crisis).

Table 13 presents the results of the estimation of the following equation.

$$\Delta \ln TFP^{WLP}_{ijct} = \alpha + \beta \text{Foreign}_{ijct} + \delta \ln L_{ijct-1} + \kappa \text{Age}_{ijct} + \gamma_t + \gamma_j + \gamma_c + \epsilon_{ijct} \qquad (3)$$

We regress *WLP-TFP* growth of firm i in industry j in country c at time t on a dummy indicating foreign ownership, the lagged size of the firm, the age of the firm, and a set of industry, country, and year dummies. In the first column we find that the growth rate of total factor productivity of foreign-owned firms is about three percentage points higher than that of matched domestic firms. This is mainly driven by majority foreign-owned firms that record growth rates that are about four percentage points higher than that of domestic firms, whereas minority foreign-owned firms that record growth rates that are about 1.2 percentage points higher. The difference between foreign and domestic firms is larger for services industries according to point estimates, but the difference is not statistically significant. The last column shows that the difference in growth rates between foreign and domestic firm narrows during the crisis period but the combination of the foreign dummy and its interaction with the crisis variable (which equals one from 2008 onwards) suggests that foreign firms still outperform domestic counterparts by a little more than two percentage points in terms of *TFP* growth.

In Table 14 we compare the *WLP-TFP* level of parents with only domestic affiliates with that of parents with at least one foreign affiliate, i.e. an affiliate located in another European country than that of the parent. From *EUMULNET* we retain all parents and information on whether or not they own a foreign affiliate (*ForAff*). We then run the following regression for parents p in industry j in country c at time t:

$$\ln TFP^{WLP}_{pjct} = \alpha + \beta \text{ForAff}_{pjct} + \delta \ln L_{pjct-1} + \gamma_t + \gamma_j + \gamma_c + \epsilon_{ijct} \qquad (4)$$

In the first column of Table 14 we consider all parents when estimating (4). We find parents that have at least one foreign affiliate to be 28% more productive than parents with only domestic affiliates. In columns two to four we restrict the sample by considering only firms with less than 21, 6, or exactly 1 affiliate. The *TFP* premium decreases, but when we focus on parents with only a single affiliate (foreign or domestic), we still find a productivity premium of 22% for parents with a single foreign affiliate. This evidence is in line with Antràs and Helpman (2004) who show that only the most productive domestic firms will set up a foreign affiliate. When we introduce the number of foreign affiliates as additional variable

in the specification in column five (the sample is restricted to parents with at most five affiliates), we find that an additional foreign affiliate is associated with an eight percentage points larger *TFP* premium over parents with only domestic affiliates. Column six finally shows that the productivity premium decreases to 16.5% when we only consider parents in manufacturing industries.

Table 15, finally, considers other performance indicators than productivity. For ease of comparison the first two lines repeat the results of Table 10 for *WLP-TFP* and value added per worker. All performance indicators have been trimmed in a similar way as indicated for *TFP* above. The results reported in column (1) cover a sample of all firms and industries in period 1995–2011; the results in column (3) are obtained from an analysis using only manufacturing firms that employ at least 20 employees in period 2003–2007. We first focus on the results for the full sample of firms. Aside from being about 50% more productive on average, foreign firms create 60% more value added, have 56% more operating revenues, and generate a 64% larger cash flow. They do so by employing about 70% more employees than domestic firms, using 23% more capital per employee, 26% more intangibles (based on the subsample of firms reporting strictly positive intangibles). Foreign firms' total assets are on average 75% larger. All this results in profits that are slightly less than 80% larger (based on the subsample of firms reporting strictly positive profits). Their implicit tax rate, calculated as profits before tax minus profits after tax divided by profits before tax, is on average 2.5 percentage points smaller than domestic firms' average implicit tax rate. When we restrict the sample to manufacturing firms with at least 20 employees in 2003-2007, we confirm the premiums of foreign over domestic firms for all performance indicators. As can be seen from column (3), the premiums are generally smaller. From Table 10 we know that this is more likely to be driven by the exclusion of services firms than by the narrower period considered. Only for capital per worker and intangible fixed assets the premium slightly increases. The difference in implicit tax rates is also slightly larger for this subsample.

8 Conclusions

This paper in detail documents the build of two datasets on the basis of raw data taken from the Amadeus database issued by Bureau Van Dijk Electronic Publishing. The first dataset, *AUGAMA*, is a large panel of firms in 26 European countries. More than 18 million observations with all information needed to obtain a measure of *TFP* are available for more than 3.6 million firms. The data stretch the period 1995-2012 and for the average firm 6.1 years of data is available. The coverage for the years before 2002 is generally lower (not for all countries though) and not for all firms information for 2012 is already available. We

show that *AUGAMA* adequately approximates the structure of the European economy across countries, regions, and industries as portrayed by data from Eurostat (Structural Business Statistics) and Cambridge Econometrics. The second dataset, *EUMULNET*, is a dataset of European multinational networks with 'full' information, i.e. *TFP*, for both parents and their European affiliates. We have more than 600,000 parent-affiliate-year observations, 16% of these are links between a parent and affiliate in different countries. The period is 1997-2012, but as with *AUGAMA* the coverage is lower for earlier years and information for 2012 is not available for all parent-affiliate combinations. The advantage of *AUGAMA* and *EUMULNET* is that both datasets cover cross-country comparable firm-level data for a large number of European countries in a single dataset. This allows for cross-country research at the firm level while maintaining representativeness that is e.g. comparable to the 'distributed micro-data analysis' by CompNet ((CompNet, 2014)).

We use *AUGAMA* to estimate foreign firms' productivity premium over domestic firms. We follow Bernard and Jensen (1999) but estimate productivity premiums for foreign firms rather than for exporters using simple regression analysis. We find that across Europe on average foreign firms are 48% more productive than domestic firms. This is mainly driven by services sectors where foreign firms are on average about 56% more productive, whereas the difference amounts to 32% in manufacturing industries. Majority foreign-owned firms (>50% of the shares are foreign-owned) outperform minority foreign-owned firms, who in turn still outperform domestic firms. We also find the premium to be smaller in larger firm-size categories. Productivity premiums are typically larger in Eastern European countries than in Western European countries. For a number of other firm characteristics (e.g. value added, profitability, intangible assets) we find similar premiums for foreign firms. In a framework similar to Alfaro and Chen (2012) we find that foreign firms grow about 3%-points faster over the period 2005–2011. During the crisis period the growth gap between foreign and domestic firms becomes smaller but remains positive. Finally, in line with Antràs and Helpman (2004) we find parents with foreign affiliates to be 27% more productive than parents with only domestic affiliates. The number decreases to 14.4% when we only consider parents in manufacturing industries.

9 Appendix

Table 16: List of the NACE 2-digit industries included in the data.

Broad category	NACE 2-digit	Description
C		**Mining and quarrying**
C	10	Mining of coal and lignite; extraction of peat
C	11	Extraction of crude petroleum and natural gas
C	12	Mining of uranium and thorium ores
C	13	Mining of metal ores
C	14	Other mining and quarrying
D		**Manufacturing**
DA	15	Manufacture of food products and beverages
DA	16	Manufacture of tobacco products
DB	17	Manufacture of textiles
DB	18	Manufacture of wearing apparel; dressing and dyeing of fur
DC	19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
DD	20	Manufacture of wood and products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
DE	21	Manufacture of pulp, paper and paper products
DE	22	Publishing, printing and reproduction of recorded media
DF	23	Manufacture of coke, refined petroleum products and nuclear fuel
DG	24	Manufacture of chemicals and chemical products
DH	25	Manufacture of rubber and plastic products
DI	26	Manufacture of other non-metallic mineral products
DJ	27	Manufacture of basic metals
DJ	28	Manufacture of fabricated metal products, exc. machinery/equipment
DK	29	Manufacture of machinery and equipment n.e.c.
DL	30	Manufacture of office machinery and computers
DL	31	Manufacture of electrical machinery and apparatus n.e.c.
DL	32	Manufacture of radio/television/communication equipment/apparatus
DL	33	Manufacture of medical/precision/optical instruments, watches/clocks
DM	34	Manufacture of motor vehicles, trailers and semi-trailers
DM	35	Manufacture of other transport equipment
DN	36	Manufacture of furniture; manufacturing n.e.c.
DN	37	Recycling

Table continued on the next page

Table 16: List of the NACE 2-digit industries included in the data (Continued).

Broad category	NACE 2-digit	Description
E		**Electricity, gas and water supply**
E	40	Electricity, gas, steam and hot water supply
E	41	Collection, purification and distribution of water
F		**Construction**
F	45	Construction
G		**Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods**
G	50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel
G	51	Wholesale trade and commission trade, except of motor vehicles and motorcycles
G	52	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods.
H		**Hotels and restaurants**
H	55	Hotels and restaurants
I		**Transport, storage and communication**
I	60	Land transport; transport via pipelines
I	61	Water transport
I	62	Air transport
I	63	Supporting and auxiliary transport activities; activities of travel agencies
I	64	Post and telecommunications
K		**Real estate, renting and business activities**
K	70	Real estate activities
K	71	Renting of machinery and equipment without operator and of personal and household goods
K	72	Computer and related activities
K	73	Research and development
K	74	Other business activities

Table 17: Definition of Nomenclature of territorial units for statistics (NUTS) of the European Union. Minimum and maximum population thresholds indicated.

NUTS level	Description	Min.	Max.
NUTS 1	Major socio-economic regions	3 million	7 million
NUTS 2	Basic regions for application of regional policies	800,000	3 million
NUTS 3	Small regions for specific diagnoses	150,000	800,000

See also "Regions in the European Union; Nomenclature of territorial units for statistics, NUTS 2006/EU-27". NUTS favours administrative divisions. `http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-020/EN/KS-RA-07-020-EN.PDF`.

NUTS regions: Example for Belgium

Level	Code	Description
NUTS 0	BE	Belgique / België
NUTS 1	BE1	Region de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest
NUTS 2	BE10	Region de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest
NUTS 3	BE100	Arrondissement Bruxelles-Capitale / Arrondissement Brussel-Hoofdstad
NUTS 1	BE2	Vlaams gewest
NUTS 2	BE21	Provincie Antwerpen
NUTS 3	BE211	Arrondissement Antwerpen
NUTS 3	BE212	Arrondissement Mechelen
NUTS 3	BE213	Arrondissement Turnhout
...		

Table 18: Number of firms with at least 20 employees and *WLP-TFP* available in *AUGAMA* (period 1996–2011).

	avg. # years	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
AT	4.3	118	243	331	403	331	266	952	1,349	1,793	1,699	2,074	2,006	1,875	1,768	1,982	1,703
BE	10.0	5,322	5,356	5,478	5,450	5,640	5,691	5,740	5,884	6,037	6,132	6,311	6,470	6,547	6,587	6,681	6,630
BG	6.5	3,921	3,260	3,745	6,158	6,240	6,194	3,412	3,693	3,767	4,238	7,248	9,540	11,466	11,793	11,534	6,599
CZ	7.1	153	147	119	425	4,152	5,608	7,710	9,228	11,456	12,522	13,584	14,574	12,148	15,401	14,604	10,455
DE	3.7	1,073	1,304	1,834	2,593	2,862	3,176	5,794	8,445	9,954	13,426	15,682	15,495	15,359	15,541	13,384	12,561
EE	8.7	220	1,620	1,939	2,217	2,405	2,218	2,375	2,398	2,085	2,138	2,162	2,215	2,116	1,789	1,748	1,689
ES	9.7	23,376	25,347	28,042	29,847	33,563	36,217	39,636	41,510	43,317	45,206	47,160	46,816	46,790	48,029	45,200	34,593
FI	7.6	2,275	3,212	3,698	3,779	3,984	3,969	4,203	4,363	4,349	4,385	4,533	4,553	3,961	4,017	3,635	3,826
FR	8.1	31,019	35,325	47,357	52,495	53,110	44,176	43,341	43,606	43,778	43,343	38,054	36,518	31,714	30,353	34,671	38,268
GR	2.2			4,566	2,126	3,734	1,890										
HR	8.4	18	27	2,833	3,583	392	647	3,646	3,964	3,990	4,105	4,341	4,372	4,380	4,637	4,492	4,326
HU	4.2	407	2,028	2,754	2,571	2,034	1,093	4,423	1,066	1,119	2,558	3,046	9,831	5,555	9,905	5,759	4,940
IT	8.1	28,824	31,175	32,775	32,409	34,513	34,414	38,071	36,453	36,872	29,778	39,469	40,557	37,030	37,194	34,074	49,012
LV	4.3					14	33	43	56	68	75	82	90	73	64	134	134
NL	3.5	236	244	246	501	575	661	932	943	1,003	1,079	1,111	1,212	1,266	1,246	984	474
NO	7.5	3,709	1,932	1,999	2,295	2,492	2,553	2,530	2,722	5,877	6,018	6,092	6,461	6,704	6,858	6,748	6,736
PL	5.0	921	2,063	2,251	4,538	5,651	6,652	6,718	8,161	8,611	9,649	13,394	11,948	13,789	22,246	10,821	6,477
PT	5.5	1,472	1,442	1,804	474	669	618	660	798	793	810	14,119	14,407	14,756	14,761	13,934	12,780
RO	8.9	12,682	13,510	14,353	15,736	17,726	17,639	18,480	20,262	21,465	22,832	20,004	23,985	22,835	21,280	21,500	21,472
SE	6.8		2,229	5,370	5,690	6,085	5,861	6,390	5,435	4,134	4,445	4,453	4,133	1,612	1,651	1,630	1,673
SI	7.6		963	458	401	709	825	1,704	1,877	2,041	2,131	2,304	2,304	2,273	2,324	2,782	2,688
SK	5.9	10	49	107	167	364	744	1,152	1,686	1,689	3,007	3,553	3,606	3,021	4,178	3,660	2,293

Table 19: Number of *foreign* firms with at least 20 employees and *WLP-TFP* available in *AUGAMA* (period 1996–2011).

	avg. # years	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
AT	4.9	36	52	60	87	71	52	206	345	491	461	552	519	484	434	507	429
BE	12.1	993	1,012	1,039	1,026	1,050	1,047	1,100	1,170	1,232	1,270	1,290	1,369	1,366	1,370	1,418	1,380
BG	8.1	249	223	312	557	639	683	487	541	574	638	869	1,041	1,142	1,145	1,089	705
CZ	7.7	32	33	14	97	845	1,226	1,571	1,816	2,225	2,380	2,589	2,738	2,367	2,911	2,839	2,136
DE	4.8	104	114	178	256	255	311	568	826	960	1,654	2,193	2,349	2,430	2,422	2,130	1,403
EE	8.8	65	379	431	501	555	535	570	569	487	503	516	524	500	435	429	416
ES	10.4	1,445	1,649	1,674	1,738	1,922	2,001	2,118	2,177	2,261	2,262	2,352	2,363	2,366	2,462	2,380	1,703
FI	9.1	241	299	331	317	339	372	422	452	456	472	497	480	408	427	396	384
FR	9.0	2,297	2,725	3,359	3,528	3,622	3,120	3,175	3,197	3,173	3,172	2,917	2,864	2,290	2,180	2,588	2,451
GR	2.2			348	135	236	152										
HR	8.9	2	3	226	273	20	41	304	346	378	395	405	425	436	454	450	445
HU	6.5	89	355	436	400	300	187	719	212	206	390	299	613	552	656	659	580
IT	10.4	1,961	2,075	2,188	2,161	2,249	2,229	2,294	2,272	2,393	2,223	2,451	2,504	2,438	2,398	2,335	2,341
LV	5.5					2	10	13	17	25	28	28	29	32	25	33	32
NL	4.4	55	61	69	133	145	161	226	222	217	237	248	268	311	332	247	121
NO	8.1	279	132	141	183	213	211	213	228	522	544	544	534	565	566	593	608
PL	6.1	177	335	437	893	988	1,175	1,285	1,519	1,608	1,918	2,260	2,148	2,527	3,454	2,229	1,285
PT	6.0	81	64	95	40	51	48	55	66	66	58	563	613	730	756	708	615
RO	9.0	2,778	3,031	3,452	3,933	4,251	3,872	4,039	4,397	4,728	5,119	4,605	5,231	5,013	4,781	4,767	4,783
SE	7.3		151	330	354	380	380	430	304	199	215	223	206	164	169	165	175
SI	9.3		121	61	63	100	128	235	255	280	296	321	325	329	333	352	346
SK	6.5	2	10	20	33	80	161	248	339	323	648	748	726	631	765	678	477

References

Ackerberg, D., Caves, K., and Frazer, G. (2008). Structural Identification of Production Functions. mimeo.

Alfaro, L. and Chen, M. X. (2012). Surviving the Global Financial Crisis: Foreign Ownership and Establishment Performance. *American Economic Journal: Economic Policy*, 4(3):30–55.

Antràs, P. and Helpman, E. (2004). Global Sourcing. *Journal of Political Economy*, 112(3):552–580.

Becker, S., Egger, P., and von Ehrlich, M. (2010). Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics*, 94:578–590.

Becker, S., Egger, P., and von Ehrlich, M. (2012). Too much of a good thing? on the growth effects of the EU's regional policy. *European Economic Review*, 56(4):648–668.

Bernard, A. B. and Jensen, J. B. (1999). Exceptional exporter performance: cause, effect, or both? *Journal of International Economics*, 47(1):1–25.

Bureau van Dijk Electronic Publishing (various issues). Amadeus database. Brussels, Belgium.

Castellani, D. and Giovannetti, G. (2010). Productivity and the international firm: dissecting heterogeneity. *Journal of Economic Policy Reform*, 13(1):25–42.

CompNet (2014). Micro-Based Evidence of EU Competitiveness: The CompNet Database. Working Paper No. 1634, European Central Bank, Frankfurt.

European Commission (2011). Regions in the European Union—nomenclature of territorial units for statistics nuts 2010. Working Paper ISSN 1977-0375, European Commission, Brussels.

Gal, P. N. (2013). Measuring Total Factor Productivity at the Firm–Level using OECD–ORBIS. Working Paper No. 1049, Organisation for Economic Co-operation and Development, Paris.

Griliches, Z. and Mairesse, J. (1995). Production Functions: The Search for Identification. Working Paper No. 5067, The National Bureau of Economic Reserach, Cambridge.

Helpman, E., Melitz, M. J., and Yeaple, S. R. (2004). Exports versus FDI with Heterogenous Firms. *American Economic Review*, 94(1):300–316.

Javorcik, B. S. (2004). Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages. *American Economic Review*, 94(3):605–627.

Levinsohn, J. and Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2):317–341.

Markusen, J. R. (1995). The Boundaries of Multinational Enterprises and The Theory of International Trade. *Journal of Economic Perspectives*, 9(2):169–189.

Merlevede, B., Schoors, K., and Spatareanu, M. (2014). FDI Spillovers and Time since Foreign Entry. *World Development*, 56:108–126.

Olley, S. G. and Pakes, A. (1996). Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297.

Petrin, A. and Levinsohn, J. (2012). Measuring Aggregate Productivity Growth using Plant-Level Data. *The RAND Journal of Economics*, 43(4):705–725.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2):326–365.

Timmer, M. P. (2012). The World Input-Output Database (WIOD): Contents, Sources and Methods. Working Paper No. 10, World Input-Output Database, Groningen. (Downloadable at http://www.wiod.org/publications/papers/wiod10.pdf).

Wooldridge, J. M. (2009). On Estimating Firm-Level Production Functions using Proxy Variables to Control for Unobservables. *Economics Letters*, 104:112–114.

Yasar, M., Garcia, P., Nelson, C., and Rejesus, R. (2007). Is there Evidence of Learning-by-Exporting in Turkish Manufacturing Industries? *International Review of Applied Economics*, 21(2):293–305.