DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

Sectoral Determinants of Foreign Affiliate Sales

by

Eddy BEKKERS
Indre MACSKASI

Working Paper No. 1507
October 2015

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

eddy.bekkers@wti.org
phone: +41316313270

Sectoral Determinants of Foreign Affiliate Sales*

Eddy Bekkers

University of Bern

Indre Macskasi

Johannes Kepler University Linz

ABSTRACT: We test recent theories on sectoral determinants of foreign affiliate sales employing European foreign affiliate sales statistics (FATS). On the one hand, we test hypotheses by Oldenski (2012) that foreign affiliate sales are less likely in sectors with complex tasks and more likely in sectors where communication with customers is important. On the other hand, we test the hypothesis by Keller and Yeaple (2013) that the force of gravity is stronger in more complex sectors. Employing Poisson and Negative Binomial estimators, we find support for the hypotheses in Oldenski (2012) and contradict the hypothesis in Keller and Yeaple (2013).

Keywords: Affiliate Sales, European Affiliate Sales Data, Task Complexity, Communication with Customers, Gravity of Knowledge

JEL codes: F23

printdate: October 7, 2015

*We thank Michael Landesmann for useful comments. Address for correspondence: Eddy Bekkers, University of Bern, World Trade Institute, Hallerstrasse 6. 3012 Bern, Switzerland. +41316313270. Email: eddy.bekkers@wti.org.

Sectoral Determinants of Foreign Affiliate Sales

ABSTRACT: We test recent theories on sectoral determinants of foreign affiliate sales employing European foreign affiliate sales statistics (FATS). On the one hand, we test hypotheses by Oldenski (2012) that foreign affiliate sales are less likely in sectors with complex tasks and more likely in sectors where communication with customers is important. On the other hand, we test the hypothesis by Keller and Yeaple (2013) that the force of gravity is stronger in more complex sectors. Employing Poisson and Negative Binomial estimators, we find support for the hypotheses in Oldenski (2012) and contradict the hypothesis in Keller and Yeaple (2013).
Keywords: Affiliate Sales, European Affiliate Sales Data, Task Complexity, Communication with Customers, Gravity of Knowledge

JEL codes: F23

1 Introduction

Established theory on the determinants of foreign affiliate sales of multinational firms distinguishes between vertical and horizontal FDI. The knowledge capital model (Markusen (2002)) and extended knowledge capital model (Bergstrand and Egger (2007)) nest the different types of FDI. A host of country-level, bilateral and sectoral determinants like income, income-differences, skill abundance-differences, distance and the importance of plant level and firm level scale effects emerge from this theory. Exploring the effect of these variables on foreign affiliate sales empirically, researchers find support for both types of FDI. Recent literature on the determinants of multinational activity has shifted away from the distinction between horizontal and vertical FDI, focusing attention on sectoral determinants of FDI related to the costs of transferring knowledge and information. Oldenski (2012) claims that sectors with higher costs related to communicating complex information, measured by the intensity of non-routine tasks, will display less affiliate sales relative to exports. The more intensive a sector is in non-routine tasks, the more expensive the transfer of information between an affiliate and its parent company. Firms thus prefer exports over affiliate sales. Instead in industries where communication with customers is important, firms will prefer affiliate sales over exports. Keller and Yeaple (2013) argue that gravity forces exert a stronger negative influence on affiliate sales in more complex industries, as measured by R&D-intensity. In more complex sectors foreign affiliates rely more on production inputs from their headquarters. Therefore, increasing trade costs have a stronger impact on an affiliate's marginal costs and thus depress affiliate sales more.[1]

Both Oldenski (2012) and Keller and Yeaple (2013) test their theories using data on affiliate sales of multinational firms with headquarters located in the US, finding support for their hypothesis. Also other research on affiliate sales has worked with the American data (Bergstrand and Egger (2007), Chen (2009), Yeaple (2009), Ramondo, et al. (2011), Antras and Yeaple (2013)). This raises the question of the generalizability of these findings to other countries. A particular drawback of the American data is that there is only one parent country, which is moreover always bigger than the host country. Besides the above point on the generalizability of the findings to different countries, we observe that Keller and Yeaple (2013) seem to omit an important econometric specification in their work. They only include distance as a measure of gravity and the interaction between distance and measures of complexity to test their theory. But measures of complexity are not included as separate regressor.

In this paper we explore the robustness of the findings in the recent literature on the sectoral determinants of affiliate sales to the use of a different dataset and different econometric specifications using European affiliate sales data. We work with inward foreign affiliate sales data collected by Eurostat which is part of FATS, Foreign Affiliates Statistics (Eurostat Metadata, 2012). The data consist of both local affiliate sales and exports in 22 countries by foreign affiliates from 38 countries. Since 2007 data collection is mandatory and henceforth we work with FATS data for 2008-2011. A potential drawback of our dataset is that about 20% of the industry-country pairs is missing, mainly due to confidentiality issues.[2] We do not think this invalidates our analysis. On the one hand, our results are in line with previous literature on the conventional drivers of affiliate sales like source and destination GDP, third country GDP, distance and skill differences. On the other hand we show that the missings are hardly related to variation across sectors and are thus not likely to create a big selection problem for our analysis on sectoral determinants. Moreover, we correct for selection bias in the estimation. To our knowledge only one study has thus far used the FATS data to analyse affiliate sales, Fukui and Lakatos (2012). These scholars employ data for the period 2003-2007, when reporting was still voluntary, and moreover they do not address the sectoral determinants which are the main

[1] Also Helpman, et al. (2008) focus on a sectoral determinant of affiliate sales, the productivity dispersion of firms, but we do not test this recent theory due to a lack of productivity dispersion data at the sectoral level for the countries in our sample.

[2] Data are confidential when direct or indirect identification of statistical units is possible due to too few reporting enterprises.

focus of our paper.[3]

In the analysis we employ two estimators used frequently in gravity estimations of international trade flows, the (Poisson) Pseudo Maximum Likelihood estimator (PPML) and the Zero Inflated Negative Binomial estimator (ZINB). Since the study by Silva and Tenreyro (2006) pointing at the inconsistency of OLS estimators in the presence of heteroskedasticity after log transformation of the data, PPML has become the standard in gravity estimation. ZINB is appropriate for our data with a very large number of zeros. Egger and Staub (2014) show in a Monte Carlo study that these two estimators outperform other estimators. Since affiliate sales are similar to international trade flows and thus subject to the same heteroskedasticity-problem after log transformation, we think the employed estimators are better then the OLS estimator employed thus far in the analysis of affiliate sales.

Our empirical analysis generates three main sets of results. First, explanatory variables following from the established literature are significant and have the expected sign. GDPs in the host and source country have a positive and significant coefficient and GDP in the rest of the world and an index for FDI restrictiveness display negative and siginficant coefficients. Remarkable is that distance between source- and host-country has a significant negative impact on affiliate sales, lending support for theories of vertical FDI. Second, different measures of complexity of tasks in a sector have a negative and significant impact on affiliate sales, whereas measures of communication with customers have a positive impact, both in line with Oldenski (2012). We derive this result not only employing the American O*NET data on task complexity and communication, but also the recently released task data from OECD (2014). The effect of communcation with customers is not robust to the inclusion of a dummy for services. Third, the interaction term between distance and R&D intensity or task complexity is positive when including R&D intensity or task complexity also as a separate variable in the regression, contradicting the findings in Keller and Yeaple (2013). Without including these variables separately we find a significant negative coefficient on the interaction terms between distance and both R&D intensity and task complexity, in line with the findings in Keller and Yeaple (2013). These authors do not include R&D intensity separately in any of their robustness checks. Our empirical analysis indicates that affiliate sales are discouraged in more complex sectors as claimed by Oldenski (2012) and the effect of distance is not stronger in such sectors as claimed by

[3] Cravino and Levchenko (2013) and Alvariez (2013) both also deviate from working with the American data, employing respectively ORBIS firm level data and unpublished OECD data.

Keller and Yeaple (2013) but weaker, i.e. we find a positive interaction term of distance and sector-complexity. We thus find support for our assertion that the econometric specification matters. The theory as well as the empirics in Keller and Yeaple (2013) are on firm-level affilaite sales, whereas our empirical analysis is based on aggregate (industry-level) data. We show in a webappendix that the prediction on the enhanced effect of distance on affiliate sales in more complex sectors also holds at the aggregate (sectoral) level. Also along the extensive margin distance has a stronger effect in more complex sectors. We then describe how an extension of the theory in Keller and Yeaple (2013) with both horizontal and vertical FDI would be able to reconcile our findings with the theory in Keller and Yeaple. We argue that a positive interaction term for aggregate sales can go together with a negative interaction term at the firm level, if there is a composition effect between horizontal and vertical affiliate sales from changes in distance and if the impact of distance on affiliate sales is different for horizontal and vetical affiliate sales. Our explanation for this effect will be presented in section 5.3.

Before discussing the data into more detail in section 3, in the next section we first review the tested theories. In section 4 we outline the estimation methods and in section 5 we present the empirical findings. Section 6 concludes.

2 Review of Theory

In this paper we concentrate on the sectoral determinants of affiliate sales proposed in recent work by Oldenski (2012) and Keller and Yeaple (2013). As control variables we include the determinants of FDI following from established theory as summarized by the knowledge capital model (Markusen (2002)). Therefore, we first briefly describe these determinants following the knowledge and physical capital model in Bergstrand and Egger (2007) allowing for third-country effects. We derive five determinants of foreign affiliate sales from this theory and we refer the reader to Bergstrand and Egger (2007) for further motivation of these. First, a larger size of home and host country raises affiliate sales with the former having a bigger impact. Second, the size of third countries has a negative impact on affiliate sales as long as the third country is bigger than the combined size of home and host country. Third, affiliate sales rise in transport costs and trade barriers, as the alternative way of serving the foreign market, exporting, becomes more expensive. Fourth, investment costs are negatively related to FAS. And fifth, countries with a regional trade agreement will display less bilateral affiliate sales.

Oldenski (2012) argues that both the complexity of tasks performed within firms and the importance of communication with customers are important industry-level determinants of affiliate sales relative to exports. First, we discuss the complexity of tasks carried out within firms. For more complex tasks cross-border communication within a firm between headquarters and foreign affiliate will be more costly. So, the hypothesis is that in industries with more complex tasks affiliate sales are less likely relative to exports. Second, Oldenski (2012) points out that not only communication within a firm is important, but also communication with outside customers. In industries with a strong need for direct contact with customers, like service sectors, there is a strong incentive for a foreign firm to be physically present and thus have an affiliate instead of selling services cross-border. Hence, the hypothesis is that affiliate sales are more likely in industries where communication with customers is important.

Keller and Yeaple (2013) explore how the impact of gravity forces on multinational activity varies with the knowledge intensity of sectors. Affiliates produce final goods using intermediate inputs and these inputs can be produced locally by the affiliate or they can be sourced from the headquarters. If the affiliates produce the intermediates themselves, knowledge transfers from the headquarters are required and knowledge moves in disembodied form. If the intermediates are sourced from the headquarters, knowledge moves in embodied form. In more complex or knowledge-intensive industries, knowledge is less codifiable and therefore more difficult to transfer. This makes local production more costly. As a result, more intermediates are sourced from the headquarters. As shipping intermediates goes along with trade costs, the gravity forces have a stronger impact in more complex sectors. The affiliate´s costs of production hence rise more with trade costs in complex sectors and therefore affiliate sales fall more heavily with trade costs. So the hypothesis is that foreign affiliate activity is decreasing with trade costs and the decrease is larger in more complex or knowledge-intensive industries.[4]

3 Data

3.1 Affiliate Sales Data

Most of the previous studies on multinational activity are based on data of affiliate sales of US parent companies or foreign affiliates located in the US, as pointed out in the introduction.

[4]With their model Keller and Yeaple also derive the hypothesis that the share of intermediates sourced from the headquarters falls more slowly with trade costs in more complex sectors. In our empirics we do not explore this hypothesis further, as it would require data on intra-firm trade.

In contrast, we employ the inward Foreign Affiliate Sales (FAS) data from Eurostat which include sector level affiliate sales (local sales and exports) by source country for many European countries. FAS data are a part of the European Foreign Affiliate Sales Statistics (FATS) which are collected yearly by Eurostat (Eurostat Metadata (2012)).

Inward FAS focus on majority owned foreign affiliates, so with the parent company having more than 50% of the voting rights, and describes the amount of sales, the number of workplaces, value of R&D and other characteristics generated by foreign investors from the particular source country in a given European host country. The FATS data are split in reference years 1996-2002, 2003-2007 and 2008-. In 2007 with EC-Regulation No. 716/2007 the FATS data collection became mandatory and therefore we concentrate our analysis on the data since 2008, also because almost 45% of the data before 2008 are missing observations. In the dataset since 2008 about 20% of the observations are missing. The missing observations consist both of not-reported values and confidential data. Data are considered confidential when direct or indirect identification of statistical units is possible due to too few reporting enterprises.

Table 1: Foreign affiliate sales observations

Type	**No. Observations**	**Share**
Missing	33 883	21.5%
Zero	101 118	64.3%
Positive	22 167	14.1%
Total	157 168	

Source: Eurostat inward FATS database, from 2008 onwards datasets

As table 1 shows 64% of the country-pair-sector combinations display zero affiliate sales. These are conventional zeros due to the fact that there is a relatively large number of sectors. There are 47 sectors in our sample, 38 source countries and 22 reporting host countries. Sectors and countries are listed in a table in the webappendix. In the webappendix we also analyze to what extent missings and zeros can be explained by source-country, host-country, or sectoral variation. It turns out that sectoral variation is the weakest determinant of the incidence of zeros in the data and explains only 6% of the variation in positive observations. This suggests that the zeros are not likely to create a big selection problem for our analysis on sectoral determinants. Still, we do account for biases as a result of zeros with the selection of our estimation methods.

3.2 Task Data

To construct measures for task complexity and the importance of communicating with customers we follow Oldenski (2012). Three measures from the American O*NET dataset are used to proxy task complexity: "Thinking Creatively"; "Making Decisions and Solving Problems"; and "Communicating with Supervisors, Peers, or Subordinates." To proxy the importance of communicating with customers we employ the task "Performing for or Working Directly with the Public". These indicators range from 0 to 100, 100 indicating high importance of the task for an occupation. The indicators are measured at the occupation level and henceforth have to be converted to the industry level. Following Oldenski (2012), we aggregate these scores to the industry level by employing data on the shares of occupations used in the industry from the US Bureau of Labor Statistics Occupational Employment Statistics. The importance of work activity a in an industry i, M_{ai} is defined as

$$M_{ai} = \sum_{c} \alpha_{ic} l_{ac}$$

c indicates occupation, α_{ic} is the share of occupation c employed in industry i and l_{ac} is the importance score of a task for an occupation. To get the share of a task in the total task inputs in an industry, I_{ai}, we can divide M_{ai} by the sum of importance scores for each task in the industry:

$$I_{ai} = \frac{M_{ai}}{\sum_{b} M_{bi}}$$

The tasks are matched to the sectors classified by NAICS activity classification. However, the affiliate sales data and other sectoral data are based on NACE Rev. 2. We match NAICS to NACE using correspondence tables provided by Eurostat. The detailed sector matching is provided in the webappendix.

As an alternative for the American complexity and communication measures we employ the OECD Skills survey data collected in 2010-2012. The survey covers 22 countries. Due to missing sector classifications we can only use the data from 16 countries. Because the number of observations is small in the same sector-country pairs and because we can only employ information from 16 countries, we use sectoral averages across countries. Since the task composition is relatively similar across countries in the data, this is a reasonable approach. Because of the nature of the data, survey data, we cannot extract information on how important

Table 2: Cross-correlation table

Variables	ThinkCrea	MakeDec	CommColl	Complex	CommCust	Sell
ThinkCrea	1.000					
MakeDec	0.639	1.000				
CommColl	0.367	0.260	1.000			
Complex	0.637	0.559	0.073	1.000		
CommCust	-0.081	-0.258	0.125	0.012	1.000	
Sell	-0.014	-0.206	0.025	0.096	0.771	1.000

Notes: ThinkCrea stands for thinking creatively, MakeDec - making decisions and solving problems, CommColl - communicating with supervisors, peers, or subordinates and CommCust - working with public. These measures based on O*NET data are percentage shares of tasks in total task input requirement of a given sector.
Complex and Sell, based on OECD data, are the mean of how often the complex tasks and selling are performed relative to the average across sectors.

complexity is in overall task composition. Instead we use how often complex tasks are performed in a certain sector.[5] To measure the importance of communication with customers we use the variable Selling, indicating how often employees have to sell.

Table 3: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
ThinkCrea	2.57	0.335	2.085	3.6
MakeDec	3.298	0.162	2.889	3.722
CommColl	3.578	0.136	3.347	3.961
Complex	1	0.4	0.356	2.159
CommCust	2.063	0.773	1.084	3.861
Sell	1	0.778	0.109	3.38

Number of observations: 153 032.

The survey outcomes range from 1 to 5, with 1 indicating that the task is never performed, 2 less than once a month, 3 less than once a week but at least once a month, 4 at least once a week but not every day and 5 every day. We rescale these numbers into the average number of times per day the task is performed. So 1 becomes 0, 2 becomes 1/90, 3 becomes 2/37, 4 becomes 1/4 and 5 becomes 2. In robustness checks we repeat the estimations with the original scaling and a different scaling where the arbitrary values, less than once a month and more than once a day get different values.

Table 2 shows the pairwise correlations of the task variables and table 3 displays summary statistics. The correlation between the different complexity measures from O*NET (ThinkCrea,

[5]The exact question asked is: how often are/were you usually confronted with more complex problems that take/took at least 30 minutes to find a good solution? The 30 minutes only refer to the time needed to THINK of a solution, not the time needed to carry it out.

MakeDec, and CommColl) and OECD (Complex) is fairly high and the same holds for the communication measures CommCust (O*NET) and Sell (OECD).

3.3 Other Data

In this section we outline the data sources of the explanatory variables. Table 4 provides an overview of the explanatory variables, their data sources and the summary statistics. To measure market size we use GDP, measured in current dollars, taken from the World Bank World Development Indicators. Third country market size is denoted by GDP RoW. It is calculated by subtracting the GDP of host and source countries from world GDP. Variation in this variable is small as the GDP of two single countries is small compared to world GDP for most countries.

Table 4: Summary statistics

Variable	**Source**	**Units**	**Mean**	**Std. Dev.**
Foreign affiliate sales	Eurostat	$ million	152	2 033
GDP, source	The World Bank	$ billion	1 170	2 550
GDP, host	The World Bank	$ billion	756	985
GDP RoW	The World Bank	$ billion	62 340	5 303
Distance	CEPII	Km	3 081	3 968
Common language	CEPII	0 or 1	0.04	0.20
RTA	De Sousa (2012)	0 or 1	0.81	0.39
FDI restrictiveness	OECD	0 or 1	0.02	0.09
Skill difference	ILOSTAT	Skill ratio	-0.015	0.23
R&D intensity	OECD STAN Database	R&D to value added ratio	0.043	0.083

Notes: Eurostat data (FAS) are in euros. We convert the values using the average yearly exchange rate from IMF Exchange Rate Archives by Month.

Trade costs are proxied by geographical distance between countries and by common language, both taken from CEPII (Mayer and Zignago (2011)). RTA is a dummy variable equal to one if a country pair has signed a bilateral regional trade agreement. RTAs include free trade agreements, customs unions, and Economic Integration Agreements. (WTO). The data on RTAs are from De Sousa (2012). The average for RTA is high, 0.81, since the majority of the country pairs is member of the EU.

FDI restrictiveness is an index that measures statutory restrictions on FDI in different sectors of host countries. It ranges from zero to one, where one indicates that FDI is fully restricted in a sector. This index takes into account foreign equity limits, screening and prior approval requirements at national level, restrictions on key foreign personnel, and other restrictions on the operation of foreign controlled entities. (Kalinova, et al. (2010)) The data on FDI restrictiveness are from OECD. We use the 2010 FDI restrictiveness index. The small mean values of 0.03 indicates low regulatory barriers on FDI. The most restricted sectors in the data are transportation, financial intermediation, and real estate activities.

Skill difference is defined as the skilled to unskilled labor ratio in the source country subtracted by the same ratio in the host country (Bergstrand and Egger (2007), Fukui and Lakatos (2012)). Data on skill composition are from the ILOSTAT database with skilled labor defined as managers, professionals, and technicians. The data is yearly. The mean of this variable is very close to zero, which reflects the similarity between average host and source countries. R&D-intensity is defined as R&D-expenditures over gross value added. The data are from the OECD STAN Database. Due to a broader sector classification used for the R&D-data and the fact that the data are not available for Bulgaria, Cyprus, Ireland, Lithuania, Luxemburg, Latvia, Malta, Sweden there are less observations in the estimations with R&D-intensity included.[6]

4 Estimation Method

Two features of the affiliate sales data guide us in our choice of the appropriate estimation method. First, a large fraction of bilateral affiliate sales is zero in our dataset and second, the data are log transformed. In the recent literature gravity equations of trade flows are estimated almost exclusively employing Poisson Pseudo Maximum Likelihood (PPML). This estimation method is favored because it can properly account for the log transformation of variables as suggested by Silva and Tenreyro (2006). PPML can also account for zeros, but the number of zeros is very large in our data. Therefore, we also present results with the zero-inflated negative binomial (ZINB) estimator. This estimator consists of a logit and a negative binomial component. It allows for two types of zeros, zeros in the binary process and zeros in the count process (Cameron and Trivedi (2010)). As described in the previous section our data also contain two types of zeros, missings due to confidentiality and zeros reflecting the absence of affiliate sales. Egger and Staub (2014) show that the negative binomial estimator performs best (together with PPML) in Monte Carlo studies on gravity estimation. Some scholars argue

[6]In these estimations we drop two sectors (Electricity, gas, steam and air conditioning supply; water supply; sewerage, waste management and remediation activities) due to a different classification in the OECD STAN Database.

that ZINB estimates are not invariant to changes in the scale of measurement of the dependent variable. Varying this scale the estimation results hardly changed.[7] In the estimations with PPML we dropped the missings and only retained real zeros, i.e. zeros reflecting the absence of affiliate sales between two countries in a sector. ZINB can account for two types of zeros, so there we retained both of them.

To control for common trends across countries and sectors, time fixed effects are included in every estimation. In the main analysis of the sectoral determinants of affiliate sales source-country-time and host-country-time fixed effects are included. In robustness checks presented in a webappendix we also include pairwise-time fixed effects. The webappendix also contains the results of the main regressions employing OLS estimation, since this has been the main estimation method in the literature on affiliate sales thus far.

5 Estimation Results

In this section we present our estimation results, starting each time with PPML estimates and then moving on to ZINB.[8] Columns 1 of tables 5 and 6 display the results of regressing affiliate sales on the established determinants of affiliate sales respectively with PPML and ZINB, omitting the host- and source-country-time fixed effects to be able to identify the effect of host- and source-country GDP. The subscript s indicates source country, r host country, i industry and t time. With both estimators both GDP in the home and host country have a large and very significant positive impact on affiliate sales. GDP in the rest of the world (ROW) is significantly and negatively related with affiliate sales, as expected from theory. Distance has a highly significant negative effect on affiliate sales, lending support for models of vertical FDI and the model of Keller and Yeaple (2013) and contradicting theories of horizontal FDI with a proximity-concentration tradeoff like Bergstrand and Egger (2007). Common language has a positive effect on affiliate sales although it is only significant with PPML. The negative and weakly significant coefficient on regional trade agreements (RTAs) is in line with the model of Bergstrand and Egger (2007). As expected the FDI restrictiveness index has a negative and highly significant impact on affiliate sales. Finally, the effect of skill-differences between home and host country is not robust across the two estimation methods: it is positive and

[7] As pointed out by Egger and Staub (2014) the negative binomial estimator is sensitive to scale in a two-step estimation approach and not with one-step estimation, as we are using.

[8] In the estimation of PPML the regressand is in levels, whereas the regressors are in logs, except for index variables and dummies which are also in levels.

Table 5: Estimation results of regressing affiliate sales on measures of task complexity and communication with customers employing PPML

	(1) FAS	(2) FAS	(3) FAS	(4) FAS	(5) FAS	(6) FAS
Ln(GDP_st)	1.080***					
	(28.01)					
Ln(GDP_rt)	0.793***					
	(30.25)					
Ln(GDP RoW_rst)	-4.054***	-43.11**	-40.72**	-40.40**	-40.48**	-33.97*
	(-5.28)	(-2.86)	(-2.69)	(-2.67)	(-2.66)	(-2.25)
Ln(Distance_rs)	-1.085***	-1.225***	-1.223***	-1.222***	-1.216***	-1.199***
	(-22.65)	(-13.32)	(-13.29)	(-13.29)	(-13.25)	(-13.48)
Comm Lang_rs	0.531***	0.354**	0.348**	0.349**	0.343**	0.349**
	(5.83)	(2.87)	(2.82)	(2.83)	(2.78)	(2.93)
RTA_rs	-0.219*	-1.553	-1.551	-1.551	-1.553	-1.560
	(-2.13)	(-1.66)	(-1.65)	(-1.66)	(-1.66)	(-1.66)
FDI Restrict_ir	-4.409***	-6.948***	-7.318***	-6.530***	-7.668***	-5.725***
	(-9.38)	(-14.46)	(-14.81)	(-13.46)	(-15.56)	(-13.37)
Skill Diff_rst	0.821***	25.11*	23.39*	23.15*	23.27*	18.70
	(5.17)	(2.30)	(2.13)	(2.10)	(2.11)	(1.70)
ThinkCrea			-0.393***			
			(-9.55)			
MakeDec				-1.552***		
				(-16.03)		
CommColl					-0.868***	
					(-9.23)	
Complex						-0.0798*
						(-2.49)
CommCust			0.207***	0.115***	0.245***	
			(8.53)	(5.06)	(10.01)	
Sell						0.714***
						(26.56)
Time dummies	Yes	No	No	No	No	No
Time varying country dummies	No	Yes	Yes	Yes	Yes	Yes
Observations	119884	119884	119884	119884	119884	119884
R^2	0.095	0.113	0.115	0.118	0.115	0.182

t statistics in parentheses

All estimations include a constant.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

significant with PPML and negative and significant with ZINB. In column 2 we include both source-country-time and host-country-time fixed effects and redo the regressions from column 1, omitting GDP. ROW-GDP, distance, common language and fdi-restrictiveness keep the same sign and remain significant or become significant in the case of common language. RTAs instead becomes significant and skill differences switches sign with ZINB from negative to positive.

5.1 Complexity and Communication with Customers

In columns 3 to 6 of tables 5 and 6 we move to the complexity and communication measures from O*NET and OECD. As expected from Oldenski's theory, the complexity measures thinking creatively, making decisions, communication with colleagues (all from O*NET) and complex tasks (from OECD) have a negative and significant impact on affiliate sales. Communication with customers (O*NET) and selling (OECD) display positive coefficients, also in line with theory. Still, in one of the specifications with the ZINB estimator communication with customers changes sign indicating that the effect of this variable is not robust across different specifications. Oldenski examines the effect of complexity and communication on affiliate sales relative to exports. Since we want to keep a uniform specification in testing Oldenski's and Keller and Yeaple's theories, we decided to work with affiliate sales alone. In the robustness checks we present the results on the effects of complexity and communication on affiliate sales relative to exports.

5.2 The Role of Services

In table 7 we analyse the robustness of the results on complexity and communication with customers by evaluating to what extent the results are driven by differences between services and manufacturing sectors. To do so, we include a dummy for service sectors. The table shows estimation results both with PPML and ZINB. We can draw three main conclusions from this analysis. First, the O*NET complexity measures are robust to inclusion of the service dummy, whereas the OECD measure is not. Affiliate sales are significantly smaller in more complex sectors as measured by O*NET also when including a services dummy. Instead, the coefficient on the OECD ComplexTasks becomes insignificant employing the ZINB estimator in column 8 and positive and significant with PPML. So the negative coefficient on this variable in the baseline results in tables 5 and 6 is partially the result of the distinction between manufactures and services.

Table 6: Estimation results of regressing affiliate sales on measures of task complexity and communication with customers employing ZINB

	(1) FAS	(2) FAS	(3) FAS	(4) FAS	(5) FAS	(6) FAS
FAS						
Ln(GDP_st)	0.382***					
	(21.16)					
Ln(GDP_rt)	0.637***					
	(35.75)					
Ln(GDP RoW_rst)	-3.716***	-46.65***	-41.82***	-40.27***	-42.70***	-46.32***
	(-7.22)	(-4.50)	(-4.24)	(-4.14)	(-4.23)	(-5.12)
Ln(Distance_rs)	-0.607***	-1.191***	-1.170***	-1.195***	-1.163***	-1.074***
	(-16.01)	(-21.81)	(-21.35)	(-21.97)	(-21.33)	(-21.35)
Comm Lang_rs	0.00137	0.203**	0.206**	0.242***	0.174*	0.292***
	(0.02)	(2.60)	(2.74)	(3.31)	(2.30)	(4.32)
RTA_rs	-0.406**	0.0704	0.0129	-0.261	-0.209	0.224
	(-2.82)	(0.07)	(0.01)	(-0.27)	(-0.21)	(0.21)
FDI Restrict_ir	-1.621***	-1.905***	-2.081***	-2.086***	-2.070***	-1.916***
	(-15.17)	(-22.52)	(-24.21)	(-23.31)	(-23.58)	(-22.05)
Skill Diff_rst	-0.383***	22.27**	19.03*	18.45*	20.30**	21.42**
	(-3.71)	(2.79)	(2.49)	(2.44)	(2.58)	(2.92)
ThinkCrea			-1.243***			
			(-25.21)			
MakeDec				-3.658***		
				(-29.56)		
CommColl					-2.565***	
					(-19.37)	
Complex						-0.340***
						(-7.75)
CommCust			0.181***	-0.184***	0.157***	
			(8.05)	(-7.15)	(6.92)	
Sell						0.564***
						(34.81)
Time dummies	Yes	No	No	No	No	No
Time varying country dummies	No	Yes	Yes	Yes	Yes	Yes
Observations	153032	153032	153032	153032	153032	153032
R^2						

t statistics in parentheses

All estimations include a constant. Standard errors are robust. In all estimations Inflate part of the regression includes the variables from the first specification and time dummies.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Estimation results of regressing affiliate sales on measures of task complexity and communication with customers including service dummy

	PPML				ZINB			
	(1) FAS	(2) FAS	(3) FAS	(4) FAS	(5) FAS	(6) FAS	(7) FAS	(8) FAS
main								
Ln(GDP RoW_rst)	-40.90**	-40.52**	-40.85**	-34.08*	-46.37***	-44.56***	-45.28***	-39.32***
	(-2.73)	(-2.70)	(-2.70)	(-2.24)	(-5.07)	(-4.87)	(-4.60)	(-4.74)
Ln(Distance_rs)	-1.223***	-1.225***	-1.214***	-1.198***	-1.181***	-1.210***	-1.170***	-1.069***
	(-13.32)	(-13.34)	(-13.26)	(-13.44)	(-22.16)	(-22.46)	(-21.80)	(-23.15)
Comm Lang_rs	0.356**	0.356**	0.349**	0.346**	0.271***	0.309***	0.203**	0.263***
	(2.90)	(2.91)	(2.85)	(2.89)	(3.72)	(4.31)	(2.73)	(4.18)
RTA_rs	-1.542	-1.541	-1.537	-1.582	0.197	-0.173	-0.149	0.116
	(-1.64)	(-1.64)	(-1.64)	(-1.68)	(0.21)	(-0.18)	(-0.15)	(0.11)
FDI Restrict_ir	-7.911***	-7.217***	-8.302***	-4.625***	-2.191***	-2.150***	-2.121***	-1.137***
	(-15.24)	(-13.21)	(-15.89)	(-11.10)	(-25.58)	(-22.99)	(-24.04)	(-13.40)
Skill Diff_rst	23.47*	23.14*	23.47*	18.86	21.55**	21.01**	21.83**	17.14*
	(2.15)	(2.12)	(2.14)	(1.71)	(3.00)	(2.90)	(2.82)	(2.51)
ThinkCrea	-1.024***				-1.791***			
	(-19.15)				(-27.38)			
MakeDec		-2.680***				-4.573***		
		(-17.58)				(-27.69)		
CommColl			-1.452***				-2.958***	
			(-14.00)				(-20.26)	
Complex				0.117**				0.0556
				(3.02)				(1.40)
CommCust	-0.351***	-0.478***	-0.0598		-0.216***	-0.618***	-0.0539	
	(-7.75)	(-9.56)	(-1.92)		(-5.90)	(-13.63)	(-1.47)	
Sell				0.918***				1.059***
				(41.47)				(54.38)
Service Dummy	1.144***	1.087***	0.710***	-0.617***	0.787***	0.720***	0.413***	-1.363***
	(11.89)	(11.19)	(9.37)	(-11.78)	(13.02)	(12.23)	(7.17)	(-36.60)
Time varying country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	119884	119884	119884	119884	153032	153032	153032	153032
R^2	0.130	0.136	0.126	0.175				

t statistics in parentheses

All estimations include a constant. Standard errors are robust. In all estimations of ZINB model Inflate part of the regression includes the variables from the first specification of Table 6, respective complexity measure, service dummy, and time dummies.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Second, the effect of communication with customers as measured by O*NET seems to be driven by the distinction between services and manufactures, since the sign of the coefficient on CommCust switches from positive to negative in all specifications and is significant in two of the three specifications. Additional regression results in the webappendix running the regressions separately for manufacturing and services sectors indicate that this is the case within both of these broad sectors. So, within services and within manufacturing sectors where communication with customers as measured by O*NET is larger display less affiliate sales. Instead the coefficient on the OECD measure Selling remains negative and significant indicating that also within services and manufactures, sectors where Selling is more important are characterized by more affiliate sales.

Third, the services dummy is positive and highly significant when including the O*NET measures in the regression. Instead, if we include the OECD measures the sign of the services dummy coefficient reverses and becomes significantly negative. This result reflects that the OECD-variable Selling is picking up the distinction between services and manufactures: conditioning on the variable Selling the services dummy is significantly negative.

5.3 The Gravity of Knowledge

In tables 8 and 9 we present estimation results on the interaction effect between distance and both complexity and R&D-intensity. We use one complexity measure from O*NET, thinking creatively, and the complexity measure from OECD. We always include source-country-time and host-country-time fixed effects. In columns 1, 3 and 5 we do not include complexity and R&D intensity as a separate regressor, following the approach in Keller and Yeaple (2013). In the other columns we do. In columns 1, 3 and 5 we find strong support for the findings in Keller and Yeaple (2013): the negative effect of distance on affiliate sales is bigger in more complex sectors and more R&D-intensive sectors. The picture changes completely when including complexity and R&D-intensity as separate regressors. Thinking creatively in column 2, complex tasks in column 4 and R&D-intensity in column 6 all have a negative and significant impact on affiliate sales, but the interaction terms with distance display positive and significant coefficients.[9] In the webappendix we find the same results when including sector fixed effects instead of the complexity measures as separate regressors. The interaction term changes from being negative and significant to positive and significant.

[9] In the webappendix we show that the findings are robust to the use of alternative measures of complexity

Table 8: Estimation results of regressing affiliate sales on the interaction of distance with R&D-intensity and task complexity employing PPML

	(1) FAS	(2) FAS	(3) FAS	(4) FAS	(5) FAS	(6) FAS
Ln(GDP RoW_rst)	-42.39**	-42.71**	-42.83**	-43.10**	-13.65	-11.67
	(-2.82)	(-2.84)	(-2.84)	(-2.86)	(-0.73)	(-0.63)
Ln(Distance_rs)	-1.082***	-1.533***	-1.209***	-1.328***	-0.858***	-0.909***
	(-11.59)	(-12.69)	(-13.16)	(-13.18)	(-6.10)	(-6.36)
Comm Lang_rs	0.350**	0.351**	0.353**	0.352**	0.652***	0.662***
	(2.84)	(2.84)	(2.87)	(2.86)	(3.53)	(3.59)
RTA_rs	-1.566	-1.562	-1.559	-1.552	-0.313	-0.345
	(-1.67)	(-1.67)	(-1.66)	(-1.66)	(-0.55)	(-0.60)
FDI Restrict_ir	-6.726***	-6.725***	-6.811***	-6.794***	-3.829***	-3.950***
	(-14.31)	(-14.46)	(-14.32)	(-14.32)	(-6.21)	(-6.49)
Skill Diff_rst	24.65*	24.88*	24.93*	25.12*	9.259	3.005
	(2.25)	(2.28)	(2.28)	(2.30)	(0.14)	(0.05)
ThinkCrea		-1.343***				
		(-5.05)				
Ln(Dist)*ThinkCrea	-0.0555***	0.122**				
	(-10.12)	(3.28)				
Complex				-0.941***		
				(-4.29)		
Ln(Dist)*Complex			-0.0155***	0.108***		
			(-3.36)	(3.65)		
R&D intensity						-16.67***
						(-3.71)
Ln(Dist)*R&D int					-0.495***	1.653**
					(-4.55)	(2.81)
Time varying country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	119884	119884	119884	119884	33803	33803
R^2	0.113	0.113	0.113	0.113	0.140	0.143

t statistics in parentheses

All estimations include a constant.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These results seem to indicate that the interaction terms of complexity and R&D-intensity with distance without including these variables as separate regressors are picking up the independent effect of complexity and R&D-intensity described in Oldenski's theory: in more complex and R&D-intensive sectors firms avoid affiliate sales, because communication is difficult. Keller and Yeaple (2013) do not include complexity as a separate regressor in any of their robustness checks. They argue that in more complex sectors distance has a more discouraging effect on affiliate sales, since most intermediates will be sourced from the home country driving up production costs. Our results do not support this theory. Distance and trade costs have a less discouraging effect on affiliate sales in more complex sectors.

Before pointing out a possible way to reconcile our findings with those in Keller and Yeaple (2013), we discuss two other differences between our estimations and the estimations in Keller and Yeaple (2013) that could explain the different findings. First, Keller and Yeaple (2013) use measures of trade costs employing data on fob-cif margins and tariffs and we use distance. The work on gravity in international trade suggests that distance is a prime determinant of trade costs. So, if Keller and Yeaple's theory would hold for trade costs as measured by fob-cif margins it should also hold for distance. Second, Keller and Yeaple (2013) use firm-level data, whereas we are working with sectoral data. This means that we are merging the intensive and extensive margin effect. It is easy to show that the hypothesis in Keller and Yeaple (2013) also holds for affiliate sales at the aggregate (sectoral) level. At the aggregate level the effect becomes even stronger, because the intensive and extensive margin effect work in the same direction. The intensive margin effect described in Keller and Yeaple (2013) states that firms in more complex sectors will import more from the headquarters and therefore trade costs will have a bigger impact on marginal costs and thus on sales. A similar logic holds for the extensive margin. In more complex sectors the costs of affiliate sales rise more with distance and henceforth the negative impact of distance on the presence of affiliates will be stronger. In the webappendix we derive these claims formally employing the analysis on aggregate variables in Keller and Yeaple (2013).

So we need a different way to reconcile our empirical results with Keller and Yeaple (2013). These authors only model the decision of firms between sales through local affiliates (horizontal FDI) and sales through exporting. In our data instead, output of affiliates might also be shipped back to the source country and affiliates activity is then part of value-chain production taking place in multiple countries to exploit differences in factor costs (vertical FDI). Since our data

Table 9: Estimation results of regressing affiliate sales on the interaction of distance with R&D-intensity and task complexity employing ZINB

	(1)	(2)	(3)	(4)	(5)	(6)
	FAS	FAS	FAS	FAS	FAS	FAS
FAS						
Ln(GDP RoW_rst)	-41.02***	-44.07***	-44.73***	-46.12***	-46.76***	-39.18**
	(-4.02)	(-4.30)	(-4.38)	(-4.45)	(-3.76)	(-3.16)
Ln(Distance_rs)	-0.774***	-1.592***	-1.147***	-1.313***	-0.997***	-1.055***
	(-13.81)	(-12.01)	(-20.87)	(-18.36)	(-12.78)	(-13.41)
Comm Lang_rs	0.212**	0.199*	0.209**	0.206**	0.0580	0.0816
	(2.72)	(2.57)	(2.69)	(2.64)	(0.53)	(0.76)
RTA_rs	-0.0420	0.00117	0.0874	0.0692	-0.244	-0.000660
	(-0.05)	(0.00)	(0.09)	(0.07)	(-0.49)	(-0.00)
FDI Restrict_ir	-1.933***	-1.941***	-1.917***	-1.920***	-2.200***	-2.224***
	(-22.96)	(-23.27)	(-22.59)	(-22.71)	(-17.33)	(-17.54)
Skill Diff_rst	18.58*	20.79**	20.88**	21.98**	121.4**	91.22*
	(2.37)	(2.64)	(2.64)	(2.75)	(2.75)	(2.08)
ThinkCrea		-2.345***				
		(-7.08)				
Ln(Dist)*ThinkCrea	-0.164***	0.153***				
	(-23.10)	(3.30)				
Complex				-1.258***		
				(-3.56)		
Ln(Dist)*Complex			-0.0532***	0.122*		
			(-7.88)	(2.46)		
R&D intensity						-10.69***
						(-6.04)
Ln(Dist)*R&D int					-0.450***	0.994***
					(-13.24)	(4.22)
Time varying country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	153032	153032	153032	153032	44437	44437
R^2						

t statistics in parentheses

All estimations include a constant. Standard errors are robust. In all estimations Inflate part of the regression includes the variables from the first specification of Table 6, respective interaction term (and complexity measure), and time dummies.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

are on all affiliate sales (in local markets, to third markets and back to the source country) they also contain vertical affiliate sales. With a composition effect between horizontal and vertical affiliate sales, both horizontal and vertical affiliate sales individually can display a negative interaction effect of distance and complexity, whereas the overall interaction effect is positive.

Suppose we allow for both affiliate sales in the local market and affiliate sales back home. We define the former as HAS, since it reflects horizontal FDI and we define the latter as VAS, since it reflects vertical FDI. So total affiliate sales AS are given by:

$$AS(\tau, \phi) = HAS(\tau, \phi) + VAS(\tau, \phi) \tag{1}$$

Keller and Yeaple (2013) only allow for market-seeking horizontal FDI. To model vertical FDI we would have to include both multiple stages of production and differences in factor costs across countries and factor intensities across the stages of production. We do not model such a setup, but limit ourselves to showing that the interaction effect of trade costs and complexity on total affiliate sales could be positive, although the impact on both horizontal and vertical affiliate sales separately is negative. This can happen if two conditions are satisfied. First, the elasticity of horizontal affiliate sales with respect to trade costs is less negative (or even positive) than the elasticity of vertical affiliate sales with respect to distance and second, the share of horizontal affiliate sales should rise with the complexity of the sector. Before discussing the reasonableness of these two conditions, we first show formally that these two conditions are sufficient to generate a positive interaction effect of distance and complexity on total affiliate sales. Log differentiating equation (1) with respect to trade costs τ generates the elasticity of affiliate sales with respect to trade costs τ:

$$\frac{\partial \ln AS}{\partial \ln \tau} = s_H \frac{\partial \ln HAS}{\partial \ln \tau} + s_V \frac{\partial \ln VAS}{\partial \ln \tau} \tag{2}$$

s_H is the share of HAS in AS, $s_H = \frac{HAS}{AS}$ and $s_V = 1 - s_H$.

Differentiating equation (2) with respect to complexity ϕ gives:

$$\begin{aligned}\frac{\partial^2 \ln AS}{\partial \ln \tau \partial \ln \phi} &= s_H \frac{\partial^2 \ln HAS}{\partial \ln \tau \partial \ln \phi} + s_V \frac{\partial^2 \ln VAS}{\partial \ln \tau \partial \ln \phi} + \frac{\partial s_H}{\partial \ln \phi}\frac{\partial \ln VAS}{\partial \ln \tau} + \frac{\partial s_V}{\partial \ln \phi}\frac{\partial \ln VAS}{\partial \ln \tau} \\ &= s_H \underset{(-)}{\frac{\partial^2 \ln HAS}{\partial \ln \tau \partial \ln \phi}} + s_V \underset{(-)}{\frac{\partial^2 \ln VAS}{\partial \ln \tau \partial \ln \phi}} + \underset{(+)}{\frac{\partial s_H}{\partial \ln \phi}} \underset{(+)}{\left(\frac{\partial \ln HAS}{\partial \ln \tau} - \frac{\partial \ln VAS}{\partial \ln \tau}\right)}\end{aligned}$$

In the second line we have used $\frac{\partial s_V}{\partial \ln \phi} = -\frac{\partial s_H}{\partial \ln \phi}$.

So, even if the interaction terms for VAS and HAS separately are negative, the overall effect could be positive if the elasticity of distance with respect to HAS is less negative (or positive) than the elasticity with respect to VAS and if the share of horizontal affiliate sales rises with complexity. This would generate a composition effect: in more complex sectors the share of horizontal affiliate sales is larger and horizontal affiliate sales display a less negative (or positive) trade cost elasticity.

A negative trade cost elasticity of vertical sales and a positive trade cost elasticity of horizontal sales seem reasonable: vertical sales require the to-and-forth shipping of intermediates, whereas it is the goal of horizontal affiliate sales to avoid trade costs. A rising share of horizontal affiliate sales in the complexity of the sector is also reasonable: vertical affiliate sales require more coordination between the different stages of production and will thus be less prevalent in more complex sectors.

5.4 Robustness

We conduct three sets of robustness checks of which the results are presented in the webappendix. First, we redo the main analysis of tables (5)-(9) using OLS. We find similar results on the effect of source and host GDP, third country GDP, distance and FDI restrictiveness as with PPML and ZINB. Also the analysis on the interaction effects of distance with complexity and R&D-intensity leads to the same results. The coefficients are negative without including these variables as separate regressors and the effects are positive when we do include them. The effect of complexity is negative and significant as in the main analysis. The effect of communication with customers as measured by O*NET and employing OLS is at odds with the results employing PPML and ZINB. We selected these methods for a reason, i.e. to confront heteroskedasticity and to account for the large number of zeros. Therefore, we do not interpret these estimation results as contradicting Oldenski's hypothesis on the effect of communication with customers on affiliate sales.

Second, we estimate the main specifications on the sectoral determinants including time-varying pairwise fixed effects. Including these fixed effects enables us to control for pairwise unobserved heterogeneity. The estimation results provide support for our main analysis: the complexity variables all have a negative and significant coefficient and the communication with customers variables a positive coefficient. The interaction terms of distance with complexity and

R&D-intensity have a negative effect without including these variables as separate regressors, but they become positive and significant when including the variables separately. This confirms our estimation results in the previous section.

Third, we examine the effect of the complexity and communication measures on exports relative to affiliate sales. The results show that the O*NET complexity measures have the expected positive sign, although only one of them is significant. The OECD complexity measure instead has a negative sign, at odds with the hypotheses.[10] The communication measures have an expected negative and significant sign.

6 Concluding Remarks

We have analysed the sectoral determinants of affiliate sales put forward in recent models of multinational activity employing European data. The Eurostat data on affiliate sales between 2008 and 2011 and thus since their collection is compulsory have not been used before in the literature. We found broad support for the theories in Oldenski (2012) that affiliate sales are smaller in sectors with a higher degree of task complexity and larger in sectors where communication with customers is important, although the effect of communication was not robust in all specifications. The empirical results were at odds with the hypothesis in Keller and Yeaple (2013) that the negative effect of distance on affiliate sales is stronger in more complex sectors, as measured by R&D-intensity or task complexity. Once including these variables as separate regressors, the coefficients of the interaction terms turned from negative into positive. We pointed out that our results could be reconciled with Keller and Yeaple's findings if our results on sector-level affiliate sales are driven by a composition effect between horizontal and vertical affiliate sales. Our findings call for a formal theory on the varying effect of trade costs on affiliate sales in industries with different levels of complexity. Our proposed explanation to reconcile our findings with the ones in Keller and Yeaple contains useful building stones to set up such a theory.

[10]Since we are evaluating the effect on trade relative to affiliate sales, the expected signs of the coefficients are opposite to the ones in the main text on the size of affiliate sales.

References

Antras, P., & Yeaple, S. R. (2013). Multinational Firms and the Structure of International Trade. NBER Working Paper No. 18775.

Bergstrand, J. H., & Egger, P. (2007). A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises. Journal of International Economics, 73(2), 278-308.

Bureau of Labor Statistics (2007 and 2010). Occupational Employment Statistics [National industry-specific and by ownership]. Retrieved from http://www.bls.gov/oes/tables.htm

Burger, M., Van Oort, F. V., & Linders, G. J. (2009). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-Inflated Estimation. Spatial Economic Analysis, 4(2), 167-190.

Cameron, A. C., & Trivedi, P. K. (2010). Microeconometrics Using Stata (Revised ed.). College Station, Texas: Stata Press.

Chen, M., X. (2009). Regional Economic Integration and Geographic Concentration of Multinational Firms. European Economic Review 53, 355-375.

De Benedictis, L., & Taglioni, D. (2011). The Gravity Model in International Trade. Chapter 4 of The Trade Impact of the European Union Preferential Policies: an Analysis through Gravity Models. Ed. Luca De Benedictis and Luca Salvatici.

De Sousa, J. (2012). The currency union effect on trade is decreasing over time. Economics Letters 117(3), 917-920. Retrieved from http://jdesousa.univ.free.fr/data.htm

Egger, Peter H. and Kevin E. Staub (2014). GLM estimation of trade gravity models with fixed effects. Mimeo ETH Zurich.

Eurostat metadata (2012). Foreign controlled EU enterprises - inward FATS. Retrieved from http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/fats_esms.htm#accuracy

Eurostat (2013a). Foreign controlled EU enterprises - inward FATS [Foreign control of enterprises by economic activity and a selection of controlling countries (2003-2007 and from 2008 onwards) (fats_g1a_03 and fats_g1a_08)]. Retrieved from http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing

Eurostat (2013). Index of Correspondence Tables [NAICS 2002 - NACE Rev. 1.1 and NAICS 2007 - NACE Rev. 2]. Retrieved from http://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL

Franko A. (2007). NACE: How the revision of a classification can affect quality of statistics. Eurostat: NACE - European Conference on Quality 2008.

Fraser Institute (2012). Economic Freedom of the World [2012 dataset]. Retrieved from http://www.freetheworld.com/release.html

Fukui, T., & Lakatos, C. (2012). A Global Database of Foreign Affiliate Sales. Office of Economics Working Paper, No. 2012-08A.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. Quarterly Journal of Economics 123, 441-487.

International Monetary Fund (2013). Exchange Rate Archives by Month. Retrieved from http://www.imf.org/external/np/fin/data/param_rms_mth.aspx

Kalinova, B., Palerm, A., & Thomsen, S. (2010). OECD's FDI Restrictiveness Index. OECD Working Papers on International Investment 2010-03.

Keller, W. &. Yeaple, S. R. (2013). The Gravity of Knowledge. American Economic Review 103(4),1414-44.

Kleinert, J., & Toubal, F. (2010). Gravity for FDI. Review of International Economics, 18(1), 1-13.

LABORSTA Internet (2013). Statistics by topic [Total employment, by occupation]. Retrieved from http://laborsta.ilo.org/

Markusen, J. R. (2002). Multinational Firms and the Theory of International Trade. Boulder: The MIT Press.

Martin, Will and Cong S. Pham (2008). Estimating the Gravity Model When Zero Trade Flows Are Frequent. Mimeo World Bank.

Mayer, T., & Zignago, S. (2011), Notes on CEPII's distances measures (GeoDist), CEPII Working Paper 2011-25.

O*NET Resource Center (2012). O*NET 17.0 Database. Retrieved from http://www.onetcenter.org/database.html?p=2

OECD (2013). FDI Regulatory Restrictiveness Index [Data by measure and year]. Retrieved from http://www.oecd.org/investment/fdiindex.htm

OECD (2013). Technical Report of the Survey of Adult Skills (PIAAC).

OECD (2014). Survey of Adult Skills (PIAAC). Retrieved from http://www.oecd.org/site/piaac/publicdataandanalysis.htm

Oldenski, L. (2012). Export versus FDI and the Communication of Complex Information. Journal of International Economics, 87(2), 312-322.

Ramondo, N., Rappoport, V., & Ruhl, K. J. (2011). Horizontal vs. Vertical FDI: Revisiting Evidence from U.S. Multinationals. Working Papers 11-12, New York University, Leonard N. Stern School of Business, Department of Economics.

Silva, J. S., & Tenreyro, S. (2006). The Log of Gravity. The Review of Economics and Statistics, 88(4), 641-658.

The World Bank (2013). World Development Indicators [Economic Policy & External Dept]. Retrieved from http://data.worldbank.org/data-catalog/world-development-indicators

WTO (2013). Regional Trade Agreements: Scope of RTAs. Retrieved from http://www.wto.org/english/tratop_e/region_e/scope_rta_e.htm

Yeaple, S. (2009). Firm Heterogeneity and the Structure of U.S. Multinational Activity: An Empirical Analysis. Journal of International Economics 78(2), 206-215.