Model uncertainty and variable selection: an application to the modelization of FDI determinants in Europe

DOCTORAL DISSERTATION

Doctorado en Economía Internacional y Turismo

By:

Laura Montolio Breva

Supervisors:

Dr. Mariam Camarero Olivas

Dr. Cecilio Tamarit Escalona

September 2019

A mis padres

Agradecimientos

Quisiera dedicar estas líneas a expresar mi agradecimiento a todas aquellas personas e instituciones que, de alguna manera, han contribuido a la realización de esta Tesis doctoral.

En primer lugar, a mis directores de Tesis, Mariam Camarero Olivas y Cecilio Tamarit Escalona,por depositar en mí su confianza y acogerme en el grupo de investigación *Joint Research Unit in Economic Integration* (INTECO). Su profesionalidad, constante disponibilidad y paciencia han sido indispensables para la consecución de esta Tesis doctoral. Gracias por vuestro apoyo y dedicación durante todo este tiempo.

Dar las gracias también al Ministerio de Economía y Competitividad de España y, en particular a los proyectos ECO2014-58991-C3-2-R y ECO2017-83255-C3-3-P por financiar la realización de esta Tesis doctoral a través de un contrato predoctoral para la Formación del Personal Investigador (FPI) (Referencia BES-2015-072395). A la Fundación Banco Sabadell por la financiación otorgada para la realización del capítulo 3 a través de una de las Ayudas Predoctorales a la Investigación 2017.

Agradecer especialmente a las personas que nos han facilitado los recursos y datos necesarios para la elaboración de la presente Tesis doctoral. Al Centro de Proceso de Datos de la Universidad de Valencia por poner a nuestra disposición el superordenador LluisVives v2 con el cual se han realizado parte de los cálculos del capítulo 3. A Matilde Mas Ivars del Instituto Valenciano de Investigaciones Económicas (Ivie) por proporcionarnos los datos de capital acumulado en infraestructuras públicas por regiones utilizados en el capítulo 2. A

Isaac Barbero y Emilio Carmona del Ministerio de Economía y Competitividad de España por proporcionarnos la versión extendida y desagregada a nivel regional de los datos de stock de IED de *DataInvex* que nos ha permitido llevar a cabo el análisis regional en el capítulo 2. A Astrit Sulstarova de la UNCTAD por proporcionarnos la versión extendida de los datos de stock de IED de la base de datos de estadísticas bilaterales de IED de la UNCTAD utilizados en los capítulos 3 y 4.

Asimismo, me gustaría expresar mi agradecimiento a Inmaculada Martínez Zarzoso por su hospitalidad, amabilidad y los consejos que me ha brindado durante mi estancia en la University of Göttingen (Alemania). De igual modo, expresar mi gratitud a Stephan Klasen, Felicitas Nowak-Lehmann y demás profesores del departamento de economía de la University of Göttingen, así como a los compañeros del grupo de investigación *"Globalization and Development"* (GlaD) por la cálida acogida y por todos los medios que han puesto a mi alcance. A Yunzhi, por compartir esta experiencia conmigo.

De igual modo, me gustaría agradecer a los compañeros del Departamento de Estructura Económica de la Universidad de Valencia por acogerme en un entorno amigable e intelectualmente enriquecedor. Agradecer también a los compañeros del pasillo E, con los que he tenido la suerte de compartir el día a día, y especialmente a Mercedes Beltrán, por su amabilidad y apoyo ofrecido cuando lo he necesitado.

Asimismo, me gustaría extender mi agradecimiento al Instituto de Economía Internacional. En especial, a Gloria Berenguer Contrí no solo por su profesionalidad y dedicación al programa de doctorado, sino por su comprensión y apoyo en momentos difíciles. Dar las gracias también a todo el personal administrativo y en particular, a Carmen Planells por su implicación y calidad humana.

Quisiera expresar asimismo mi gratitud a todas las personas cuya amistad, compañerismo y apoyo han sido indispendables a lo largo de estos cuatro años. A los compañeros de doctorado (Adriana, Carlos, Danny, Eli, Ernesto, Fede, Jesús, Jorge, Jose, Luis, Mafi, Marta, Marta S., Mauricio, Paula y Rubén): gracias por todos los buenos momentos compartidos. A Mariola y a Adrián, por las innumerables ocasiones en que me han ofrecido su apoyo: gracias por brindarme vuestra amistad y llenar esta etapa de momentos inolvidables.

A mis amigas, por acompañarme en ésta y otras muchas aventuras: Alba, Andrea, Carol, Eva, Julia, Julia N., María, Mireia, Noelia y Patricia.

Finalmente, quisiera expresar mi agradecimiento a mi familia sin cuyo cariño y apoyo esta Tesis no hubiera sido posible. A mis padres, Ana y Desi, por inculcarme el valor del esfuerzo, la superación y la generosidad. A mi hermana Silvia por acompañarme y guiarme con su ejemplo en todo momento. A Edu, por creer en mí y brindarme su apoyo incondicional en éste y otros muchos aspectos de mi vida.

A todos, gracias

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The growth of Foreign direct investment (FDI) throughout the world economy has accelerated since 1990 and become a key driver of globalization. Due to the removal of barriers and the rise in international exchanges, more and more firms have decided to reorganize and reallocate their activities. Nonetheless, FDI has been highly unevenly spread, with the bulk of foreign investments concentrated in developed countries, both in outward and inward FDI terms. In fact, developed countries accounted for around 78% of total world outward FDI stock in 2015, while developing economies hold just the 21%. Nevertheless, investments held by developing economies have increased notably in recent years, reflecting a relocation of FDI from developed to developing countries offering more competitive conditions. This growth is partly explained by the successful integration of China into the international production networks. The global pattern of foreign direct investment stocks and its evolution over time is presented in Tables 1.1 and 1.2.

FDI from and to European Union member countries has seen a sharp rise associated with a second wave of dynamic economic integration triggered by the creation of the Single Market. The EU accounted for 43% of world outward FDI stocks in 1990, a share that has steadily increased with the acceleration of the EU economic integration process since the launching of the euro. The growth of intra-EU investments gave rise to increasing and persistent external imbalances. The elimination of the exchange rate risk combined with optimistic expectations of peripheral countries' convergence to the core resulted in a credit boom in the periphery, leading to a large-scale reallocation of capital from the core to peripheral EU countries. In this respect, Spain became the largest capital importer, as opposed to Germany, which has been the main investor. Thereby, because the pattern of investment differs across countries, increasing attention has been paid to the question of why FDI goes to some countries and not to others.

Table 1.1 Global Outward Foreign Direct Investment Stocks

(*current value, $ billion*)

	1985	1990	1995	2000	2005	2010	2015
Developed							
USA	386,35	731,76	1.363,79	2.694,01	3.638	4.809,59	6.007,77
Japan	43,97	201,44	238,45	278,44	386,58	831,08	1.228,77
EU-28	310,95	975,78	1.720,70	2.907,12	5.058,43	9.136,66	9.377,76
Germany	59,91	308,74	505,69	483,95	794,20	1.364,56	1.349,85
Spain	4,46	15,65	35,03	129,19	305,43	653,24	492,50
Total	**818,58**	**2.113,95**	**3.677,67**	**6.699,29**	**10.558,27**	**17.554,73**	**19.530,25**
Developing							
Latin America	48,24	52,05	68,06	53,53	206,78	457,19	713,98
Asia	23,08	67,06	211,48	597,07	947,92	2.465,54	4.681,63
China	0,9	4,455	17,77	27,77	57,21	317,21	1.097,86
Africa	11,87	21,25	32,60	39,88	44,18	134,35	238,81
Total	**83,23**	**140,40**	**312,37**	**690,73**	**1.199,61**	**3.059,93**	**5.644,30**
World	**901.81**	**2.254,90**	**3.993,68**	**7.409,63**	**11.900,46**	**20.981,76**	**25.514,31**

Source: UNCTAD.

Table 1.2 Global Inward Foreign Direct Investment Stocks

(*current value, $ billion*)

	1985	1990	1995	2000	2005	2010	2015
Developed							
USA	220	539,60	1.005,73	2.783,24	2.817,97	3.422,29	5.709,66
Japan	4,74	9,85	33,53	50,32	100,90	214,88	174,15
EU-28	277,04	883,89	1.329,45	2.322,12	4.363,35	7.357,41	7.933
Germany	36,93	226,55	312,58	470,94	640,06	955,88	775,68
Spain	8,94	65,92	105,72	156,35	384,54	628,34	543,88
Total	**616,24**	**1.685,88**	**2.711,01**	**5.782,41**	**8.524,26**	**13.480,30**	**16.384,15**
Developing							
Latin America	66,50	107,19	179,92	338,77	738,64	1.629,25	1.884,95
Asia	259,89	339,67	571,68	1.052,67	1.629,11	3.881,16	6.020,75
China	6,06	20,69	101,10	193,35	272,09	587,82	1.220,90
Africa	42,91	59,99	88,63	152,80	282,04	598,29	747,91
Total	**370,37**	**508,80**	**842,66**	**1.546,08**	**2.653,17**	**6.123,09**	**8.677,87**
World	**986,61**	**2.196,33**	**3.564,64**	**7.380,45**	**11.427,73**	**20.279,39**	**25.664,96**

Source: UNCTAD.

FDI has important economic effects on both host and source countries. First, from the host country perspective, FDI brings new technologies leading to positive spillovers to local firms and creates new job opportunities, boosting domestic demand and enhancing growth. Similarly, the source country benefits from access to foreign markets in the sense of spreading the cost of production across more consumers as well as technology transfers.

In view of the beneficial effects of FDI, it is important for countries to understand the factors underlying investors' decisions across different countries or regions. Traditionally, researchers have relied on the gravity equation when modelling the cross-country patterns of FDI. The gravity equation, originally derived by Tinbergen in 1962 as an analogy of Newton's Law of Gravity, states that international flows (such as trade or FDI) among two countries will be directly proportional to their economic sizes and inversely proportional to distance (in the sense of trade frictions or investment costs). Initially, the gravity model for

FDI lacked a theoretical foundation and it was frequently applied by resemblance to trade international flows. It was not until recently that economists, notably Bergstrand and Egger (2007) and Head and Ries (2008), derived a general equilibrium theory for FDI bilateral flows.

Nowadays, while the theoretical foundation of the FDI gravity model is well-established, there exists considerable uncertainty on its empirical application. Even though the literature devoted to analyze the determinants of FDI is vast, the results are inconclusive due to the variety of model specifications and estimation methods applied by researchers. The conventional practice in the literature has been to consider and ad hoc set of regresors associated with different FDI theories. Nevertheless, this approach has been proved to provide uncertain or even contradictory results (Eicher et al. (2012)). Particularly, ignoring uncertainty in the model specification might lead to misleading estimates due to the exclusion of relevant variables or the inclusion of irrelevant variables (Blonigen and Piger (2014)).

In the same vein, researchers have applied a variety of different estimation techniques either for the additive or the mutltiplicative functional form estimation of the gravity model. The seminal work of Silva and Tenreyro (2006) argued that the gravity model should be estimated in its multiplicative form by means of non-linear estimators and in particular, using the Poisson Pseudo Maximum Likelihood (PPML) estimator. Despite the PPML has been frequently applied in the literature, recent contributions have suggested that alternative estimators appear to be more appropriate than the PPML (see Martínez-Zarzoso (2013), Gómez-Herrera (2013) and Egger and Staub (2016), among others). In this context, the literature on the determinants of FDI faces a model uncertainty problem that involves two main challenges. On the one hand, there is a need for robust statistical techniques that allow to select the relevant set of potential FDI determinants. On the other hand, alternative estimators should be compared in order to identify the proper specification of the model.

In this dissertation, we aim to contribute to the debate by addressing the problem of model uncertainty in the modelization of FDI in order to provide a broader understanding of FDI determinants. We focus in the European Union and, in particular, in the cases of Spain and Germany, as they represent one of the largest FDI recipient and investor countries within the EU, respectively. The Thesis is divided into three chapters. In chapter 2, we examine the long-run inward FDI determinants in Spanish regions and implement an exploratory factor analysis (EFA) to address the collinearity problem that arises when including a large share of potential FDI determinants in the model. Then, in chapter 3, we analyze the long-run determinants of German outward FDI and apply a Bayesian Model Averaging (BMA) analysis to deal with the variable selection problem. Lastly, in chapter 4, we confront the uncertainty in the model specification of German outward FDI by comparing the performance of alternative estimators for the FDI gravity model.

This dissertation presents a number of contributions to the literature. First, the three chapters included in this dissertation rely on FDI stock data as an alternative to flows as we hypothesize that FDI stock data better approximates the long-run allocation patterns of FDI. To the best of our knowledge, we provide one of the first attempts on examining the determinants of inward FDI in Spain from a regional standpoint by using FDI stock data instead of flows. Second, from a methodological point of view, we implement several techniques to deal with model uncertainty in FDI that allow us to uncover the relevant FDI determinants in the European Union. We argue that, when there is uncertainty about the relevant set of explanatory variables to be included in a regression model, the model selection problem should be explicitly addressed prior to the estimation of the model. Our approach to deal with model uncertainty stems from recent contributions in Bayesian Variable selection problems, namely, the BMA approach developed in the R package BayesVarSel (García-Donato and Forte (2015)). Our results point towards a parsimonous FDI specification, as opposed to the conventional practice in the literature. Finally, we provide a comprehensive understanding of the German outward FDI determinants by taking into account the heterogeneity of the recipient countries and comparing several alternative estimators.

In what follows, we describe the logical steps followed in the research that lead to this dissertation. Following this introduction, chapter 2 focuses on the modelization of FDI from a regional perspective. To this end, we consider that Spain provides a very interesting case study. As highlighted at the beginning of this introduction, within peripheral EU countries, Spain attracted most of the FDI that arose since the introduction of the euro. Nevertheless, the allocation of FDI differs across regions. Even though a regional analysis is an invaluable tool for regional governments aiming at designing policies to attract investments, the empirical evidence at the regional level is limited, primary due to the lack of data. For our analysis, *DataInvex* provides FDI data on FDI stocks disaggregated by regions. We are grateful to the Spanish Ministry of Economy for providing us with the extended version of FDI stock data. An important drawback of our data is that the sectoral breakdown of FDI stock data is not available and hence, we are not able to conduct a more disaggregated analysis.

In this respect, we estimate a gravity equation to model the FDI distribution in Spanish regions from 40 source countries over the 2004-2013 period. We consider as regressors a set of variables highlighted by the literature as potential FDI determinants and show that they suffer from collinearity problems. Consequently, we propose to implement an exploratory factor analysis (EFA), pioneered in this strand of literature by Villaverde and Maza (2012). The basic idea behind EFA is to reduce the number of initial variables to a manageable set of non-collinear factors. Specifically, we identified three factors: *Economic or Market potential*, *Productive capacity* and *Competitiveness and agglomeration effects*. Once the appropiate set of variables is identified, we estimate our model using the Poisson Pseudo-Maximum Likelihood (PPML) estimator including as regressors the three extracted factors. Our results highlight the key role of *Competitiveness and agglomeration effects* in attracting FDI to the Spanish regions, suggesting efficiency-seeking FDI motivations (i.e. vertical

FDI). Furthermore, we find that inward FDI to Spanish regions is highly determined by the degree of industrial specialization and the geograhical location of the regions. This analysis is intended to serve as an empirical evidence to assist local and regional policy makers to implement measures in order to further strengthen FDI attractiveness to a region.

In the context of external imbalances in Europe, Germany plays an important role as previously mentioned. Thereby, chapter 3 addresses the variable selection problem in FDI determinants for the particular case of Germany. As located at the core of the EU, Germany has traditionally attracted a large share of inward FDI. Nonetheless, the motivation for the case study of Germany lies in the fact that it is considered one of the major investors throughout the world. In this respect, the focus of our study will be to examine the pattern of outward FDI. Our primary objective in this chapter is twofold. First, because an ad hoc selection of regressors with respect to certain FDI theories might negatively affect subsequent inferences, we implement a Bayesian Model Averaging (BMA) approach to deal with the intrinsic uncertainty in FDI determinants.

Standard statistical practice ignores model uncertainty. Data analysts typically select a model and proceed as if the selected model had generated the data. This approach, however, ignores the uncertainty in model selection, leading to over-confident inferences. Accordingly, BMA provides a coherent mechanism to overcome this problem (Hoeting et al. (1999)). In the statistical literature, early work related to model averaging includes Roberts (1965) or Leamer (1978). Although BMA is an intuitively attractive solution to the problem of model uncertainty, it is not yet part of the standard data analysis tool kit. To some extent, this is due to the fact that implementation of BMA presents several difficulties. In this respect, the BMA approach was basically neglected in economic applications until the late 1990s and 2000s, when the "BMA revolution" in economics took place. In recent times, several methods for implementing BMA have emerged. This is due to the availability of more powerful computers and new developements in numerical methods such as Markov chain Monte Carlo model composition that allow applied researchers to overcome the problems encountered when implementing BMA in the context of large model spaces.

The state of BMA research in economics can be found in Raftery (1995) and Fernández et al. (2001). Particularly, empirical growth economics is, without any doubt, the most active field in which model averaging techniques have been applied from the beginning. The seminal papers on model averaging and growth are Fernández et al. (2001) and Sala-i Martin et al. (2004). More recently, model averaging has also been successfully considered in applications in macroeconomics, finance or health economics.

In the context of the determinants of FDI, model uncertainty was first put forward by Chakrabarti (2001) by means of Extreme Bounds Analysis (EBA). Nevertheless, only more recently, with the contribution of Blonigen and Piger (2014), model uncertainty has been recognised as a major econometric problem. The pioneering work of Blonigen and Piger (2014) examines the determinants of FDI using a cross sectional data set of US outward FDI stock and recommend to follow a Bayesian Model Averaging (BMA) approach to address model uncertainty. Their research influenced the study of Eicher et al. (2012) who extends the BMA approach to solve for selection bias in the presence of zero FDI observations using a panel data of bilateral FDI flows.

Our study is closely linked to Camarero et al. (2015) that applies to the analysis of the nexus between growth and energy consumption the BMA approach developed in the R package BayesVarSel by García-Donato and Forte (2015). This package provides a very flexible framework under different priors that makes it especially suitable for the empirical economic analysis. Based on the developments of Bayarri et al. (2012), among others, this approach allows us to choose the best model among all possible configurations within our data set by computing the posterior distributions for each parameter. An important aspect of this methodology is that it provides summaries of the posterior distributions, namely the inclusion probabilities, which are helpful when there is a large set of potential explanatory variables.

Furthermore, we are able to compute average estimated coefficient of the parameters that provide us with a hint of the direction and effect of the identified FDI determinants.

Second, given that the recipient countries of German FDI are very heterogenous, we propose to conduct the analysis across different country-groups. In doing so, we avoid the so-called aggregation bias that would arise by the omission of relevant information regarding the diversity of motives for FDI in different regions or country-groups (see, Mitchell et al. (2011)). For this purpose, we assembled a panel dataset that includes outward FDI stock data from Germany to 59 recipient countries - 38 developed and 21 developing - over the period 1996-2012. We are grateful to the UNCTAD for providing us with the extended version of FDI stock from the UNCTAD's Bilateral FDI Statistics database. As regressors we consider 61 potential FDI determinants according to the theoretical and empirical literature.

Our results suggest that a proper FDI specification is more parsimonous than previously suggesed in the literature. Only a small share of the potential FDI determinants appear to be robust in the sense of displaying a high posterior inclusion probability. Furthermore, we find that the decision to invest abroad involves a mixture of FDI motivations that differ across country-groups. For developed countries, determinants linked to market-seeking or HFDI are found to be relatively more important; whereas those related to VFDI prevail in developing countries. A further disaggregation of developing countries into Latin American and Asian countries reveals that both HFDI and VFDI strategies coexist together with institutional factors. Nonetheles, market-seeking FDI appears to prevail in Asian countries; whereas efficiency-seeking FDI plays a key role in Latin American countries. Additionally, we find that market access is the main strategy for FDI in "core" EU countries, while vertical multinational enterprises (MNEs) are dominant in peripheral EU countries. In summary, our analysis proves to be a useful approach to account for model uncertainty and provides policy makers with evidence of the factors that should be strengthen in order to attract German investments.

The last chapter of this dissertation aims to select the estimation method of the parameters of the FDI determinants previously identified in chapter 3. In this respect, we address the uncertainty surrounding the specification of the gravity model by comparing several alternative Generalized Linear Model (GLM) estimators. A GLM framework allows us to tackle the main econometric issues that arise in the estimation of the gravity model. First, GLMs estimate the gravity model in its original multiplicative functional form and thereby, avoid the bias caused when estimating the log-linear or additive functional form of the gravity model. Second, GLM estimators fully account for zero-valued bilateral FDI observations, as the dependent variable is included in levels. Specifically, we compare the most frequently used GLM estimator, the Poisson Pseudo Maximum Likelihood (PPML) estimator, with alternative estimators recently proposed by the literature to outperform the PPML for certain applications; the Gamma Pseudo Maximum Likelihood (GPML) and the Negative Binomal Pseudo Maximum Likelihood (NBPML) estimators. We include also the Gaussian GLM for comparative purposes. Our model selection approach is based on several goodness-of-fit statistics and graphical techniques that allow us to select the best performing estimator for our data set. Our results appear to suggest that NBPML is the best performing estimator for this application, followed by GPML. Furthermore, the estimated parameters confirmed the results provided by the BMA analysis conducted in chapter 3.

Once the appropiate estimator is considered, our findings provides further evidence on the determinants of German outward FDI across country-groups. In particular, our results highlight the key role of countries' participation in GVCs together with institutional factors for German FDI directed towards developed countries. The latter findings are in line with modern literature that stresses the positive impact of institutional quality on FDI location through the promotion of a good investment climate, which faces protection of property rights, political stability and less corruption (see Wheeler and Mody (1992), Wei (2000), Globerman and Shapiro (2002), Daude and Stein (2007) or Bénassy-Quéré et al. (2007), among others).

In this respect, the institutional approach has highlighted the role of institutional factors as relevant FDI determinants together with the variables previously posited by other theoretical approaches. Furthermore, FDI appears to be particularly affected by the EU integration process. On the other hand, we find that German FDI into developing countries is mainly driven by labour factor endowments availability, market access and institutional factors. Concretely, market-seeking FDI prevails in Asian countries; whereas efficiency-seeking FDI is the dominant in Latin American economies. As regards German investment in its European neighbours, we find that the market-seeking motive prevails for FDI in "core" EU countries, whereas investment in peripheral countries is linked to the increasing fragmentation of production and the countries' integration into the global value chains network. Overall, these findings provide policy makers with a thorough comprehension of the factors that should be emphasized to attract German FDI.

Introducción

El crecimiento de la Inversión Extranjera Directa (IED) en la economía mundial se ha acelerado desde 1990 y se ha convertido en un motor clave de la globalización. Debido a la eliminación de barreras y al aumento de los intercambios internacionales, cada vez más empresas han decidido reorganizar y reasignar sus actividades. No obstante, la IED se ha extendido de manera muy desigual, con la mayor parte de las inversiones extranjeras concentradas en los países desarrollados, tanto en términos de IED saliente como entrante. De hecho, los países desarrollados representaron alrededor del 78% del stock total de IED saliente mundial en 2015, mientras que las economías en desarrollo representaron solo el 21%. Sin embargo, las inversiones en las economías en desarrollo han aumentado notablemente en los últimos años, lo que refleja una reubicación de la IED desde los países desarrollados a los países en desarrollo que ofrecen condiciones más competitivas. Este crecimiento se explica en parte por la integración exitosa de China en las redes internacionales de producción. El patrón global de stock de inversión extranjera directa y su evolución a lo largo del tiempo se presenta en las Tablas 1.3 and 1.4.

La IED desde y hacia los países miembros de la Unión Europea (UE) ha experimentado un fuerte aumento asociado con una segunda ola de integración económica dinámica desencadenada por la creación del Mercado Único. La UE representó el 43% del stock mundial de IED en 1990, una proporción que ha aumentado constantemente con la aceleración del proceso de integración económica de la UE desde el lanzamiento del euro. El crecimiento de las inversiones intra-UE dio lugar a crecientes y persistentes desequilibrios externos. La eliminación del riesgo cambiario combinado con expectativas optimistas de convergencia de los países periféricos a los países centrales de la UE resultó en un boom crediticio en la periferia, conduciendo a una reasignación de capital a gran escala desde los países centrales a los países periféricos. A este respecto, España se convirtió en el mayor importador de capital, a diferencia de Alemania, que ha sido el principal inversor. De este modo, debido a que el patrón de inversión difiere entre países, se ha prestado cada vez más atención a la pregunta de por qué la IED se dirige a algunos países y no a otros.

Table 1.3 Stock global de Inversión Extranjera Directa saliente

(*Billones de dólares corrientes*)

	1985	1990	1995	2000	2005	2010	2015
Desarrollados							
EEUU	386,35	731,76	1.363,79	2.694,01	3.638	4.809,59	6.007,77
Japón	43,97	201,44	238,45	278,44	386,58	831,08	1.228,77
UE-28	310,95	975,78	1.720,70	2.907,12	5.058,43	9.136,66	9.377,76
Alemania	59,91	308,74	505,69	483,95	794,20	1.364,56	1.349,85
España	4,46	15,65	35,03	129,19	305,43	653,24	492,50
Total	**818,58**	**2.113,95**	**3.677,67**	**6.699,29**	**10.558,27**	**17.554,73**	**19.530,25**
En desarrollo							
América Latina	48,24	52,05	68,06	53,53	206,78	457,19	713,98
Asia	23,08	67,06	211,48	597,07	947,92	2.465,54	4.681,63
China	0,9	4,455	17,77	27,77	57,21	317,21	1.097,86
África	11,87	21,25	32,60	39,88	44,18	134,35	238,81
Total	**83,23**	**140,40**	**312,37**	**690,73**	**1.199,61**	**3.059,93**	**5.644,30**
Mundo	**901.81**	**2.254,90**	**3.993,68**	**7.409,63**	**11.900,46**	**20.981,76**	**25.514,31**

Fuente: UNCTAD.

Table 1.4 Stock global de Inversión Extranjera Directa entrante

(*Billones de dólares corrientes*)

	1985	1990	1995	2000	2005	2010	2015
Desarrollados							
EEUU	220	539,60	1.005,73	2.783,24	2.817,97	3.422,29	5.709,66
Japón	4,74	9,85	33,53	50,32	100,90	214,88	174,15
EU-28	277,04	883,89	1.329,45	2.322,12	4.363,35	7.357,41	7.933
Alemania	36,93	226,55	312,58	470,94	640,06	955,88	775,68
España	8,94	65,92	105,72	156,35	384,54	628,34	543,88
Total	**616,24**	**1.685,88**	**2.711,01**	**5.782,41**	**8.524,26**	**13.480,30**	**16.384,15**
En desarrollo							
América Latina	66,50	107,19	179,92	338,77	738,64	1.629,25	1.884,95
Asia	259,89	339,67	571,68	1.052,67	1.629,11	3.881,16	6.020,75
China	6,06	20,69	101,10	193,35	272,09	587,82	1.220,90
África	42,91	59,99	88,63	152,80	282,04	598,29	747,91
Total	**370,37**	**508,80**	**842,66**	**1.546,08**	**2.653,17**	**6.123,09**	**8.677,87**
Mundo	**986,61**	**2.196,33**	**3.564,64**	**7.380,45**	**11.427,73**	**20.279,39**	**25.664,96**

Fuente: UNCTAD.

La IED tiene importantes efectos económicos tanto en los países receptores como en los de origen. En primer lugar, desde la perspectiva del país receptor, la IED trae nuevas tecnologías que conllevan *spillovers* positivos para las empresas locales y crea nuevas oportunidades laborales, aumentando la demanda interna y mejorando el crecimiento. De manera similar, el país de origen se beneficia de acceso a mercados extranjeros en el sentido de distribuir el coste de producción entre más consumidores, así como de transferencias de tecnología.

En vista de los efectos beneficiosos de la IED, es importante para los países entender los factores subyacentes de las decisiones de los inversores en diferentes países o regiones. Tradicionalmente, los investigadores se han basado en la ecuación de gravedad al modelar los patrones de IED entre países. La ecuación de gravedad, originalmente derivada por Tinbergen en 1962 como una analogía de la Ley de Gravedad de Newton, establece que los flujos internacionales (como el comercio o la IED) entre dos países serán directamente proporcionales a sus tamaños económicos e inversamente proporcionales a la distancia (en el sentido de fricciones comerciales o costes de inversión). Inicialmente, el modelo de gravedad para la IED carecía de una base teórica y con frecuencia se aplicaba por semejanza a los flujos comerciales internacionales. Solo recientemente los economistas, principalmente Bergstrand and Egger (2007) y Head and Ries (2008), han derivado una teoría de equilibrio general para los flujos bilaterales de IED.

Actualmente, a pesar de que los fundamentos teóricos del modelo de gravedad de la IED están bien establecidos, existe una incertidumbre considerable sobre su aplicación empírica. Aunque la literatura dedicada a analizar los determinantes de la IED es amplia, los resultados no son concluyentes debido a la variedad de especificaciones del modelo y métodos de estimación utilizados por los investigadores. La práctica convencional en la literatura ha sido considerar un conjunto ad hoc de regresores asociados con diferentes teorías de la IED. Sin embargo, se ha demostrado que este enfoque proporciona resultados inciertos o incluso contradictorios (Eicher et al. (2012)). En particular, ignorar la incertidumbre en la especificación del modelo podría conducir a estimaciones engañosas debido a la exclusión de variables relevantes o la inclusión de variables irrelevantes (Blonigen and Piger (2014)).

En la misma línea, los investigadores han aplicado una variedad de técnicas de estimación diferentes ya sea para la estimación de la forma funcional aditiva o multipliciva del modelo de gravedad. El trabajo seminal de Silva and Tenreyro (2006) argumentó que el modelo de gravedad debería estimarse en su forma multiplicativa por medio de estimadores no lineales y, en particular, utilizando el estimador de Pseudo Máxima Verosimilitud de Poisson (*Poisson Pseudo Maximum Likelihood*; PPML). A pesar de que PPML se ha aplicado con frecuencia en la literatura, contribuciones recientes han sugerido que estimadores alternativos parecen ser más apropiados que PPML (ver Martínez-Zarzoso (2013), Gómez-Herrera (2013) y Egger and Staub (2016), entre otros). En este contexto, la literatura sobre los determinantes de la IED enfrenta un problema de incertidumbre del modelo que conlleva dos desafíos principales.

Por un lado, existe la necesidad de técnicas estadísticas robustas que permitan seleccionar el conjunto relevante de determinantes potenciales de la IED. Por otro lado, se deben comparar estimadores alternativos para identificar la especificación adecuada del modelo.

En esta disertación, nuestro objetivo es contribuir al debate abordando el problema de incertidumbre del modelo en la modelización de la IED con el fin de proporcionar una comprensión más amplia de los determinantes de la IED. Nos centramos en la Unión Europea y, en particular, en los casos de España y Alemania, ya que representan uno de los mayores países receptores e inversores de IED dentro de la UE, respectivamente. La presente Tesis doctoral se divide en tres capítulos. En el capítulo 2, examinamos los determinantes a largo plazo de la IED entrante en las regiones españolas e implementamos un análisis factorial exploratorio (*exploratory factor analysis*; EFA) para abordar el problema de colinealidad que surge cuando se incluye una gran proporción de determinantes potenciales de la IED en el modelo. Después, en el capítulo 3, analizamos los determinantes a largo plazo de la IED saliente de Alemania y aplicamos un análisis de Promedio Bayesiano de Modelos (*Bayesian Model Averaging*; BMA) para tratar el problema de selección de variables. Por último, en el capítulo 4, afrontamos la incertidumbre en la especificación del modelo de IED saliente de Alemania comparando el desempeño de estimadores alternativos para el modelo de gravedad de la IED.

Esta disertación presenta una serie de contribuciones a la literatura. Primero, los tres capítulos incluidos en esta disertación se basan en datos de stock de IED como una alternativa a los flujos, ya que proponemos que los datos de stock de IED aproximan mejor los patrones de distribución de la IED a largo plazo. Hasta donde sabemos, proporcionamos uno de los primeros intentos de examinar los determinantes de la IED entrante en España desde un punto de vista regional mediante el uso de datos de stock de IED en lugar de flujos. En segundo lugar, desde un punto de vista metodológico, implementamos varias técnicas para abordar la incertidumbre del modelo en la IED que nos permiten descubrir los determinantes relevantes de la IED en la Unión Europea. Argumentamos que, cuando existe incertidumbre sobre el conjunto de variables explicativas relevantes que se incluirán en un modelo de regresión, el problema de selección del modelo debe abordarse explícitamente antes de la estimación. Nuestro enfoque para lidiar con la incertidumbre del modelo proviene de contribuciones recientes en problemas bayesianos de selección de variables, concretamente, el enfoque BMA desarrollado en el paquete de R BayesVarSel (García-Donato and Forte (2015)). Nuestros resultados apuntan hacia una especificación parsimoniosa de la IED, en oposición a la práctica convencional en la literatura. Finalmente, proporcionamos una comprensión integral de los determinantes de la IED saliente de Alemania teniendo en cuenta la heterogeneidad de los países receptores y comparando varios estimadores alternativos.

A continuación, describimos los pasos lógicos seguidos en la investigación que condujeron a esta disertación. Tras esta introducción, el capítulo 2 se centra en la modelización de la IED desde una perspectiva regional. Con este fin, consideramos que España proporciona un caso de estudio muy interesante. Como se destacó al comienzo de esta introducción, dentro de los países periféricos de la UE, España atrajo la mayor parte de la IED que surgió desde la introducción del euro. Sin embargo, la distribución de la IED difiere entre las regiones. Si bien un análisis regional es una herramienta invaluable para los gobiernos regionales con el objetivo de diseñar políticas para atraer inversiones, la evidencia empírica a nivel regional es limitada, principalmente debido a la falta de datos. Para nuestro análisis, *DataInvex* proporciona datos de stock de IED desagregados por regiones. Agradecer al Ministerio de Economía de España por proporcionarnos la versión extendida de los datos de stock de IED. Un inconveniente importante de nuestros datos es que el desglose sectorial de los datos de stock de IED no está disponible y, por lo tanto, no podemos realizar un análisis más desagregado.

A este respecto, estimamos una ecuación de gravedad para modelar la distribución de la IED en las regiones españolas procedente de 40 países de origen durante el período 2004-

2013. Consideramos como regresores un conjunto de variables resaltadas por la literatura como posibles determinantes de la IED y mostramos que sufren problemas de colinealidad. En consecuencia, proponemos implementar un análisis factorial exploratorio, utilizado por primera vez en esta rama de la literatura por Villaverde and Maza (2012). La idea básica detrás de EFA es reducir el número de variables iniciales a un conjunto manejable de factores no colineales. Concretamente, identificamos tres factores: *Potencial económico o de mercado*, *Capacidad productiva* y *Competitividad y efectos de aglomeración*. Una vez identificado el conjunto de variables adecuadas, estimamos nuestro modelo usando el estimador PPML incluyendo como regresores los tres factores extraídos. Nuestros resultados destacan el papel clave del factor *Competitividad y efectos de aglomeración* en la atracción de IED a las regiones españolas, lo que sugiere inversiones motivadas por la búsqueda de eficiencia (i.e., IED vertical). Además, encontramos que la IED entrante en las regiones españolas está altamente determinada por el grado de especialización industrial y la ubicación geográfica de las regiones. El objetivo de este análisis es proporcionar evidencia empírica que sirva de soporte a los responsables políticos locales y regionales para la implementación de medidas dirigidas a fortalecer el atractivo de una región para la IED.

En el contexto de los desequilibrios externos en Europa, Alemania juega un papel importante como se mencionó anteriormente. De este modo, el capítulo 3 aborda el problema de selección de variables en los determinantes de la IED para el caso particular de Alemania. Localizado en el núcleo central de la UE, Alemania ha atraído tradicionalmente una gran parte de la IED entrante. No obstante, la motivación para el caso de estudio de Alemania radica en el hecho de que es considerado uno de los principales inversores en todo el mundo. A este respecto, el enfoque de nuestro estudio será examinar el patrón de la IED saliente. Nuestro objetivo principal en este capítulo es doble. En primer lugar, debido a que una selección ad hoc de regresores con respecto a ciertas teorías de IED podría afectar negativamente a las inferencias posteriores, implementamos un enfoque de Promedio Bayesiano de Modelos para abordar la incertidumbre intrínseca en los determinantes de la IED.

La práctica estándar de la estadística ignora la incertidumbre del modelo. Los analistas de datos generalmente seleccionan un modelo y proceden como si el modelo elegido hubiera generado los datos. Este enfoque, sin embargo, ignora la incertidumbre en la selección del modelo, dando lugar a inferencias muy confiadas. Por consiguiente, BMA proporciona un mecanismo coherente para superar este problema (Hoeting et al. (1999)). En la literatura estadística, los primeros trabajos relacionados con el promedio de modelos incluyen Roberts (1965) o Leamer (1978). Aunque BMA es una solución intuitivamente atractiva para el problema de incertidumbre del modelo, todavía no forma parte del kit de herramientas de análisis de datos estándar. Hasta cierto punto, esto se debe al hecho de que la implementación de BMA presenta varias dificultades. A este respecto, el enfoque BMA fue fundamentalmente ignorado en las aplicaciones económicas hasta finales de los años noventa y comienzos de la década del 2000, cuando tuvo lugar la "revolución BMA" en economía. Recientemente, han surgido varios métodos para implementar BMA. Esto se debe a la disponibilidad de ordenadores más potentes y nuevos desarrollos en métodos numéricos como la aproximación *Markov chain Monte Carlo model composition (MC^3)* que permite a los investigadores aplicados solventar los problemas encontrados al implementar BMA cuando el número de modelos a considerar es elevado.

El estado de la investigación sobre BMA en economía se puede encontrar en Raftery (1995) y Fernández et al. (2001). Particularmente, la literatura empírica sobre crecimiento económico es, sin lugar a dudas, el campo de investigación principal en el que las técnicas de promediación de modelos se han aplicado desde el inicio. Los trabajos más influyentes sobre promedio de modelos y crecimiento son Fernández et al. (2001) y Sala-i Martin et al. (2004). Más recientemente, el promedio de modelos también se ha considerado con éxito en diversas aplicaciones en macroeconomía, finanzas o economía de la salud.

En el contexto de los determinantes de la IED, Chakrabarti (2001) puso de manifiesto por primera vez la incertidumbre del modelo mediante el análisis de límites extremos (*Extreme Bounds Analysis*; EBA). Sin embargo, solo más recientemente, con la contribución de Blonigen and Piger (2014), se ha reconocido la incertidumbre del modelo como un gran problema econométrico. El trabajo pionero de Blonigen and Piger (2014) examina los determinantes de la IED utilizando un conjunto de datos de sección cruzada de stock de IED saliente de EEUU y recomienda seguir un enfoque BMA para abordar la incertidumbre del modelo. Su investigación influyó el estudio de Eicher et al. (2012) que amplía el enfoque BMA para resolver el sesgo de selección en presencia de observaciones cero en la IED utilizando datos de panel de flujos de IED bilaterales.

Nuestro estudio está estrechamente relacionado con Camarero et al. (2015) que aplica el enfoque BMA desarrollado en el paquete R BayesVarSel por García-Donato and Forte (2015) al análisis del nexo entre crecimiento y consumo de energía. Este paquete proporciona un marco muy flexible bajo diferentes probabilidades a priori (*priors*) que lo hacen especialmente adecuado para el análisis empírico en economía. Basado en los desarrollos de Bayarri et al. (2012), entre otros, este enfoque nos permite elegir el mejor modelo entre todas las posibles combinaciones en nuestra base de datos calculando las distribuciones a posteriori para cada parámetro. Un aspecto importante de esta metodología es que proporciona resúmenes de las distribuciones a posteriori, a saber, las probabilidades de inclusión, útiles cuando el conjunto de variables explicativas potenciales es grande. Además, podemos calcular el coeficiente estimado promedio de los parámetros que nos proporciona información de la dirección y el efecto de los determinantes de la IED identificados.

En segundo lugar, dado que los países receptores de IED de Alemania son muy heterogéneos, proponemos realizar el análisis por diferentes grupos de países. Al hacerlo, evitamos el llamado sesgo de agregación que surgiría por la omisión de información relevante con respecto a la diversidad de motivos para la IED en diferentes regiones o grupos de países (ver, Mitchell et al. (2011)). Para este propósito, creamos una base de datos de panel que incluye datos de stock de IED saliente de Alemania en 59 países receptores, 38 desarrollados y 21 en desarrollo, durante el período 1996-2012. Agradecer a la UNCTAD por proporcionarnos la versión extendida de stock de IED de la base de datos de estadísticas bilaterales de IED de la UNCTAD. Como regresores consideramos 61 determinantes potenciales de la IED de acuerdo con la literatura teórica y empírica.

Nuestros resultados sugieren que una especificación adecuada de la IED es más parsimoniosa que la sugerida previamente en la literatura. Únicamente una pequeña proporción de los posibles determinantes de la IED parece ser robusta en el sentido de mostrar una alta probabilidad de inclusión a posteriori. Además, consideramos que la decisión de invertir en el extranjero implica una mezcla de motivaciones de IED que difieren entre grupos de países. Para los países desarrollados, determinantes asociados a la búsqueda de mercados o IED horizontal son relativamente más importantes; mientras que aquellos relacionados con la IED vertical prevalecen en los países en desarrollo. Una mayor desagregación de los países en desarrollo en países latinoamericanos y asiáticos revela que tanto las estrategias de IED horizontal como vertical coexisten junto con factores institucionales. No obstante, la IED en búsqueda de mercados parece prevalecer en los países asiáticos; mientras que la IED en búsqueda de eficiencia juega un papel clave en los países latinoamericanos. Además, encontramos que el acceso a mercados es la estrategia principal para la IED en los países centrales de la UE, mientras que las empresas multinacionales (EMNs) integradas verticalmente predominan en los países periféricos de la UE. En resumen, nuestro análisis es un enfoque útil para tener en cuenta la incertidumbre del modelo y proporcionar a los responsables de formular políticas evidencia de los factores que deben fortalecerse para atraer inversiones alemanas.

El último capítulo de esta disertación tiene como objetivo seleccionar el método de estimación de los parámetros de los determinantes de la IED previamente identificados en el capítulo 3. A este respecto, abordamos la incertidumbre asociada a la especificación del modelo de gravedad comparando varios estimadores alternativos en el marco de Modelos Lineales Generalizados (*Generalized Linear Models*; GLMs). Un marco GLM nos permite abordar los principales problemas econométricos que surgen en la estimación del modelo de gravedad. En primer lugar, GLMs estiman el modelo de gravedad en su forma funcional multiplicativa original y, por lo tanto, evitan el sesgo causado al estimar la forma funcional log-lineal o aditiva del modelo de gravedad. En segundo lugar, los estimadores GLM incorporan las observaciones de IED bilaterales que toman valor cero, ya que la variable dependiente se incluye en niveles. Específicamente, comparamos el estimador GLM más frecuentemente utilizado, el estimador PPML, con estimadores alternativos recientemente propuestos por la literatura como estimadores más adecuados que PPML para ciertas aplicaciones; el estimador de Pseudo Máxima Verosimilitud de Gamma (*Gamma Pseudo Maximum Likelihood*; GPML) y el estimador de Pseudo Máxima Verosimilitud de la Binomial Negativa (*Negative Binomal Pseudo Maximum Likelihood*; NBPML). Incluimos también GLM con distribución gaussiana (*Gaussian GLM*) con fines comparativos. Nuestro enfoque de selección de modelos se basa en varios estadísticos de bondad de ajuste y técnicas gráficas que nos permiten seleccionar el estimador con mejor desempeño para nuestra base de datos. Nuestros resultados parecen sugerir que NBPML es el mejor estimador para esta aplicación, seguido de GPML. Además, los parámetros estimados confirman los resultados proporcionados por el análisis BMA realizado en el capítulo 3.

Una vez que se considera el estimador apropiado, nuestros resultados proporcionan evidencia adicional sobre los determinantes de la IED saliente de Alemania por grupos de países. En particular, nuestros resultados destacan el papel clave de la participación de los países en las cadenas globales de valor junto con factores institucionales para la IED alemana dirigida a los países desarrollados. Los últimos resultados están en línea con la literatura moderna que enfatiza el impacto positivo de la calidad de las instituciones en la localización de la IED a través de la promoción de un buen clima de inversión, caracterizado por protección de los derechos de propiedad, estabilidad política y menos corrupción (ver Wheeler and Mody (1992), Wei (2000), Globerman and Shapiro (2002), Daude and Stein (2007) or Bénassy-Quéré et al. (2007), entre otros). A este respecto, el "enfoque institucional" ha destacado el papel de los factores institucionales como determinantes relevantes de la IED junto con las variables previamente planteadas por otros enfoques teóricos. Además, la IED parece verse particularmente afectada por el proceso de integración de la UE. Por otro lado, encontramos que la IED alemana en los países en desarrollo está impulsada principalmente por la disponibilidad de recursos laborales, el acceso a mercados y los factores institucionales. Concretamente, la IED en búsqueda de mercados prevalece en los países asiáticos; mientras que la IED en búsqueda de eficiencia es la dominante en las economías latinoamericanas. Con respecto a la inversión de Alemania en sus vecinos europeos, encontramos que el motivo de búsqueda de mercados prevalece para la IED en los países centrales de la UE, mientras que la inversión en los países periféricos está vinculada a la creciente fragmentación de la producción y la integración de los países en las cadenas globales de valor. En general, estos resultados proporcionan a los responsables de formular políticas una comprensión exhaustiva de los factores que deben enfatizarse para atraer la IED de Alemania.

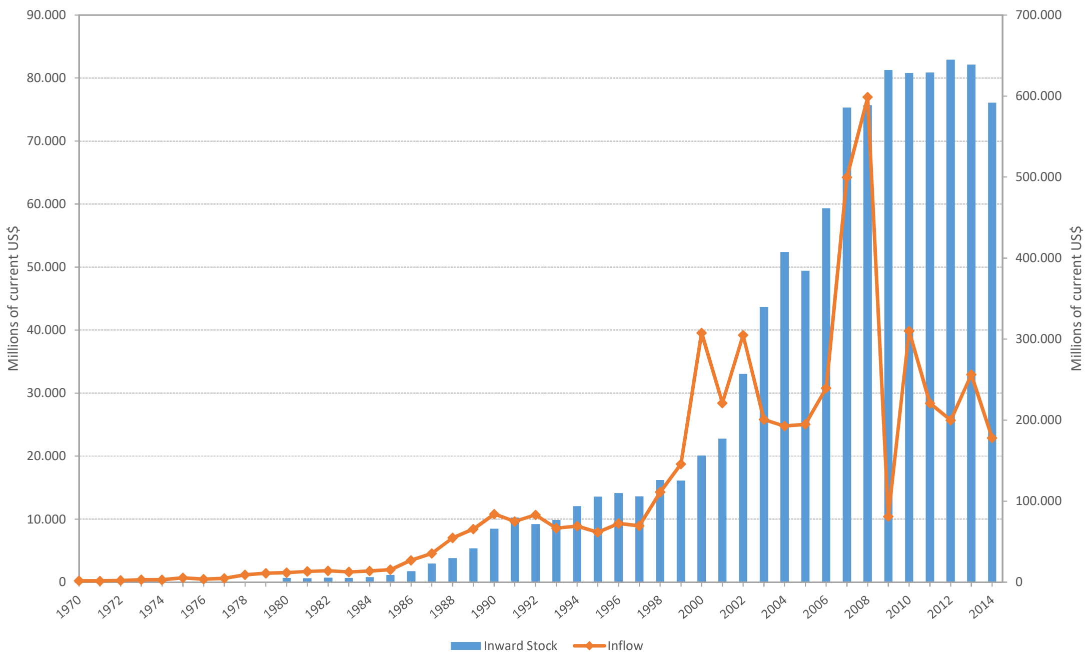
Chapter 2

Determinants of FDI for Spanish regions: Evidence using stock data.

2.1 Introduction and motivation

Foreign Direct Investment (FDI) has become the engine of the globalization process. It is broadly accepted that this type of investment has beneficial effects in terms of job creation and technological transfers. Therefore, most governments in the world, either at a national or regional level, have promoted generous policies to attract FDI on their soil. However, there is no current consensus in the literature about the drivers of FDI and, therefore, the most effective policies for its promotion. In the European case, the significant increase in FDI flows after the launching of the euro gave rise to mounting and persistent external imbalances, and examining the factors that attract direct investment has become a "hot topic" in the rebalancing debate in the Eurozone, that is far from being solved empirically.[1] Thus, there is a growing body of research that provides empirical evidence not only on the factors that determine FDI, but also about the existence of significant disparities in its distribution across countries. Indeed, the peripheral European countries became massive capital importers until the outbreak of the crisis and Spain was by far the largest capital importer in absolute terms.

Fig. 2.1 Foreign Direct Investment in Spain.



Source: UNCTAD database.

Figure 2.1 presents the inward FDI stocks and inflows to Spain during the period 1970-2014.[2] Until the mid−eighties FDI was negligibly small. Since then, FDI inflows began to increase after the Spanish entry in the EU in 1986 (and the European Monetary System

[1]According to Bertola et al. (2013), there were two primary and interrelated causes explaining these imbalances: First, the monetary union created optimistic expectations regarding the rapid convergence of the peripheral countries with the core ones in the Eurozone. Second, the introduction of the euro eliminated the exchange-rate risk and induced investors to disregard country-specific bankruptcy risks. Both causes generated an investment and credit boom in the periphery and implied a large-scale reallocation of capital from the core to the periphery that materialized as current-account deficits.

[2]Note that data on inward FDI stock is only available since 1980.

in 1989) and later with the creation of the monetary union. In particular, FDI inflows have increased from US$ 4.570,7 million in 1987 to US$ 76.992,5 million in 2008.

Exchange rate stability has been crucial in the reduction of risk and, then, the increase of the general attractiveness of the country as an investment destination. However, the impressive upward trend in FDI inflows was disrupted by the global financial crisis in 2008. Since 2008, FDI inflows have declined. However, the bar graph in Figure 2.1 shows that although the FDI inward stock has slightly declined, it remains large and suggests that Spain is still an attractive host country for FDI. Even during the crisis, the value of the stock is maintained (at least in nominal terms) and only decreases with some delay, in 2013-2014.

Although there is an extensive empirical literature at the country-level for developed countries, the evidence about the determinants of FDI at the regional level is quite recent and relatively scarce (Kandogan (2012), Villaverde and Maza (2012) and Chan et al. (2014)).[3] This scarcity mainly derives from the lack of data. Some efforts have been made, however, to collect and streamline data for regions by regional statistical offices. Studies reveal regional disparities in the distribution of FDI and, in this regard, the case of Spain clearly arises as a prominent example.

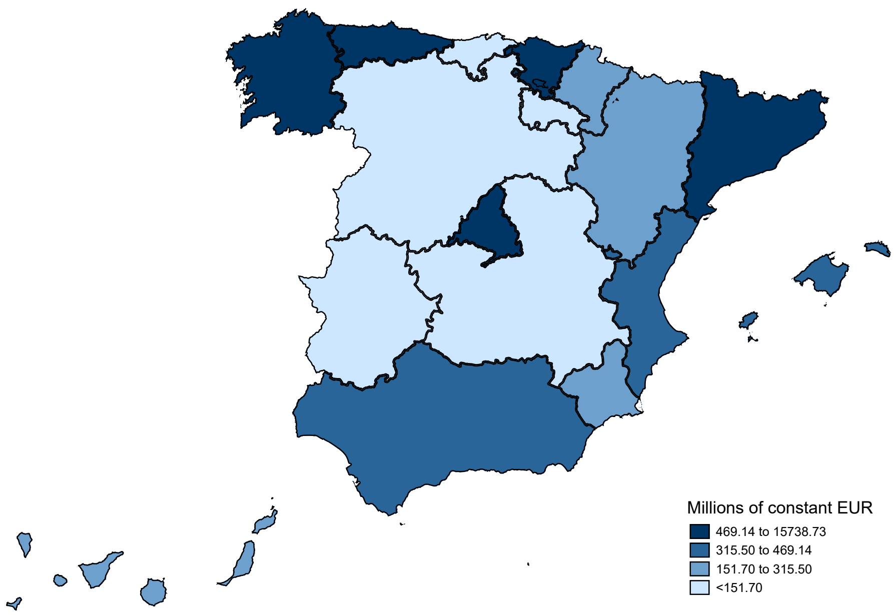
Looking at Figure 2.2, we can observe that the regional distribution of FDI is highly concentrated. The map shows total inward FDI in each of the 17 Spanish regions during the period of analysis.[4] FDI seems to be attracted mainly to four regions over the period 2004-2013. Of these, Madrid has been by far the largest recipient of inward FDI in Spain (almost 3.500 million euros), followed by Catalonia (with 1.048 millions), Asturias (304 millions) and the Basque Country (237 millions). Extremadura has attracted the lowest

[3]The empirical literature at the country-level has focused mainly in OECD countries, since traditionally they have represented a prominent share of world's FDI flows (Bénassy-Quéré et al. (2007) and Talamo (2007)).

[4]We do not account for Ceuta and Melilla due to data availability. DataInvex provides data for these autonomous cities only for 3 out of 40 of the source countries included in our sample and not for the whole time span. Furthermore, FDI data for these cities is not provided separately what could be problematic for the analysis when including specific host region characteristics.

amount of FDI (an average of 6 millions). Another stylized fact, also related to the regional distribution of FDI, has to do with the geography of FDI: with the exception of the region of Madrid, the other regions that concentrate more FDI are in the coast. In addition, potential regional spillovers may influence Aragon's and Navarre's FDI (due to their closeness to Catalonia and Valencia in the case of the former and to the Basque Country in the case of the latter).

Fig. 2.2 Spatial distribution of average inward FDI stock (2004-2014).



Source: DataInvex.

Table 2.1 Sectoral distribution of gross inward FDI flows in Spanish regions, Millions of euros, No-ETVE

	2004					2013				
	Total	Agriculture	Industry	Construction	Services	**Total**	Agriculture	Industry	Construction	Services
Andalusia	242.83	6.10	21.94	37.14	177.65	149.87	4.71	13.79	23.45	107.92
Aragon	46.28	0	33.09	4.20	8.99	104.87	0	76.19	0.08	28.60
Cantabria	1.05	0	0	0.01	1.04	15.54	2	0.01	1.23	12.31
Castile and León	5.28	0.14	1.82	0.02	3.30	171.08	0.5	81.99	1.66	86.93
Castilla-La Mancha	36.33	0	29.61	2.43	4.28	24.52	3	2.53	0.004	18.99
Catalonia	1400.78	5.09	687.65	61.19	646.84	3464.95	4.44	1395.55	86.94	1978.02
Madrid	3414.73	3.53	526.99	213.12	2671.07	7172.54	4	1517.77	279.33	5371.45
Valencian Community	659.15	5.64	176.30	9.80	467.41	160.46	4.46	20.30	14.32	121.38
Extremadura	1.76	0.06	0.86	0	0.84	12.25	12.15	0.04	0.001	0.05
Galicia	31.74	0.4	22.68	4.14	4.52	221.50	0.25	165.85	37.04	18.34
Balearic Islands	89.03	0	0.04	43.25	45.73	759.29	7.54	0.34	45.33	706.08
Canary Islands	247.09	0.55	0.02	2.09	244.44	35.99	0.25	0.18	0.89	34.66
Navarre	1.12	0.24	0.87	0	0.01	62.22	0	1.39	0	60.83
La Rioja	1530.76	0.59	1524.20	0.005	5.95	43.77	0.003	29.87	0.19	13.70
Basque Country	255.17	0.003	84.41	1.29	169.46	290.34	0	165.22	10.43	114.70
Asturias	610.12	0.001	593.99	0.13	15.99	872.50	0	851.01	0	21.49
Murcia	326.46	1.70	309.90	0.06	14.80	43.85	1.80	36.34	0.003	5.71
Spain	**8910.44**	**24.06**	**4022.37**	**379.01**	**4484.99**	**13605.93**	**45.12**	**4358.39**	**500.89**	**8701.53**

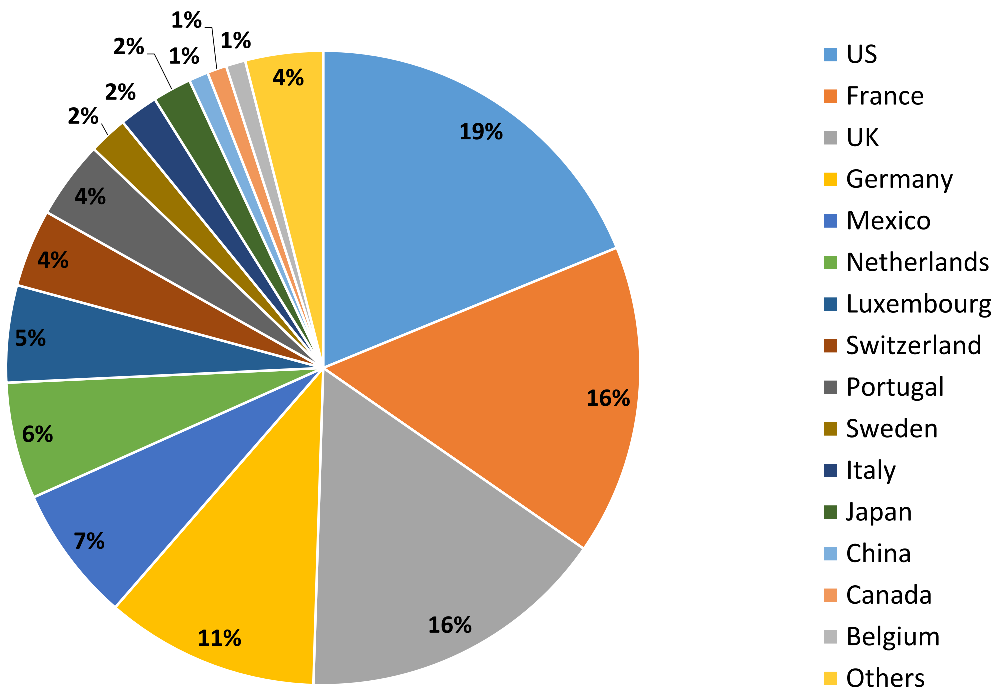
Notes: Inward foreign investment from the 40 source countries included in our study. *Source:* DataInvex.

The sectoral distribution of FDI is also very heterogeneous. Table 2.1 provides the sectoral breakdown (agriculture, industry, construction and services) for the years 2004 and 2013. Three descriptive facts can be derived from the data. First, FDI inflows are also highly concentrated across sectors. Industry and services concentrate the 95% of total FDI. Second, the sectoral distribution of FDI across regions does not vary across time. The exceptions are Castilla-La Mancha, Catalonia, Extremadura, Navarre and Basque Country. And, third, the amount of inward FDI flows allocated to services increases over time. By 2013, this sector concentrated most inward FDI, 64% of the total. Despite these facts, the sectoral breakdown of FDI stock data has not been included in this study due to data availability.[5]

Finally, Figure 2.3 shows the main source-countries of FDI in Spain. The largest stock of FDI has come from the US (2.601 million euros, on average, what represents the 19% of inward FDI stock in Spain), followed by France, the UK and Germany, that jointly represent over 60% of the total. Together with Mexico, the Netherlands, Luxembourg, Switzerland and Portugal they constitute the largest foreign investors in Spain.

[5]The Spanish Ministry of Economy, Industry and Competitiveness through its investment registry (DataInvex database) provides data on FDI flows disaggregated by regions and sectors whereas FDI stock data are available at national level since 2007. Indeed, we are indebted to Isaac Barbero and Emilio Carmona from the Spanish Ministry of Economy, for providing us with the extended and disaggregated version of FDI stock data from the DataInvex database that allows us to conduct the study at the regional level.

Fig. 2.3 Main investors in Spain: Average FDI stock for the 2004-2014 period, in percent.



Source: DataInvex.

In this context, examining the factors that attract FDI in Spain seems very pertinent. Indeed, the analysis of FDI determinants at the regional level within a country may be of special interest for countries with a federal structure where regional authorities have a key role in designing policies to encourage FDI. Moreover, the motivation for studying FDI from a regional standpoint comes from the fact that a lot of interesting characteristics are hidden at more aggregate levels. More importantly, FDI determinants and effects may be localized and, thus, a regional analysis may be more appropriate to obtain better-grounded results.

This paper seeks to contribute to the existing literature by providing further insights on the determinants of inward FDI activity from a macroeconomic point of view using Spanish regional data for the period 2004-2013. We claim that our approach is a macroeconomic one because we use aggregate FDI instead of Foreign Affiliate Sales (FAS) as the dependent variable. We think this definition of FDI is preferable to FAS, because the latter may be a measure of multinational enterprise (MNE) production and hence, the drivers and the theoretical model might be different (Bergstrand and Egger (2007)). Furthermore, Baldwin et al. (2008) highlight the drawbacks of using FAS instead of the book value. They describe the two main sources to measuring FDI: central bankers and economics ministries. Central bankers consider FDI as part of the capital account of balance of payments and they gather statistics accordingly. Economics ministries, in contrast, gather data on the number of employees, sales and assets of foreign controlled firms. Unfortunately, the production/sales data is based on surveys and thus generally subject to confidentiality requirements that make the data difficult to access for scholars. And when it is accessible, it is usually for just one nation since the various datasets are not compatible enough to pool the data.

The FDI data used in this paper has been extracted from *DataInvex*, a rich and under-exploited database that provides FDI data on FDI stocks and flows broken down by autonomous communities or regions (NUTS 2) for the Spanish economy.[6] Our aim is to shed some light on the increasing and heterogeneous patterns of inward FDI that the Spanish regions have attracted since the launching of the euro. We estimate an extended gravity model using the Poisson Pseudo-Maximum Likelihood (PPML) estimator in a gravity framework. We rely on PPML estimator because it produces unbiased and consistent estimates when the dependent variable comprises of a large proportion of zero observations as is the case in the present study.

[6]According to international recommendations, DataInvex defines FDI as those transactions through which a direct investor acquires or increases its participation in a company resident in another country so that it can exert an effective influence in its management. In practice, it is considered that the investor has the ability to influence the management of a company when he has at least 10% of the capital or voting rights. Specifically, we use in our study the investment position which represents the value of the assets that direct investors hold in companies, resident in countries other than their own, with direct investment. The position data is established from the perspective of the country that presents them (reporting country). For the purposes of this study, the position of foreign investment in Spain would be the value of the shares of non-resident investors in companies domiciled in Spain. The participations are valued on the basis of the book value of the equity of the direct investment company.

Furthermore, we use inward FDI stock data, as an alternative to flows, because stocks are much less volatile than flows. This is a salient feature of our study as our hypothesis is that this variable is the most suitable for empirical studies, as Wacker (2013) has recently corroborated. Flows are more volatile not only due to the existence of economic shocks but because they are also very dependent on individual large-scale investment decisions. Our aim is to find the determinants of the long-run allocation patterns of FDI across regions. We differ in this respect from other empirical approaches of the literature that mainly focus on short term macroeconomic variables (such as the exchange rate or the business cycle performance), to explain the FDI behavior.[7] Therefore, we are not interested in determining the effect of surges (i. ex. a sudden stop) in the evolution of FDI. All in all, for the sake of comparison, we replicate the analysis using inward FDI flows.

A complete understanding of the factors that determine FDI seems pertinent given the interest of many countries currently facing financial constraints to attract FDI in their quest to foster economic activity. In this context, FDI stocks, as an alternative to flows, provide a better approximation to the long-run behavior of investment decisions, the ones really relevant for growth.

The reminder of the paper is organized as follows. Section 2 reviews the theoretical approaches and reports the main determinants of FDI, whereas section 3 presents the empirical literature. Section 4 provides a description of the variables and the data sources. Section 5 describes the models to be estimated together with the results and, finally, Section 6 concludes.

[7] See, for instance, Russ (2009), Russ (2012) or Cavallari and D'Addona (2013).

2.2 Theoretical background

From a theoretical point of view, the so-called eclectic or *OLI paradigm* put forward by Dunning (1977, 1979) has been considered the standard workhorse theoretical framework within this strand of literature. Dunning suggests that three types of advantages influence the foreign investment decision of a MNE: Ownership, Location, and Internalization (OLI).

Here, it is worth to note that while *Ownership and Internationalization advantages* are essentially of a microeconomic nature, *Locational advantages* generally correspond to macroeconomic variables. Focusing on locational advantages, Dunning (2000) identifies four main motives that encourage MNEs to engage in foreign production: market seeking, resource seeking, efficiency seeking and strategic assets seeking. Market seeking motives correspond to FDI that aims at supplying the local market or markets in adjacent territories. The host market size, its per capita income and consumer demand (all of them to take advantage of the economies of scale) are the main reasons behind market seeking FDI. Resource seeking companies are those investing abroad in order to obtain cheap natural resources and/or unskilled labour. Hence, locational decisions depend on factor endowments differences. Efficiency seeking investment is designed to promote a more efficient division of labor or specialization of assets by MNEs. Finally, strategic asset seeking FDI searches for resources such as technology, skilled workers and assets that can support worldwide development of a firm and weaken the competitive position of its competitors.

With the incorporation of multinational firms into the general equilibrium trade models from mid 1980s onwards, it became possible to base empirical work on theoretical predictions regarding the relationship between MNE activity and home and host countries characteristics (Barba Navaretti and Venables (2004)). These theories allowed to explain the existence of two basic types of FDI, namely horizontal (market-oriented) and vertical (export-oriented) FDI. The first type of FDI is explained using the proximity-concentration hypothesis which explains the trade-off between maximizing proximity to customers and concentrating production to achieve scale economies (Horstmann and Markusen (1987)). In this regard, horizontal FDI (HFDI) may imply duplication of the entire production process in several countries. In contrast, the second type (VFDI) is explained by using the factor-proportions hypothesis which accounts for the existence of vertically integrated firms with geographically fragmented production (Faeth (2009)). In terms of HFDI, the most important factor to attract FDI is the size and growth of the host country whereas VFDI mainly looks for cost competitiveness. VFDI is conducted in order to minimize production costs in the host country and then to export the output produced to the home country or to third countries. Hence, the most important location factor for VFDI is resource endowment.Helpman (1984, 1985) showed that countries' differences in relative factor endowments (the so-called factor-proportions hypothesis) explained VFDI.

Combining vertical and horizontal motivations for FDI, Ethier and Markusen (1996) and Markusen and Maskus (2002) formulated the *knowledge-capital model*. Markusen and Venables (1998) explain the knowledge-capital model through two tradeoffs. The first and key tradeoff is between the scale economy gains that come from dividing up production and spatially dispersing it to be near customer concentrations. The second trade-off concerns productive factors. The so-called VFDI strives to place each stage of production in the nation where it is cheapest. However, it is important to make some remarks at this point. It can be misleading to think of the former literature as a proper FDI theory. In fact, it is a FAS literature as qualified by Markusen and Maskus in their 2002 book.[8] More recent contributions to this theoretical approach allow for heterogeneous firms. Helpman et al. (2004) consider firms that have different productivities and fixed costs of establishing "beachhead" in various markets. The most competitive firms tend to sell much more and thus tend to find the transport-cost-saving aspect of FDI especially attractive. This would explain the widely spread phenomenon

[8]Markusen and Maskus (2002).

that FDI is dominated by large firms. Carr et al. (2001) provided the first empirical test of the knowledge-capital model's hypotheses. Using a panel of data (US outbound and inbound affiliate sales in many nations from 1986 to 1994), they find evidence for both the horizontal and vertical motivations for FDI. Blonigen et al. (2003) question their econometrics, which, when corrected, no longer support the vertical motivations for MNE activity. Overall, under the knowledge-capital model, similarities in market size, factor endowments and transport costs were determinants of HFDI, while differences in relative factor endowments determined VFDI. The knowledge-capital model has recently been extended to explain other forms of FDI such as export-platform FDI (see Ekholm et al. (2007); Bergstrand and Egger (2004)) which is used to serve the neighboring markets of the host country. To sum up, while recent Eaton-Kortum (Ricardian) type models have been extended to motivate gravity equations for multinational production of firms either in isolation from trade flows (Ramondo (2014)) or with trade flows (Ramondo and Rodríguez-Clare (2013)), theoretical foundations for FDI per se are limited primarily to Bergstrand and Egger (2007).[9]

Indeed, a third strand of the theoretical literature on FDI determinants is the one based on the *gravity approach* to FDI. Theoretical foundations for FDI are limited basically to Bergstrand and Egger (2007) and Head and Ries (2008). In principle, these papers provide general equilibrium theories for FDI and not FAS.[10] In this case as in gravity models applied to trade flows, the main explanatory variables are geographical location (due to its effects on transport costs) and the size of the country, crucial when economies of scale are recognized to exist in the activities of MNEs. The gravity model states that the closer two countries are (geographically, economically and culturally) the higher will be the FDI activity between them. Recent papers have provided some micro foundations for the gravity FDI specification. Head and Ries (2008) develop a model of cross-border mergers and

[9]While Markusen and Maskus (2002) knowledge-capital model is about FAS, Bergstrand and Egger (2007) is about both, FAS and proper FDI.

[10]Recently, theory has been directed, as well as empirical work, to FAS, starting with Brainard (1997) and continuing to Ramondo et al. (2015).

acquisitions (M&A) activity where a mother company has a randomly assigned advantage in controlling the company in the host country, but faces a disadvantage in monitoring technology that gets more severe with geographical distance. Bergstrand and Egger (2007) add internationally mobile capital to the knowledge-capital model and find that a "modified" gravity model fits the data better. Similarly to Head-Ries, the Bergstrand-Egger model stresses the importance of relative distance. In this case, the amount of expected FDI between two nations depends upon the bilateral distance relative to some measure of the host and home nation's distance to alternative FDI sources and destinations. Kleinert and Toubal (2010) provide the theoretical underpinnings of the gravity equation applied to the analysis of FAS showing that gravity equations can be used to discriminate between different theoretical approaches, namely, two proximity-concentration models of HFDI with homogenous (Brainard (1997)) or heterogenous firms (Helpman et al. (2004)) and a two-country factor proportions model of fragmentation that explains VFDI. [11]

In this very similar theoretical framework, the relationship between FDI and trade has been put forward, among others, in Helpman et al. (2004) and from an empirical point of view in Camarero and Tamarit (2004) and Camarero et al. (2018). Therefore, many of the FDI drivers can be connected to the ones for trade. In this very same vein, changes in interest rates and exchange rates have been identified as additional FDI determinants. The role of these variables is explained by the risk diversification hypothesis of MNE, as firms are risk averse and, therefore, are trying to diversify business risk.

Lastly, it is important to consider also the influence of political variables on FDI. The strategies adopted by companies and their performance on international markets are largely determined by institutions (i.e. the "rules of the game" as in Busse and Hefeker (2007)). In this context, foreign investment can be regarded as a "game" in which the players are the multinational firm and the government of the host country, or as a contest between

[11] The factor-proportions model is based on Venables (1999).

governments to attract FDI (Faeth (2009)). Policy variables such as corporate tax rates, tax concessions, tariffs and other fiscal and financial investment incentives have thus been posited to have an effect on FDI.

Briefly stated, the literature suggests a variety of theoretical models explaining FDI that do not necessarily replace each other (see Blonigen (2005), Faeth (2009) and Assunçao et al. (2011) for a thorough literature review on FDI determinants). Therefore, in the next Section we review the wide range of factors that can be considered in empirical studies in order to find the determinants of FDI. These factors involve both micro (e.g., organizational aspects) and macro (e.g., resource allocation) dimensions and call for a very accurate empirical validation.

2.3 A brief survey of the empirical literature on regional FDI determinants

The location determinants of FDI have been largely explored in the empirical literature using different approaches. Blonigen and Piger (2014) make an extensive overview of the empirical determinants of FDI literature. As we have already stressed before, although sometimes difficult, it is important to distinguish between proper FDI and FAS, being the latter derived from the Multinational Production (MP) theory. Even if in this review we may not always make this distinction explicitly, we are aware of its relevance and will use it in the selection of the econometric specification as well as in the choice of the variables used in Section 5. Since this study aims to identify the factors that best explain inward FDI to a particular region, it concentrates on the locational variables.

This section will emphasize the revision of the regional analyses. However, in order to set up a taxonomy of the empirical research on FDI determinants, we will single out a few recent studies on specific areas (OECD, EU...) to focus immediately on country-specific analyses. The latter are developed both at national and regional levels. It is important to stress that the variables involved at national and regional level studies may be different due to data availability but also to theoretical considerations. More specifically, *market seeking* variables, related to the size of the host market are less relevant at a regional level, as the FDI project will be serving the whole nation, so the specific location of the FDI within the country must be due to other variables more closely related to *endowment differences* or distinct *geographical or political characteristics*. At the same time, many variables, such as the exchange rate, country-risk or business cycle (output gap) considerations have full sense at the national level but not at the regional one. In order to uncover the specific variables at play at the regional level, we will review the empirical literature at our disposal. It is worth mentioning that studies carried out at the national level have been the most prolific. Therefore, first, we will review the main studies that analyze FDI determinants for groups of countries; second, we will report the most significant studies at the country level; then we will summarize the main results of previous regional studies and finally, we will review the literature for the Spanish case.

The lion's share of the literature dealing with the FDI determinants focuses on the OECD since traditionally they have been representing an outstanding share of the world FDI inflows. Recent research by Economou et al. (2017), using both static and dynamic panel techniques over the period 1980-2012, has identified lagged FDI, market size, gross capital formation and corporate taxation as robust FDI determinants for OECD countries. Moreover, Bruno et al. (2016) estimate Ordinary Least Squares (OLS), Poisson and Heckman models on a gravity framework for 34 OECD countries over 1985-2013. They conclude that GDP, GDP per capita and EU membership affect positively FDI inflows. Although they focus on modelling flows, they also replicate the estimations for FDI stocks and find qualitatively similar results.

Other recent country-group studies have focused on European Union countries and, in particular Central and Eastern European Countries (CEEC) after the 2004 enlargement. They try to unravel the specific importance of this economic integration process on FDI. Bevan and Estrin (2004) use panel data techniques in a gravity framework to examine the determinants of FDI in CEEC for the period 1994-2000. They identify unit labor costs, gravity factors, market size and proximity as the most important drivers of FDI. Besides, announcements about EU accession proposals have been found to have an impact on FDI for the future member countries.

Similarly, Demekas et al. (2007) use both FDI flows and stocks to explain inward FDI in 16 European transition economies, applying cross-section as well as dynamic panel data techniques. In addition to gravity factors, they control for host country policy variables as FDI determinants. They find that gravity factors, trade and foreign exchanges liberalization as well as infrastructure reforms encourage FDI, while high unit labor costs, corporate taxes and import tariffs discourage it.

For the case of EU countries, Canton and Solera (2016) estimates a Heckman two-step selection procedure in the context of a gravity model for the period 2003-2014. His findings suggest that business climate and product market regulations play a key role on attracting greenfield investment in the EU.

Far less research has been carried out concerning the factors that attract FDI within a particular country, that is, at a *regional level*. Moreover, most of these studies at the subnational level have been conducted in large countries (most of them in China, but also in other countries such as the US), and only rarely in smaller countries. Using provincial panel data from China, Chan et al. (2014) examine all possible flows of causality involving FDI and a set of potential determinants, both in the short and in the long run. In a context of error correction models with Granger causality tests they show that in both the short and the long run, GDP growth directly influences FDI, while growth in local infrastructure and local investment have indirect but not direct influence.

In the case of the US, Kandogan (2012) investigates within-country regional locational decisions of multinationals. He uses states within the US as regions and applies multiple regression analysis for both FDI stock and flows as dependent variable. They identify unemployment rate, market size, per capita income, income growth and state regulations as the main determinants of FDI.

Although to a lesser extent, regional studies within small countries can also be found in the empirical literature. This is the case of Chidlow et al. (2009) that, using a multinomial logit model, investigate the location determinants of FDI inflows in the Polish regions. The authors claim that knowledge seeking factors alongside market and agglomeration factors act as the main drivers for FDI inflows into the Mazowieckie region, while efficiency and geographical factors encourage FDI to other areas of Poland. Similarly, Dimitropoulou et al. (2013) analyze FDI projects in UK regions to identify the main determinants of the location choices of these investments. Using multinomial and conditional logit models, they find that existing regional specialization is the single most important determinant of inward FDI. Focusing on Croatian regions, Kersan-Skabic and Tijanic (2014) apply static (random effects estimator) and dynamic panel data methodologies to analyze their attractiveness as FDI locations. They conclude that FDI inward stocks are attracted to Croatia by education, infrastructure, the manufacturing industry and the capital city region, while unemployment and EU-border regions have a negative effect on FDI.

In the case of Spain, although most of the empirical evidence has been obtained at the national level, since the early 1990's the regional perspective has also been brought to the forefront of this field of research.[12] Egea Román and López Pueyo (1991) conduct a cluster

[12]Some examples of these studies are Bajo-Rubio and Sosvilla-rivero (1994), Pelegrín (2003) and Martínez-Martín (2011), among others.

analysis for the period 1985-1989 and conclude that per capita and per employee income, human capital and the productive structure are the main drivers of FDI. In contrast, the unemployment rate, infrastructure endowment and subsidies are not found significant. Pelegrín (2002), using three different methodologies (OLS, LS with fixed effects and GLS), estimates a FDI equation for the period 1993-1998. Three key determinants of FDI location are identified, namely, market size, human capital and public incentives; however, infrastructures are not significant while labour costs have a positive coefficient. In another paper, Pelegrín and Bolancé (2008) find that agglomeration economies and the concentration of research and development (R&D) activities are important drivers for manufacturing FDI in a model estimated using GLS. Using the same methodology together with instrumental variables, Rodríguez and Pallas (2008) show that real GDP, human capital, sectoral export potential and the differential between labor productivity and the cost of labor are key factors in explaining the regional distribution of FDI during the period 1993-2002.

More recently, Villaverde and Maza (2012) analyze the regional distribution of FDI in Spain and its main determinants between 1995 and 2005/2008 by means of panel data techniques, namely by GLS and two stage GLS. The econometric analysis reveals that factors such as economic potential, labour conditions and competitiveness are important for attracting FDI both at aggregate and sectoral levels.

Finally, Gutiérrez-Portilla et al. (2016) estimate a FDI equation by GMM and GLS over the period 1997-2013, and show that FDI inflows in Spain are mainly determined by market size, the level of human capital in interaction with wages, and the characteristics of Madrid as capital of the nation. The research is conducted not only for the whole period and total FDI but also for two sub-periods (pre-crisis and crisis) and areas of origin (Europe and America).

Our paper contributes to the existing literature in different respects. First, we use a testing framework based on the gravity model embedding competing theoretical approaches. Second, despite the amount of empirical work that has analyzed the factors that determine FDI, the majority has focused on the evolution of flows. To the best of our knowledge, there are no empirical papers dealing with regional FDI determinants based on stock data. Overall, from the above literature review it can be inferred that the main FDI determinants in Spain are those related to market seeking (i.e.market size) and resource seeking (i.e.human capital, labor market conditions and physical infrastructure endowment). Third, unlike baseline estimates in earlier studies using OLS or Heckman two-stage selection procedure, we use the PPML estimator proposed by Silva and Tenreyro (2006) which provides robust results.

In the next two sections, adopting a regional perspective, we provide further evidence on the performance of the PPML estimator using FDI stock data. The existence of an updated and largely unexploited database developed by the Spanish Ministry of Economy (DataInvex) calls for further research in this field.

2.4 Data and stylized facts

In this section, and prior to present the empirical results, we describe the variables we have chosen for the analysis based on the previous literature discussion, both from a theoretical and empirical point of view. In addition, we describe their sources and point to some stylized facts. More detailed variable definitions and data sources can be found in Table 3.A.1, while Tables 2.A.2 and 2.A.3 report some basic descriptive statistics and the correlations, respectively.

The dependent variable is bilateral inward FDI stock from the origin countries towards the Spanish regions for the period 2004-2013.[13] They have been obtained from *DataInvex* from the Spanish Ministry of Economy, Industry and Competitiveness that provides data on FDI disaggregated by regions. Unlike the vast majority of the literature on this field, we

[13] We do not include FDI information of ETVE (Empresas de Tenencia de Valores Extranjeros-brokers) firms because they are considered instrumental companies whose existence obeys to fiscal optimization strategies within a business group and in many cases their investments lack direct economic effects.

use bilateral FDI stock as our dependent variable.[14] We argue that stocks are much less volatile than flows, especially in relatively small countries.[15] This volatility has its origin not only in the existence of economic shocks, but also on individual large-scale investment decisions. However, we have also applied the analysis to inward flows for the sake of comparison. Furthermore, FDI stocks (as a valuation of the cumulative FDI) provide a better approximation to the long-run behavior of investment decisions, the ones really relevant to capture growth and the dynamic effects of economic integration. In the same vein, Baldwin et al. (2008) argue that factors such as stock market fluctuations or exchange rate volatility cause short-run variability on FDI flows that may not always be linked to the explanatory variables and therefore lead to worse model fit for flows than for stocks.

Our dataset is annual and covers the period 2004-2013. Hence, we have a balanced panel with dimension n=680 (17 regions x 40 countries, that is, all possible bilateral relationships) and T=10. The number of observations is nxT=6800. Table 3.A.2 reports the countries included in the study. We assume zero FDI in case of non-reported data as in Canton and Solera (2016) and treat negative values of FDI stock as zero following Gouel et al. (2012).[16]

In search of the main determinants of inward FDI in the Spanish regions, we have chosen a set of explanatory variables that capture the main factors likely to attract FDI, considering not only the theoretical models but also the empirical studies previously discussed. These variables include factors describing labour market characteristics, the degree of openness,

[14]Exceptions are Wacker (2013) and Blonigen and Piger (2014).

[15]Bénassy-Quéré et al. (2007), p.769.

[16]This last point can be tricky. Following the advice of one referee we elaborate on this point to justify our position. According to UNCTAD (2018): "FDI flows are presented on a net basis, i.e. as credits less debits. Thus, in cases of reverse investment or disinvestment, FDI may be negative." Thereby, negative FDI flows have real economic meaning. On the contrary, although analytically a negative sign on FDI stocks indicates that at least one of the three components of FDI flows (i.e. equity capital, reinvested earnings or intra-company loans) is negative and is not been offset by positive amounts of other components, from an economic perspective this negative sign lacks real economic meaning and are usually considered the consequences of accounting methods (see Gouel et al. (2012), Bae and Jang (2014), Baronchelli and Uberti (2018) and Petkova et al. (2018)). Thereby, replacing the negative FDI stocks with a zero have become a common practice among some recent empirical studies (see, for instance Bae and Jang (2014) and Petkova et al. (2018)).

as well as capital endowment, both physical and human.[17] The variables considered are the following:

Market Size (MS_{it}): proxied by GDP per capita. Most of the literature considers this variable a robust determinant of FDI. From Dunning's OLI framework, to the HFDI theory (i.e. the proximity-concentration hypothesis) as well as the knowledge-capital model. As market's size increases, so do the prospects of higher demand (greater purchasing power), better market opportunities for the firms and potentially higher returns on their capital. In this regard, market seeking is among the main motives for investors to undertake FDI. Larger markets in the host region do not only denote good economic performance, but also allow for a reduction in the cost of entry through the exploitation of economies of scale. Hence, we would expect to find a positive relationship between market size and FDI.

Labour Productivity (LP_{it}): defined as GDP per total employment. The role of this variable would be also in line with the OLI framework, HFDI and the knowledge-capital model. The expected sign on FDI is ambiguous. We would expect a negative association when labour productivity in the host country is low due to capital scarcity, thereby the marginal return to capital is relatively high and FDI is attractive. On the other hand, a positive relationship might be expected if labour productivity indicates favourable factors for FDI, such as market size and good business climate conditions (see Canton and Solera (2016) and Razin et al. (2008)).

Wage (W_{it}): defined as employee's compensation per hour worked. The effect of wages in the host country is somewhat ambiguous in the literature: If VFDI activities are the dominant driving force it should be expected a negative relationship. However, if the driving force is HFDI it would be expected a positive relationship between the wage level and FDI (indicating the need for a qualified workforce in the foreign affiliate production and higher sales). Given

[17] It should be noted that the choice of variables was somewhat restricted by the availability of disaggregated Spanish data.

that the sectoral breakdown of FDI stock data could not be included in this study due to data availability, wages are not disaggregated by sectors either. Table 2.2 reports the gross value added (GVA) across sectors and regions together with the level of wages. It can be observed that the Spanish sectoral structure is the typical one for developed countries, with the services sector generating the highest value added, followed by industry. As regards the level of wages, the industry is the highest-wage sector, followed by services. This structure is maintained across regions.

Table 2.2 Gross value added (GVA) and wages by sector and region, 2010.

	GVA			Wages		
	Industry	Construction	Services	Industry	Construction	Services
Andalusia	16.614.034	13.548.900	96.529.409	17,02	14,55	15,08
Aragon	7.381.593	2.964.839	19.581.313	19,79	15,24	16,25
Cantabria	2.570.365	1.255.569	7.722.694	20,32	15,28	15,11
Castile and León	10.961.019	4.582.546	32.752.292	18,81	13,53	15,30
Castilla-La Mancha	7.226.701	4.032.910	22.367.302	16,78	13,52	15,79
Catalonia	38.933.793	14.301.827	130.732.544	21,27	16,10	16,95
Madrid	19.017.886	12.333.904	149.783.526	21,88	16,33	19,18
Valencian Community	16.261.288	9.682.957	65.685.541	16,68	14,56	15,19
Extremadura	2.110.368	1.968.227	11.346.594	14,27	11,63	15,21
Galicia	10.421.849	5.403.221	33.897.441	16,04	13,32	14,72
Balearic Islands	2.053.226	2.101.212	19.738.430	16,71	13,19	15,47
Canary Islands	3.472.665	2.939.545	30.750.777	15,92	13,38	15,30
Navarre	5.077.599	1.406.313	9.685.817	22,22	19,70	16,95
La Rioja	2.060.606	697.774	4.181.096	17,67	14,46	15,47
Basque Country	16.390.237	5.317.238	38.015.463	23,64	18,62	17,87
Asturias	4.855.996	2.073.347	13.695.672	22,18	17,23	15,31
Murcia	4.436.025	2.692.043	17.478.263	15,52	13,57	14,93

Notes: Gross value added (GVA) is expressed in thousand of euros (current values) whereas wages is measured as employee's compensation per hour worked (current EUR). *Source:* INE and Eurostat.

Unemployment Rate (UR_{it}): defined by the annual unemployment rate. The effect of the unemployment rate on FDI could be either positive or negative. High levels of unemployment may draw in efficiency seeking FDI by increasing the availability of labour and the willingness of employees to work harder and for lower wages. However, unemployment can also reduce FDI by restricting incomes and spending power in host country markets. Furthermore, higher unemployment could also signal less competitive conditions and a lower quality of life that tend to discourage foreign investors.

Human Capital (HC_{it}): proxied by the share of population with tertiary education. Again, the effect of human capital on FDI can be argued to be positive or negative. The positive relationship accords with the OLI framework and the proximity-concentration HFDI theory. All other things being equal, regions with highly-skilled workers would be expected to compete more favorably than others in attracting FDI. Indeed, qualified workers attract FDI oriented to industrial sectors with higher demand and technology. Hence, higher human capital is expected to have a positive effect on FDI. However, if FDI was oriented to activities with a very low value added (following VFDI theories), it would seek cheap and less qualified workers. Hence, a negative association between skilled labour and FDI could also be expected.

Trade Openness (OP_{ijt}): defined by the sum of bilateral exports plus imports over GDP. According to the OLI framework, the HFDI and the knowledge-capital model, a reduction of barriers to external trade and, in general, a business-friendly economic climate would increase investment in general, thus attracting FDI as well. In addition, it is expected that MNEs would invest in trade-partner markets with whom they are already familiar. Numerous empirical studies suggest that trade (imports and exports) complements rather than substitutes FDI. Much of FDI is export-oriented and may also require the import of complementary, intermediate and capital goods. In either case, the volume of trade is enhanced and thus trade openness is generally expected to be a positive and significant determinant of FDI.

Differences in Relative Factor Endowments (RFE_{ijt}): proxied by differences in per capita GDP between the host and the source country. Following Mitze et al. (2008), the variable takes a minimum of zero for equal factor endowments in the two regions. According to the knowledge-capital model, differences in relative factor endowments determine VFDI. Vertical or export-oriented FDI is conducted in order to minimize production costs in the host country and then to export the output produced to the home country or to third countries. In this regard, the most important location factors are resource endowments.

Infrastructure (RI_{it},GCF_{it}): proxied by road infrastructure (RI_{it}) and gross capital formation (GCF_{it}). According to market seeking and/or efficiency seeking FDI, multinationals would look for regions with good infrastructure as it is needed for market access and it leads to higher productivity. The empirical evidence usually supports a positive relationship between infrastructure variables and FDI.

Agglomeration Effects ($L.FDI_{ijt}$): proxied by one-year lagged FDI stock. A positive and significant coefficient of lagged FDI stock means the presence of foreign-specific agglomeration. According to the theory of agglomeration economies, once a country attracts the first mass of foreign investors, the process will become self-reinforcing, without needing a change in policies. The self-reinforcing effect of foreign investment allows new investors to benefit from positive spillovers by locating next to existing MNEs (Campos and Kinoshita (2003)).

It is worth to note that the above variables can be subject to sparse cross-dependencies and cross-correlations that may lead to collinearity problems in the estimation of the model giving to misleading results. In our case, the correlations matrix as well as some preliminary analysis indicated the presence of some collinearity. Hence, following Villaverde and Maza (2012), we implement an exploratory factor analysis (EFA) to reduce the initial number of explanatory variables to a set of non-collinear factors. Following Hair et al. (2010), firstly, we examine the data adequacy for conducting factor analysis by computing the Kaiser Meyer Olkin statistic (KMO). The KMO ranges from 0 to 1, with 0,50 considered the minimum threshold for a suitable for factor analysis. In our case, the KMO statistic is found to be 0.6 which indicates that the dataset is adequate for conducting factor analysis. Furthermore, in order to conduct EFA, the literature establishes that the sample size should be 100 or larger and that the ratio of observations per variable would be preferably 10:1 (Hair et al. (2010)). Our sample fulfills both criteria.

After diagonalizing the correlation matrix we obtain new variables, the factors, non-correlated among them. Deriving from the analysis of the eigenvalues and using as a criterium the cut-off value of greater than 1, we can get the number of factor to be retained (see Table 2.3). Our results show that the ten FDI determinants initially considered can be reduced to three significant factors explaining 62.07% of the cumulative variance of the nine original variables. As a next step, we examine the Pattern Matrix to identify the three factors with their constituent parts in Table 2.4.

When examining the factor loadings greater than 0.5 in absolute value we can easily identify their location and provide names for each factor. It can be observed that gdp per capita, labour productivity, human capital, wage and gross capital formation belong to the first factor. It appears that all the first factor elements look like they are directly related to the region's economic potential. That is why we call this factor *Economic or Market potential*. The second factor includes unemployment rate and road infrastructure . We call this factor *Productive capacity*. Finally, the third factor consists of trade openness, differences in relative factor endowments and the lagged FDI stock and we call it *Competitiveness and agglomeration effects (Comp.& agglom.)*. We can now extract each factor scores and run the regression model specification with the factors as additional determinants of FDI.[18] In the next Section we turn to the specification of the empirical model used in this paper, the empirical methodology and the results found.

[18]Notice that the results are interpreted considering these three new dimensions: region's economic potential, productive capacity and competitiveness and agglomeration effects.

Table 2.3 Factor analysis. Total variance explained.

Factors	Eigenvalue	% Variance	Cumulative % variance
1	**3.59615**	**0.3596**	**0.3596**
2	**1.35632**	**0.1356**	**0.4952**
3	**1.25446**	**0.1254**	**0.6207**
4	0.90239	0.0902	0.7109
5	0.84590	0.0846	0.7955
6	0.76595	0.0766	0.8721
7	0.69710	0.0697	0.9418
8	0.39292	0.0393	0.9811
9	0.15156	0.0152	0.9963
10	0.03725	0.0037	1.0000

Notes: The three extracted factors are shown in bold.

Table 2.4 Factor analysis. Rotated component matrix.

Variable	Factor 1 (Economic potential)	Factor 2 (Productive capacity)	Factor 3 (Comp. & agglom.)	Communalities
MS_{it}	**0.8650**	-0.4063	0.0596	0.9168
LP_{it}	**0.8044**	0.0717	0.0944	0.6611
HC_{it}	**0.8338**	-0.1360	0.0343	0.7149
OP_{ijt}	0.0706	-0.0240	**0.7553**	0.576
RFE_{ijt}	0.1053	0.1937	**-0.6216**	0.435
UR_{it}	-0.0627	**0.8012**	-0.1345	0.6639
W_{it}	**0.9231**	0.0907	-0.0043	0.8603
RI_{it}	-0.1524	**0.6808**	0.1400	0.5062
GCF_{it}	**-0.5113**	0.4095	-0.0053	0.4292
$L.FDIin_{ijt}$	0.3444	0.1739	**0.5429**	0.4436

Notes: The variables loading on each factor are shown in bold.

2.5 Econometric specification, methodology and estimation results

Although the gravity model (Anderson and van Wincoop (2003)) has been extensively employed by the trade literature to explain bilateral trade flows, its use to study bilateral FDI flows and FAS has been quite restricted until recently with the exceptions of the seminal papers by Eaton and Tamura (1994) and Wei (2000). One reason is that the transposition of the gravity model to study overseas investments was not supported by the theory. As our variable of interest is FDI and not FAS, we focus on the developments made by Head and Ries (2008), Bergstrand and Egger (2007) and Kleinert and Toubal (2010). In fact, Kleinert and Toubal (2010) refer to three different theoretical models of FAS to derive gravity equations that can yield an aggregate FDI equation. In particular, they rely on an horizontal model where firms can serve the foreign market j either by producing abroad or by exporting. They represent aggregate sales of foreign affiliates from firm i in j as follows:

$$n_i p_{ij} x_{ij} = n_i p_{ii}^{1-\sigma} \tau_{ij}^{(1-\sigma)(1-\in)} (1-\mu) Y_j P_j^{(\sigma-1)} \tag{2.1}$$

where n_i is the number of firms, p_{ij} the good price of firm i; x_{ij} is country j's consumption of variety from country i, τ_{ij} are the distance costs, Y_j the market size of country j and P_j the price index of country j.

According to Kleinert and Toubal (2010), the home country's market capacity can be denoted $s_i = n_i p_{ii}^{1-\sigma}$, while country's j's equivalent is $m_j = (1-\mu) Y_j P_j^{\sigma-1}$, and $AS_{ij} = n_i p_{ij} x_{ij}$ is bilateral foreign affiliates production. They express distance costs (τ_{ij}) as an increasing function of geographical distance between i and j, that is, $\tau_{ij} = \tau D_{ij}^{\eta_1}$.

Then, equation 2.1 can be rewritten as:

$$AS_{ij} = s_i(\tau D_{ij}^{\eta 1})^{(1-\sigma)(1-\in)} m_j \tag{2.2}$$

where τ represents the unit distance costs and $\eta_1 > 0$.

The gravity equation can be then obtained by log-linearizing equation 4.1:

$$ln(AS_{ij}) = \alpha_1 + \zeta_1 ln(s_i) - \beta_1 ln(D_{ij}) + \xi_1 ln(m_j) \tag{2.3}$$

where $\alpha_1 = (1-\sigma)(1-\in)ln(\tau)$ and $\beta_1 = (\sigma-1)(1-\in)\eta_1$.

This setting is enlarged using some additional variables following Blonigen and Piger (2014). Moreover, as the variable we are interested in is inward FDI, most of the relevant determinants should be related to the destination countries or regions (Blonigen et al. (2007)).

Silva and Tenreyro (2006) argue that estimating a log-linearized equation like the one shown in Equation 4.2 by OLS results in bias. The conditional distribution of the dependent variable is altered and estimation by OLS would produce misleading estimates, as the t-values of the estimated coefficients cannot be trusted.

They have proposed instead a PPML estimator which deals with this problem and provides consistent estimates of the original nonlinear model. The PPML estimator has a number of additional desirable properties. First, it is consistent under the presence of fixed effects; this is an important issue for the gravity approach since most theory-consistent models require the inclusion of fixed effects. Second, the PPML estimator naturally includes observations for which the FDI value is zero.

Actually, data in many country-pairs involve zero investment flows. In particular, in this study the proportion of zero inward FDI stock constitutes the $22,09\%$ of the total. Ignoring this number of zeros would lead to misleading results.

Model specifications like the one in Equation 4.2 do not incorporate absolute zero flows since the natural logarithm of zero is undefined and is consequently dropped. Therefore, we rely on PPML estimator proposed by Silva and Tenreyro (2006) which takes the following general form:

$$y_i = exp[X_i\beta]\varepsilon_i \tag{2.4}$$

In this equation, y_i is a dependent variable such that $y_i \geq 0$ and $E[\varepsilon_i|x_i]$. Since we are using a count variable as dependent, i.e. a variable that is discrete and non-negative, the PPML method is an appropriate estimator with an increasing recognition in the empirical literature. Therefore, we rely on this estimator in the present paper.

Following the above discussions, the PPML estimator in the context of the current study would take the following form:

$$\begin{aligned} &FDI_{ijt} = \beta_0 + \beta_{1k}X_{ikt} + \beta_2 Z_{ijt} + \beta_{3l}D_l + \psi_j + \psi_t + \varepsilon_{ijt} \\ &t = 1,...,T, k = 1,...,K, l = 1,...,L \end{aligned} \tag{2.5}$$

where FDI_{ijt} represents inward FDI stock received by region i from country j in any period t. Matrix X_{ikt} denotes all k FDI long-run macroeconomic determinants specific to the region and correspond to the factors *Economic potential* and *Productive capacity* more closely associated with HFDI, while Z_{ijt} contains bilateral determinants such as trade openness, differences in factor endowments and the lagged of FDI stock which are included in factor *Competitiveness and agglomeration effects* mostly related to VFDI. D_l stands for additional variables. We have augmented the analysis with a group of variables to capture not only the traditional gravity issues, but also the institutional differences, either internal or external, related to the European Union that are relevant from the point of view of the regions. We denote *Distance* the geographical distance between the reporting country and the specific region; *Landlocked* takes the value 1 when the region has not access to the sea, and 0 for coastal regions; *FIS* takes the value 1 for those regions with special fiscal regime (the Basque Country, Navarre and the Canary Islands) and 0 for the remaining ones; *OBJ1* is one for the regions Objective 1 according to the criteria of the European Structural Funds and 0 otherwise; *Capital* stands for a dummy variable that represent Madrid. We consider also the regional location quotient for the industry sector (*LQind*), measured as the relative share of industry GVA in the incumbent region compared to the national industry GVA share. Finally, *Crisis* is a dummy that captures the international financial crisis period. It takes the value 1 for the years 2008 and 2009 and 0 otherwise.

We also include country-origin fixed effects ψ_j to capture all those fixed effects of the investors, as well as time fixed effects ψ_t to control for business cycle effects over the sample period. ε_{ijt} is an error term such that $\varepsilon_{ijt} \sim N(0, \sigma^2)$. Note that β_{1k} and β_{3l} are two vectors of k and l coefficients, respectively, associated to the explanatory region-specific variables and the dummies.

More formally, our empirical specification and the expected signs of the FDI determinants are as follows:

$$FDI_{ijt} = f \begin{pmatrix} MS_{it} & LP_{it} & W_{it} & UR_{it} & HC_{it} & OP_{ijt} & RFE_{ijt} & RI_{it}/GCF_{it} & L.FDI_{ijt} \\ (+) & (+) & (+/-) & (+/-) & (+/-) & (+) & (+/-) & (+) & (+) \end{pmatrix} \tag{2.6}$$

Previously in Section 4 we have conducted an exploratory factor analysis in order to reduce the above-mentioned explanatory variables to a set of non-collinear factors. We identified the following factors: *Economic potential*, *Productive capacity* and *Competitiveness and agglomeration effects*.

Therefore, taking into account the theoretical and empirical surveys of the literature as well as the main stylized facts, the present study proposes the following testing hypotheses:

- For the explanatory variables:

  H1: *Economic potential* positively influences the decision of a MNE to invest, that is, $\beta_{11} > 0$)

  H2: The relationship between *Productive capacity* and inward FDI will be positive, so that $\beta_{12} > 0$.

  H3: *Competitiveness and agglomeration effects* has a positive influence on inward FDI, $\beta_2 > 0$.

- For the dependent variable:

  H4: FDI stock data are more appropriate than flows in econometric FDI analysis.

Table 2.5 reports the results for the PPML estimator for different model specifications including the factors in logarithms. Column (1) presents the estimated coefficients for the baseline model; columns (2)-(6) are alternative augmented versions of the basic model in order to test for additional FDI determinants, represented by the variables described above. According to the overall indicator for the model's "fit" , R^2, shows that all the specified models have a similar explanatory power of FDI (around $93-95\%$).

Taking a closer look at the estimated coefficients in column (1), FDI is positively and significantly related to the three factors: *Economic potential*, *Productive capacity* and *Competitiveness and agglomeration* of the regions. At this point it is worth to compare our results with earlier studies on regional FDI drivers. For the Spanish case, similar results were also found by Villaverde and Maza (2012). Although not explicitly using the same variables, they perform a factor analysis and found, like us, that economic potential, labour conditions and competitiveness are important factors for attracting FDI. More specifically, the factor they labelled "labour conditions" is comprised in our *Productive capacity* factor and the factor they coined as "competitivenes" contains a trade openness indicator which is linked to our *Competitiveness and agglomeration* factor. However, our study uses a different dataset and applies a distinct econometric methodology. Compared to other previous studies, we find that unlike Pelegrín (2002), infrastructure, which is embedded in our productive capacity factor, is found to be a significant determinant for FDI. Besides, the positive effect of human capital, a variable comprised in our *Economic potential* factor, is in line with previous literature (Pelegrín (2002) and Rodríguez and Pallas (2008)). Similarly, results for labour productivity are in line with those obtained by Rodríguez and Pallas (2008). Finally, our results support also the agglomeration or self-reinforcing effects of FDI as in Head and Ries (1996) or Cheng and Kwan (2000) for Chinese regions. Table 2.A.5 provides a summary of the Spanish regional FDI determinants.

Column (2) additionally includes two traditional gravity factors, *Distance* and *Landlocked*. Unexpectedly, *Distance* is found to have no significant impact on inward FDI unlike the traditional empirical literature on gravity models but the coefficient has the expected negative sign in models (4) to (6). On a second thought, these results seem quite sensible as the distance may be important at a national level, but at the regional one is not relevant to explain the heterogeneous location of FDI across regions. As for the estimated coefficient of the dummy variable *Landlocked*, this one is found to be negative, as expected, and statistically significant in models (4) to (6) once the headquarter effect is controlled for through a dummy variable for Madrid (*Capital*).

In column (3) we extend the model to include a dummy variable *FIS* to control for those regions with different fiscal system (i.e. the Basque Country, Navarre and the Canary Islands). We find that the dummy is significant but the sign is negative, so that differences in fiscal regime may cause a negative effect on FDI.

Table 2.5 PPLM Estimates of the Spanish inward FDI stock determinants, 2004-2013.

	Dependent variable: Inward FDI stock					
	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Economic Potential	0.313***	0.371**	0.391**	0.015	0.151	0.151
	(2.780)	(2.553)	(2.521)	(0.121)	(1.162)	(1.162)
Productive capacity	0.139**	0.139**	0.121*	0.048	0.110**	0.110**
	(1.975)	(1.972)	(1.808)	(1.098)	(2.323)	(2.323)
Comp. & agglom.	0.727***	0.736***	0.732***	0.773***	0.621***	0.621***
	(4.607)	(4.632)	(4.534)	(5.596)	(4.984)	(4.984)
Distance		0.068	0.071	-0.319	-0.214	-0.214
		(0.092)	(0.095)	(-0.433)	(-0.350)	(-0.350)
Landlocked		-0.191	-0.216	-2.162***	-2.621***	-2.621***
		(-1.129)	(-1.230)	(-4.002)	(-5.336)	(-5.336)
FIS			-0.421**	-0.140	-2.082**	-2.082**
			(-1.969)	(-0.672)	(-2.282)	(-2.282)
Capital				2.652***	7.393***	7.393***
				(4.197)	(3.296)	(3.296)
LQind					6.692**	6.692**
					(2.103)	(2.103)
Crisis						0.096
						(0.510)
R^2	0.939	0.938	0.940	0.954	0.957	0.957
Investing country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	218	218	218	218	218	218
RESET test $p-values$	0.5116	0.8925	0.9956	0.2678	0.2421	0.2421
AIC	200769.8	197324.5	196002.3	145981.1	132267.3	132267.3

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Factors are included in logarithms. OBJ1 was included in models (3) to (6), but it is dropped because of multicollinearity.

This result, that may initially seem striking as counterintuitive, can be somewhat justified when we include the dummy variable accounting for the headquarter effect as FDI attractor. We will comment on that in the next paragraph. Previous empirical evidence has introduced also fiscal variables in the analysis and obtained a significant and negative effect on FDI (see Rodríguez and Pallas (2008)). Additionally, in column (3) trying to capture other institutional effects, we also consider the role that the European structural policies may have on the regions and whether the affluence of European funds may stimulate FDI toward these regions. We have denoted this variable *OBJ1*. Yet, it is dropped from the analysis due to the existence of multicollinearity.

Column (4) includes a dummy for the capital city where the large majority of the MNE headquarters are located: Madrid (*Capital*). In columns (4) to (6) the dummy for Madrid has a positive and very significant influence on inward FDI. We should also emphasize that once this dummy is included, the effect of the fiscal variable (*FIS*) disappear. Moreover, both the first factor (*Economic potential*) and the second (*Productive capacity*) become insignificant. Yet, the latter becomes significant in models (5) and (6). The third factor, *Competitiveness and agglomeration* remains positive and significant.

In an attempt to capture industry-specific specialization, column (5) includes the regional location quotient for the industry sector (*LQind*), calculated by comparing the relative share of industry's GVA in the region with the national share. We find a positive and significant effect in models (5) and (6) in line with the evidence found in Copenhagen Economics (2006) and Dimitropoulou et al. (2013). Thereby, the higher the degree of regional specialization in the industry sector is, the higher the attraction for FDI. This variable is consistent with the idea that positive externalites, such as knowledge creation (learning and innovation) and knowledge transfer (diffusion and synergies), that arise with the agglomeration of firms specialized in a particular sector are seen as an attractive factor for MNEs location. Indeed, the largest recipient regions of inward FDI (i.e. Madrid, Catalonia, Galicia, Asturias and the Basque country, as shown in Figure 2.2) are those that accounted for around 94% of FDI directed towards the industry sector in 2013 (see Table 2.1). Furthermore, this variable is also providing indirect evidence of a complementary relationship between FDI and trade as those sectors with high location quotient are usually export-oriented.

Finally, the global financial crisis is found to be non-significant (see column (6)), suggesting some degree of persistence in the behaviour of the FDI stock: although FDI flows decreased during the crisis, the stock has maintained its total size, as reflected in the aggregated data shown in Figure 2.1.

Overall, the models in columns (4) to (6) present the highest R^2 and the lowest *AIC* becoming candidates to be our chosen specifications.

We replicate the estimations for FDI inflows for the sake of comparison. Results are presented in Table 2.6. Even though we do not consider all the alternative models to be strictly comparable, the interpretation of the estimation results is in a similar vein. We find that *Economic potential* is a positive and significant determinant of FDI. However, the second factor (*Productive capacity*) and the third one (*Competitivenes and agglomeration*) appear to be non-significant. Gravity factors included in column (2) have the expected negative sign but are statistically non-significant in this specification. Yet, *Landlocked* becomes significant in models (3) to (6). FIS also has a negative and significant impact on FDI in columns (3) to (6) and, as before, *OBJ1* is dropped from the models because of multicollinearity. Columns (4) to (6) confirm that Madrid attracts more FDI in line with the FDI regional distribution presented previously. Furthermore, *LQind* remains positive and significant in models (5) and (6). Lastly, the dummy variable included in model (6) representing the financial crisis has, this time, a significant effect but the sign is positive. This is not surprising, as the financial crisis affects the flows of FDI but the stock is only affected in a lesser extent.

The Ramsey (1969) Regression Equation Specification Error Test (RESET) can be considered a general misspecification and omitted variables test for the estimated models, both for the FDI stock and flow specifications. This is essentially a test for the correct specification of the conditional expectation, by testing the significance of an additional regressor constructed as $(x'b)^2$, where b denotes the vector of estimated parameters (Silva and Tenreyro (2006)). The corresponding p-values are reported at the bottom of Tables 2.5 and 2.6, respectively. In the specifications using FDI inflows, the test rejects the null hypothesis of a good specification. This means that these models are either inappropriate due to its functional form or that some relevant information is missing. In contrast, models estimated using FDI stocks clearly pass the RESET test. Thereby, the RESET test suggests that, for our empirical specification, FDI stock data are more appropriate than flows for the correct specification of the model of FDI long-run determinants.[19]

[19] We have also repeated the exercise of all the before-mentioned estimations applying OLS instead of PPML as a robustness check and the results obtained point to a superior performance of the PPML method compared to OLS. All these results are not reported in the paper but are available from the authors upon request.

Table 2.6 PPLM Estimates of the Spanish FDI inflow determinants, 2004-2013.

	Dependent variable: FDI inflow					
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Economic Potential	1.064***	1.347***	1.577***	1.001***	1.488***	1.488***
	(5.620)	(4.951)	(4.522)	(3.478)	(4.106)	(4.106)
Productive capacity	-0.126	-0.144	-0.183	-0.169	-0.084	-0.084
	(-0.636)	(-0.668)	(-0.833)	(-0.771)	(-0.348)	(-0.348)
Comp. & agglom.	0.141	0.091	-0.017	-0.006	-0.200	-0.200
	(0.680)	(0.352)	(-0.061)	(-0.022)	(-0.790)	(-0.790)
Distance		-1.098	-1.188	-1.502	-2.158	-2.158
		(-0.855)	(-0.812)	(-0.997)	(-1.234)	(-1.234)
Landlocked		-0.366	-0.615**	-2.504***	-4.138***	-4.138***
		(-1.148)	(-2.314)	(-4.870)	(-5.136)	(-5.136)
FIS			-3.607***	-3.136***	-8.549***	-8.549***
			(-6.421)	(-6.368)	(-4.062)	(-4.062)
Capital				2.703***	15.866***	15.866***
				(3.974)	(3.291)	(3.291)
LQind					18.397***	18.397***
					(2.649)	(2.649)
Crisis						2.412***
						(3.244)
R^2	0.737	0.740	0.734	0.734	0.738	0.738
Investing country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	244	244	244	244	244	244
$RESET$ test $p-values$	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
AIC	62199.69	61302.97	59363	58274.53	55601.82	55601.82

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Factors are included in logarithms. OBJ1 was included in models (3) to (6), but it is dropped because of multicollinearity.

2.6 Concluding remarks

In the last decades, understanding the factors that determine FDI has attracted an intense academic and policy-oriented interest. This paper feeds the discussion conducting an empirical investigation to identify the main driving forces for FDI activity directed towards the Spanish regions. In addition to the regional dimension itself, this paper makes three contributions to the literature: first, our research represents one of the first attempts to find the main determinants of inward FDI into Spanish regions using FDI stock instead of flows. We have compared, for the sake of robustness, the estimation results using the two definitions of FDI and concluded that the stock model was a superior model specification. Second, we use the PPML estimator which produces unbiased and consistent estimates when the dependent variable includes a large proportion of zero observations, as it turns out to be our case. Third, we have implemented an exploratory factor analysis to reduce the number of explanatory variables and avoid collinearity problems in the estimation.

We have identified three factors as FDI determinants that, according to the variables they include, were labeled as *Economic potential*, *Productive capacity* and *Competitiveness and agglomeration*. The first two factors are more closely related to HFDI while the third one is to VFDI. The empirical analysis revealed the following allocation patterns: as expected, FDI locational strategies in the Spanish regions are determined significantly by the relative competitiveness of the regions, the agglomeration effects and, to a lesser extent, by the productive capacity or the market size. This is a sensible outcome since FDI at the regional level cannot be expected to be market seeking but an efficiency seeking one, intending to fit in the global value chains developed by the multinational production strategies of transnational companies. Therefore, at a regional level the FDI drivers are linked to vertical strategies, where endowment differentials and trade openness have shown themselves to be vital. This finding supports the view of a complementary relationship between trade and FDI, a result common to the majority of empirical studies on trade-FDI linkages.

The agglomeration or self-reinforcing effect of FDI is found to be an important driver for FDI location. In the presence of agglomeration, investors decide to locate near existing MNEs in order to benefit from positive externalities such as knowledge spillovers, specialized labor, and intermediate inputs. Our results also show that the degree of industrial specialization of the regions is a major driving force for MNEs locational choice. The agglomeration of firms specialized in a particular sector generates positive externalities that influence the attractiveness of that region as a potential FDI location as opposed to competing regions without industry clustering. Furthermore, our analysis points out that FDI location depends upon the geographical position of each region, stressing the importance of coastal areas; this means that transport infrastructures and interconnectedness matter as FDI attractors.

Finally, although a sectoral breakdown of FDI stock data is not available and hence, we are not able to conduct a more disaggregated analysis, the findings in this paper give some clues about the factors that regions should emphasize in order to attract FDI. Results point out to regional competitiveness and agglomeration factors as the most efficient ways for a region to attract FDI. More specifically, given the loading of each variable inside this factor, we think that future policies should promote internationalization as well as factor endowment improvements. From a regional policy perspective, industry cluster formation is also important to attract foreign investors. Furthermore, a region wishing to attract FDI should promote policies encouraging an adequate transport infrastructure and the quality of labor by increasing human capital and labour productivity.

Appendix

Table 2.A.1 Data description and source

Variable name	Description	Source
$FDIin_{ijt}$	Inward FDI stock, in millions of euros (current values)	DataInvex[a]
$FDIinflow_{ijt}$	Inward FDI flows, in millions of euros (current values)	DataInvex
MS_{it}	Gross Domestic Product per capita (current EUR)	INE[b]
LP_{it}	GDP (current EUR) per total employment	INE
W_{it}	Employee's compensation per hour worked, calculated as the ratio of total compensation (current EUR) per hours worked	Eurostat
UR_{it}	Unemployment rate	INE
HC_{it}	Share of population with Tertiary Education[c]	Ivie[d]
OP_{ijt}	Sum of bilateral exports and imports as a share of GDP(%), in millions of euros (current values)	DataComex [e]
RFE_{ijt}	GDP per capita difference between host region and parent country (current EUR)	INE, World Bank
RI_{it}	Motorways kilometers	Eurostat
GCF_{it}	Gross capital formation (% of GDP) (current values)	Ivie, INE
$L.FDIin_{ijt}$	1-year lag of inward FDI stock, in millions of euros (current values)	DataInvex
Distance	Log of distance between region capital for Spain and national capital for the parent country, in km	Own elaboration
Landlocked	1 if host region is landlocked and 0 otherwise	Own elaboration
FIS	1 for those regions with special fiscal regime (the Basque Country, Navarre and the Canary Islands) and 0 otherwise	Own elaboration
OBJ1	1 if host region is considered Objective 1 according to the criteria of the European Structural Funds and 0 otherwise	European Commission[f]
Capital	1 for Madrid and 0 otherwise	Own elaboration
Crisis	1 for the years 2008 and 2009 and 0 otherwise	Own elaboration
LQind	Log of the regional location quotient for the industry sector, measured as the relative share of industry gross value added (GVA) in the incumbent region compared to the national industry GVA share.	INE

[a] Spanish Ministry of Economy and Competitiveness

[b] Spanish Statistical Institute

[c] Tertiary Education includes "FP II" , "Anteriores al Superior" and "Superiores" .

[d] Fundación Bancaja e Ivie (Instituto Valenciano de Investigaciones Económicas). Capital Humano en España y su distribución provincial. Enero de 2014. Database available on: http://www.ivie.es/es/banco/caphum/series.php

[e] Spanish Ministry of Industry, Tourism and Trade

[f] Directorate-General for Regional and Urban Policy

Table 2.A.2 Descriptive statistics

Variable		Mean	Std.dev.	Min.	Max.	Obs.
$FDIin_{ijt}$	Overall	331.3704	2084.274	0	38082	N=6800
	Between		2023.563	0	29193.1	n =680
	Within		504.7879	-8595.33	11864.77	T=10
$FDIinflow_{ijt}$	Overall	21.18571	312.3285	0	18021.7	N=6800
	Between		146.0353	0	2591.995	n =680
	Within		276.1358	-2287.903	16147.48	T=10
MS_{it}	Overall	22263.52	4389.056	13117.98	32151.92	N=6800
	Between		4204.211	15421.66	30298.7	n=680
	Within		1269.576	18085.13	24596.31	T=10
LP_{it}	Overall	58628.89	26651.64	36836.03	188178.9	N=6800
	Between		25215.97	44311.21	156920.2	n=680
	Within		8677.923	-26979.1	89887.63	T=10
HC_{it}	Overall	0.1950035	0.0429806	0.1134766	0.3000077	N=6800
	Between		0.0408146	0.1388397	0.2862831	n=680
	Within		0.0135538	0.157349	0.2263087	T=10
OP_{ijt}	Overall	0.0078201	0.0163329	2.43e-08	0.1848764	N=6800
	Between		0.0160351	1.63e-06	0.1581036	n=680
	Within		0.003159	-0.0210068	0.054379	T=10
RFE_{ijt}	Overall	13959.61	10817.48	3.105469	70319.9	N=6800
	Between		10167.72	638.0404	59443.57	n=680
	Within		3711.085	-2557.866	31233.15	T=10

(Continued)

Table 2.A.2 Descriptive statistics *(Continued)*

Variable		Mean	Std.dev.	Min.	Max.	Obs.
UR_{it}	Overall	14.87141	7.513715	4.72	36.22	N=6800
	Between		3.804143	9.797	22.764	n=680
	Within		6.481016	4.006412	28.32741	T=10
W_{it}	Overall	14.34223	2.142297	9.782425	19.53279	N=6800
	Between		1.683675	12.09357	17.93362	n=680
	Within		1.326058	11.47515	15.96533	T=10
RI_{it}	Overall	784.1412	636.51	57	2462	N=6800
	Between		626.2473	88.8	2282.4	n=680
	Within		116.0968	247.9412	1063.441	T=10
GCF_{it}	Overall	36.0575	11.6485	18.9108	59.72894	N=6800
	Between		11.39165	19.94766	54.45922	n=680
	Within		2.467734	30.85473	43.03918	T=10
$L.FDIin_{ijt}$	Overall	329.4655	2100.268	0	38082	N=6120
	Between		2042.544	0	29769.78	n=680
	Within		494.5683	-8157.646	11997.58	T=9

Notes: All the variables are in levels.

Table 2.A.3 Cross-correlation table

Variables	MS_{it}	LP_{it}	HC_{it}	OP_{ijt}	RFE_{ijt}	UR_{it}	W_{it}	RI_{it}	GCF_{it}	$L.FDIin_{ijt}$
MS_{it}	1.000									
	6800									
LP_{it}	0.6181*	1.0000								
	(0.0000)									
	6800	6800								
HC_{it}	0.7695*	0.5577*	1.000							
	(0.000)	(0.000)								
	6800	6800								
OP_{ijt}	0.1184*	0.0427*	0.1341*	1.000						
	(0.000)	(0.0004)	(0.000)							
	6800	6800	6800	6800						
RFE_{ijt}	-0.0184	0.0064	0.0096	-0.1706*	1.000					
	(0.1292)	(0.5993)	(0.4292)	(0.000)						
	6800	6800	6800	6800	6800					
UR_{it}	-0.3836*	-0.0560*	-0.1670*	-0.0731*	0.0819*	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)					
	6800	6800	6800	6800	6800	6800				
W_{it}	0.8017*	0.6061*	0.7989*	0.0867*	0.0254*	0.1393*	1.000			
	(0.0000)	(0.0000)	(0.000)	(0.0000)	(0.0364)	(0.0000)				
	6800	6800	6800	6800	6800	6800	6800			
RI_{it}	-0.2655*	-0.0370*	-0.2047*	0.0120	0.0208	0.2700*	-0.0831*	1.000		
	(0.000)	(0.0023)	(0.0000)	(0.3228)	(0.0869)	(0.000)	(0.0000)			
	6800	6800	6800	6800	6800	6800	6800	6800		
GCF_{it}	-0.5766*	-0.3777*	-0.2195*	-0.0333*	0.0300*	0.2150*	-0.3339*	0.2423*	1.0000	
	(0.000)	(0.000)	(0.000)	(0.0060)	(0.0134)	(0.000)	(0.0000)	(0.000)		
	6800	6800	6800	6800	6800	6800	6800	6800	6800	
$L.FDIin_{ijt}$	0.2075*	0.3473*	0.1697*	0.2198*	-0.0594*	-0.0301*	0.2086*	0.0109	-0.1536*	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.0185)	(0.0000)	(0.3930)	(0.0000)	
	6120	6120	6120	6120	6120	6120	6120	6120	6120	6120

Table 2.A.4 Countries included in the study

Source countries				
OECD				
Australia	Estonia	Ireland	Netherlands	Slovenia
Austria	Finland	Israel	New Zealand	Sweden
Belgium	France	Italy	Norway	Switzerland
Canada	Germany	Japan	Poland	Turkey
Chile	Greece	Korea, Republic of	Portugal	United Kingdom[g]
Czech Republic	Hungary	Luxembourg	Slovak Republic	United States
Denmark	Iceland	Mexico		
Non-OECD				
Argentina	China	Hong Kong	Singapore	
Brazil	India	Russian Federation		

[g]Since the 2016 referendum vote to leave the EU, the UK is on course to leave the EU.

Table 2.A.5 Summary of Spanish regional FDI determinants

Author(s)(year)	Level	Dependent variable	Time period	Estimation technique	Determinants	Effect
Pelegrín, A. (2002)	Regional	FDI inflows/GDP	1998-2000	OLS	Market size	(+)
				LSDV	Human capital	(+)
				GLS	Public incentives	(+)
					Labour cost	(+)
					Infraestructure	(0)
Pelegrín, A. and Bolancé, C. (2008)	Regional and sectoral	FDI inflows	1995-2000	Random Effects	Same industry activity	(+)
					Comparative Advantage	(+)
					R&D agglomeration	(+)
					Wage	(-)
Rodríguez, X.A. and Pallas, J.(2008)	Regional and sectoral	FDI inflows	1993-2002	GLS	Demand factors	(+)
				W2SLS	Labour productivity and its cost differential	(+)
					Human capital	(+)
					Export potential	(+)
					Fiscal pressure	(-)
					Inflation differential	(-)
Villaverde, J. and Maza, A. (2012)	Regional and sectoral	FDI inflows/GDP(POP)	1995-2005/2008	Factor Analysis	Economic potential	(+)
				GLS	Labour conditions	(+)
				Two-stage GLS	Competitiveness	(+)
					Market size	(0)
Gutiérrez-Portilla, P. et al. (2016)	Regional and sectoral	FDI inflows/GDP	1997-2013	GMM	Market size	(+)
				GLS	Human capital*Wages	(+)
					Capital dummy	(+)
					Lag FDI	(+)

Notes: (+) and (-) denote a positive and negative statistically significant effect, respectively. (0) denotes no statistically significant effect. *Source:* Own elaboration.

Chapter 3

What drives German Foreign Direct Investment? New evidence using Bayesian statistical techniques

3.1 Introduction and motivation

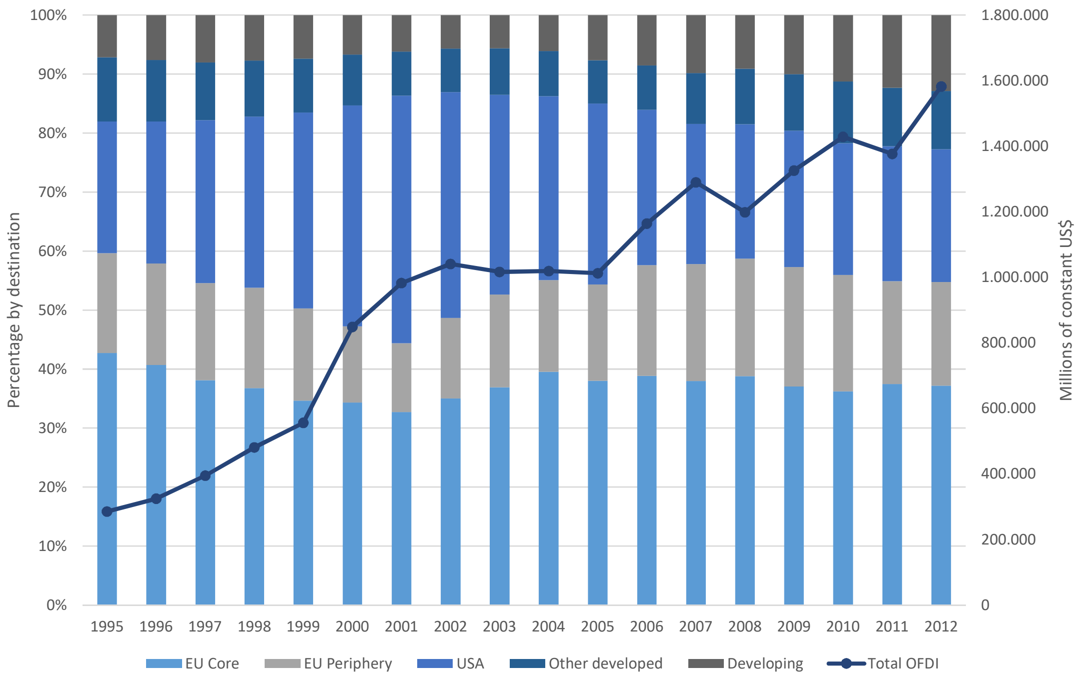
German direct investment abroad has increased significantly over the last decades. In 2016, Germany ranked the third largest investor among developed economies measured by FDI stock, only behind the United States and the United Kingdom (UNCTAD (2018)). In comparison with other OECD countries, Germany shows a clear and persistent outward orientation. Indeed, Germany's accumulated outward investment (equivalent to 40% of GDP in 2015) was much larger than its inward stock (23% of GDP in 2015) (OECD (2017)).

The substantial increase in direct investment of German enterprises abroad has developed largely in parallel with international trade, both intra European Union (EU hereafter) and with third countries. Therefore, the relationship between trade and FDI has become increasingly complex, giving rise to a conundrum with intricate issues. Trade-FDI linkages have been of continued interest in the empirical literature, where the recent results generally point to a complementarity relationship between trade and FDI, challenging to some extent the assumptions of the neoclassical trade model (see Camarero and Tamarit (2004) and Camarero et al. (2018)). However, some studies have identified a substitution relationship between exports and outward FDI for Germany (see Egger and Pfaffermayr (2004a) and Mitze et al. (2010)). Surveying several empirical contributions, Yang and Mallick (2014) show in a recent meta-paper the key role played by the returns to FDI to explain the relationship between multinationality and exporting, which is in line with the current outbreak of the Global Value Chains (GVC) and the internationalization of production.

Indeed, Germany's global investment and its engagement in international trade are closely linked to its participation in the international production networks. There appears to be three interconnected production hubs of GVCs in the world: North America (especially the United States as core country), East Asia (i.e. China, Japan and the Republic of Korea) and Europe (centered in Germany)(World Bank Group et al. (2017)). In particular, three German key industries— motor vehicles, machinery and equipment and chemicals—drive Germany's integration in complex international production networks through sourcing of intermediates from abroad (often from Central European Countries). In contrast, Germany's participation in services GVCs is driven by downstream links and the use of German intermediates in the exports of other countries (OECD (2017)).

Despite Germany's participation in GVCs, outward foreign direct investment (OFDI) of German firms is mainly concentrated in OECD countries with a special emphasis within the European Union (EU). Around 87% of the value of OFDI from Germany was held in developed countries at the end of 2012 and the EU accounted for more than half of the overall stocks. The distribution of the stock of German OFDI by regions of destination is shown in Figure 3.1. One striking feature of German OFDI is the continuos upward trend since the mid-1990s. This growth is explained mainly by the process of world globalization, the liberalization of Eastern Europe in the 1990s and the progress of EU integration.

Fig. 3.1 Germany: geographical distribution of outward FDI stock,1995-2012.

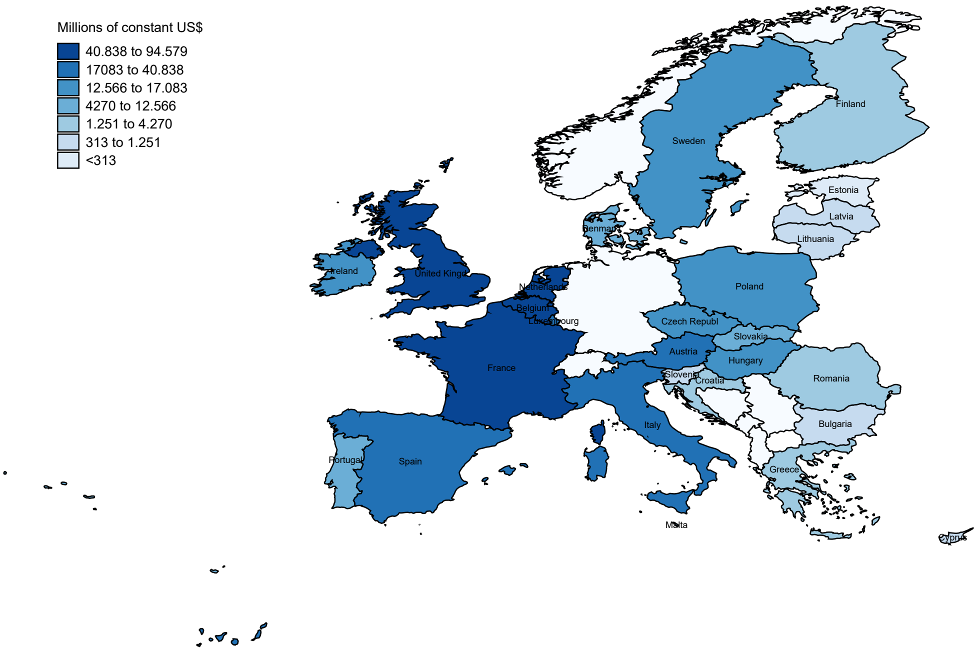


Source: UNCTAD database.

Indeed, the European Monetary Union and the introduction of the Euro in 1999 increased sharply German OFDI. Whereas the Euro area is the largest regional location for German MNEs (as it accounts for 35% of the stock in 2012), non-euro area countries accumulated only 19%. Concerning other developed countries, the United States is the largest recipient, followed by Switzerland. By contrast, OFDI in developing countries has so far not been a significant target of German MNEs activities, with the recent exception of China. Within the European Union, German OFDI distribution displays a core-periphery pattern, as shown in Figure 3.2 . The "core" accounts for 68% of Germany's total EU investment. Specifically, France, the UK and the Benelux are the most important destinations throughout the period shown. The South and North Periphery are also important locations for MNEs whereas

Eastern European countries occupied a less prominent position. A possible explanation for this difference is that German FDI increased once the latter countries joined the EU in 2004; before then, the amount of FDI was negligible.

Fig. 3.2 German FDI stock distributed across European Union, average 1995-2012.



Source: UNCTAD database.

These stylized facts raise the question of whether German MNEs have different internationalization strategies depending on the regions or areas where they operate and whether they have changed over time. Additionally, finding the robust determinants of German OFDI can provide hints for other countries to develop policy strategies helping to attract German investment.

The interest of the present study lies in two main considerations. First, the relative scarcity of studies analyzing German FDI determinants in spite of its relevance as a worldwide investor. Some exceptions are Buch et al. (2005), Egger and Pfaffermayr (2004a) and Antonakakis and

Tondl (2015). Buch et al. (2005) examines firm-level data to describe regional and sectoral patterns of FDI. Egger and Pfaffermayr (2004a) analyzes the effects of distance as a common determinant of exports and FDI in a three-factor New Trade Theory model. Antonakakis and Tondl (2015) conduct a comparative study to examine the determinants of outward FDI from four major OECD investors (the US, Germany, France, and the Netherlands) applying a Bayesian Model Averaging (BMA) approach. They constrain the analysis to developing countries as FDI destinations.[1]

A second consideration justifying the interest of the present study is that, although the literature provides a vast amount of applied research on FDI location, its main determinants are still poorly understood.[2] The reason for this is the uncertainty and ambiguity surrounding both theories and empirical approaches to FDI. Traditionally, empirical studies have controlled for the factors considered more explanatory according to certain theories usually using an *ad hoc* approach. If this is the case, the selection of variables can lack statistical motivation and disregard potentially important covariates. Moreover, a theoretical economic variable can be measured, in applied work, using different definitions. The way we discriminate among potential covariates is indeed a relevant question, as it can substantially affect the estimation results and possibly lead to multicollinearity and omitted variable problems (Blonigen (2005)). As Eicher et al. (2012) stated, numerous empirical studies estimate only subsets of particular FDI theories to produce results that are often either inconclusive or outright contradictory. When model uncertainty is not addressed comprehensively as part of the empirical strategy, traditional robustness analyses overstate significance levels and

[1] Although they also study the German case, we differ from them in three main respects: first, we focus on one single investor (Germany) and consider different destination country-groups (not only developing ones) with a geographical breakdown; second, we consider a wider set of FDI potential determinants and the most recent period available for bilateral FDI stock data, and third, we apply a different Markov Chain Monte Carlo (MCMC) method in the BMA analysis.

[2] See Blonigen (2005), Faeth (2009) and Assunçao et al. (2011) for a thorough literature review on FDI determinants.

confidence intervals. Therefore, for statisticians it is clear that we need robust statistical methods to select variables and discriminate among them.

The literature that deals with model uncertainty when there is little guidance from economic theory regarding which explanatory variables to consider goes back to Raftery (1995). Each model is defined by the specific subset of variables it includes and is treated as an unknown parameter that lies in the set of models entertained (the model space). Bayesian inference offers the tools to attach probabilities to the different possible models. Raftery (1995) showed that when there are many candidate independent variables, standard model selection criteria based on p-values can be misleading and he promoted the use of Bayesian inference to take into account model uncertainty explicitly. There is a steadily increasing bulk of empirical studies making use of the BMA methodology to address the identification of robust determinants in different contexts (see, among others, Man (2015), Ng et al. (2016), Wei and Cao (2017), Pham (2017) and Desbordes et al. (2018)). The uncertainty surrounding FDI modelling makes the BMA methodology especially suited to discriminate among the large set of candidate regressors that has been posited as possible FDI determinants by different theories. Chakrabarti (2001) was the first to put forward this uncertainty in FDI studies using Extreme Bound Analysis. More recently, Blonigen and Piger (2014) and Eicher et al. (2012) use a BMA approach to account for model uncertainty in FDI and show that the robust FDI specification is a more parsimonious one than that previously suggested in the literature. Similarly, we do not condition our study of German OFDI on a single model and instead attach probabilities to different models and thereby identify, in a robust manner, the determinants of FDI.

The aim of this paper is to use robust statistical techniques to select the main determinants of German outward FDI. To this end, we apply a probabilistic model to select the explanatory variables from a large group of potential candidates. However, given the worldwide distribution of German FDI stocks and taking into account the heterogeneity of the destinations, trying to find a general specification would be probably incorrect. Instead, we consider different country-groups taking into account German OFDI distribution. First, we distinguish between developed and developing recipient countries. Then, within developing countries, we distinguish between Latin American and Asian countries. Finally, we focus on FDI allocation within Europe and separate "core" and peripheral countries. Our aim is to shed some light on whether German internationalization strategies differ depending on the characteristics of the countries where they operate. In our case we apply the methodological approach proposed by Bayarri et al. (2012), among others, that was made available in a user-friendly R package developed by García-Donato and Forte (2015) and applied in Camarero et al. (2015) to the case of energy consumption and growth. In the present study, we adopt a BMA approach to select the main determinants of the stock of German OFDI covering the most recent period available for FDI stock data—1996 to 2012— and take into account the heterogeneity of the destinations of German FDI. Although this variable selection exercise is relevant by itself, we also provide some inference for the estimated coefficients of the models considered.

Our analysis shows that robust specification of German OFDI drivers is more parsimonious that previously suggested by the literature. Although the determinants identified highlight that the decision to invest abroad is based on a mixture of FDI theories, we can infer from the results that German investment decisions differ significantly across regions. In particular, we find evidence that determinants that are associated with horizontal FDI motives appear to be dominant for explaining bilateral FDI with developed countries while for the group of developing countries determinants associated with vertical FDI motives play a larger role. Regarding European Union countries, we find evidence pointing to the benefits associated with the proximity to large euro area markets for "core" countries, while for peripheral countries the determinants identified are mostly associated with vertical FDI motives. Finally, we find some evidence of the relevance of world GVC to explain the

FDI determinants in the different areas, from classical horizontal gravity variables, to more complex relationships that include vertical determinants and institutional variables.

The reminder of the paper is organized as follows: in Section 2 we briefly review the main theoretical approaches to FDI determination, with an emphasis in the formulated hypotheses and their differences; Section 3 presents the empirical literature on FDI classified according to the theories described in the previous section, with the purpose of completing the list of potential groups of explanatory variables; we provide a summary of the BMA methodology in Section 4, whereas Section 5 describes the data together with the results; finally, Section 6 concludes.

3.2 A Brief Literature Review

As we have already mentioned in the Introduction, there is remarkably little consensus on which variables are the most salient in explaining bilateral FDI patterns. In this section we provide a brief survey of the existing FDI theories, the hypotheses formulated and the variables that are most frequently proposed as FDI determinants. We will use the same classification of the theoretical literature proposed in this section to review the empirical work in Section 3 with the purpose of composing a wide set of potential explanatory variables for FDI.

Traditional trade economics, ownership advantages and internalization theory were brought together by Dunning (1977, 1979) who created the well-known OLI framework.

"OLI" stands for Ownership, Location and Internalization advantages, which are three types of special advantages that multinational enterprises (MNEs) have. *Ownership advantages* concern the importance of a MNE owning firm-specific assets which allow it to overcome the costs of operating in a foreign country. Examples include both tangible and intangible assets, such as pioneering technology, specific know-how and management skills. *Locational advantages* refer to all those factors a specific location owns that make it eligible for a MNE. Finally, *Internalization advantages* refer to those kinds of advantages that make more profitable for a firm to carry out transactions internally (i.e.through a wholly-owned subsidiary) rather than other entry modes such as exports, through licensing or joint-venture agreements. Focusing on locational advantages, the relevant to invest and produce abroad, Dunning (2000) distinguishes between four main motives for FDI: market seeking, resource seeking, efficiency seeking and strategic assets seeking. Market seeking motives correspond to FDI aiming at supplying the local market or markets in adjacent territories. The host market size, its per capita income and consumer demand (all of them to take advantage of the economies of scale) are the main reasons behind market seeking FDI. Resource seeking companies are those investing abroad in order to obtain cheap natural resources and/or unskilled labour. Hence, locational decisions depend on factor endowments differences. Efficiency seeking investment is designed to promote a more efficient division of labor or specialization of assets by MNEs. Finally, strategic asset seeking FDI searches for resources such as technology, skilled workers and assets that can support development of a firm's worldwide and weaken the competitive position of its competitors.

An alternative framework for analyzing FDI arose in the mid-80s with the New Trade Theory which incorporated multinational firms into the general equilibrium trade models. This theory explained the existence of two types of FDI, namely horizontal (market-oriented) and vertical (export-oriented) FDI. The early formal FDI theory put forward byHorstmann and Markusen (1987) shows that MNEs undertaking horizontal FDI (HFDI) face the trade-off between maximizing proximity to customers and concentrating production to achieve scale economies. This is what Brainard (1997) called the proximity-concentration hypothesis.

On the other hand, Helpman (1984, 1985)) showed that countries' differences in relative factor endowments (the so-called factor-proportions hypothesis) explained vertical FDI

(VFDI). Thus, the decision of firms to engage in HFDI would be driven by the size and growth of the host country whereas VFDI seeks for cost competitiveness and other factors such as institutions and infrastructure in the host country.

Combining these two motivations for FDI, Ethier and Markusen (1996) and Markusen and Maskus (2002) formulated the *knowledge-capital model*. According to this theory, similarities in market size, factor endowments and transport costs are determinants of HFDI, while differences in relative factor endowments determine VFDI. The *knowledge-capital model* has recently been extended to explain other forms of FDI such as export-platform FDI (see Ekholm et al. (2007); Bergstrand and Egger (2007)) which is used to serve the neighboring markets of the host country. These studies highlight the importance of considering regional trade agreements in the empirical approach. Additionally, Bergstrand and Egger (2007) and Head and Ries (2008) provide theoretical foundations that motivate the use of gravity equations to analyze FDI patterns. The gravity model states that the closer two countries are (geographically, economically and culturally) the higher will be the FDI activity between them. Thus, geographical location and the size of the country are considered, according to this approach, the main explanatory variables. Kleinert and Toubal (2010) provide the theoretical underpinnings of the gravity equation applied to the analysis of Foreign Affiliate Sales (FAS) showing that gravity equations can be used to discriminate between different theoretical approaches, namely, two proximity-concentration models of HFDI with homogenous (Brainard (1997)) or heterogenous firms (Helpman et al. (2004)) and a two-country factor proportions model of fragmentation that explains VFDI and based on Venables (1999).

Heterogenous firms also explain the behavior of an additional type of MNEs, that is, diversified MNEs. According to the risk diversification hypothesis, firms that are assumed to be risk averse, try to spread business risk. Moreover, based on the heterogenous-firm trade theory of Helpman et al. (2004), there have been extensions to explain how these firms expand into overseas markets either through exports or FDI. According to this approach, firms which invest in foreign markets are more productive. Hence, greater productivity and heterogeneity lead to more FDI sales relative to export sales. MNEs can shift investments between home and host countries when they are faced with changes in the macroeconomic conditions of the domestic or foreign country. Determinants suggested by this theory are, for instance, interest rates, inflation rates or exchange rates. Overseas production can therefore be a substitute for exporting (see Lankhuizen et al. (2011)). Since Helpman et al. (2004), a large body of literature has arisen focusing not only on which parent firms will choose to engage in FDI but also which domestic firms are acquired. Guadalupe et al. (2012) shows that the most productive firms ex-ante will be acquired by MNEs and this increases productivity (i.e. innovation) and exports. Higher productivity levels are mainly due to the fact that acquired firms can benefit from the access to the global market through exporting to their parent firm.

Finally, the institutional approach highlights the important role played by policy variables (such as corruption, corporate tax rates, tax concessions but also the degree of political rights and civil liberties) together with other fiscal and financial investment incentives on attracting FDI. Making a decision about investment in a foreign country requires a multidimensional evaluation. Accordingly, the expectation of earning profit in the country where investment takes place is determined by economic, social and political factors which make FDI a complex issue. In the literature that examines the institutional quality and FDI relationship, from a theoretical point of view, it is commonly accepted that low institutional quality will negatively affect the investment choices by creating a risk factor (i. ex. bribes, low law enforceability, rent-seeking activities...). However, the opposite can also occur, as corruption can speed up bureaucratic processes or gain access to publicly funded projects (see Egger and Winner (2005)). As we can see in the next Section, this question, that is mainly empirical, is far from being solved.

All in all, the various theories on FDI set out a large number of potential FDI determinants. However, the FDI literature does not have a consensus established model (Blonigen (2005)). In light of this, the presence of model uncertainty can be considered a major econometric problem that should be solved prior to specifying a model and estimating meaningful long-run relationships.

3.3 Survey of the empirical literature

In this Section, we review the empirical literature to single out the different variables that have been included in the models of previous studies depending on the theoretical approach followed. This will help us to choose the prospective variables and bundle them in different groups. Given the absence of an established theoretical framework to model FDI patterns, the common practice in empirical studies has been to focus on a particular variable or a set of variables of interest to explain a certain theory. This approach leads to an omitted variable bias problem (Blonigen (2005)) and to use determinants that are not robust across alternative specifications and theories. According to the mainstream theoretical approaches described above, this section presents a brief empirical literature review and outlines the variables identified in the major FDI studies and their statistical relationships with FDI. Following Assunçao et al. (2011), we focus on quality of infrastructure in the OLI paradigm as the other determinants that could be included in this theory are the focus of other theoretical approaches (Institutional approach and New Trade Theory) developed afterwards.

3.3.1 The OLI paradigm

A good quality of infrastructure—often considered an indicator of agglomeration effects—is found to increase potential returns to investment and hence can attract FDI inflows. However, the results concerning their significance are not conclusive neither the choice of the variables used to proxy for the quality of infrastructure distinguishing between physical, financial, and technological infrastructure. Percentage of roads paved and railway networks are often used in the literature as proxies for physical infrastructure. Agiomirgianakis et al. (2003) included these variables in a longitudinal analysis for the first time and found no effect on roads and a positive effect for railway network. These variables can also signal the level of development and the population distribution of the host country. In principle, an extensive network promotes trade within the country and helps a foreign investor to gain access to separate or different markets at the lowest cost. Di Giovanni (2005) and Stein and Daude (2007) use telephone calls as a proxy and found that this variable has a significant and positive effect on FDI. Using the number of internet connections as a proxy, Botric and Skuflic (2006) concluded that the relationship between infrastructure and FDI is negative. This striking result may be explained because internet only became widespread in these countries after 2000. Some studies, such as Demekas et al. (2007), use an Infrastructure index as a proxy and found a positive impact on FDI.

3.3.2 Institutional approach

The quality of institutions is viewed as a crucial location advantage for MNEs in the recent literature. Good institutions are supposed to promote a healthy investment climate through protection of property rights, political stability and weaken corruption. However, estimating the magnitude of the effect of institutions on FDI is difficult because the quality of institutions is difficult to measure (Blonigen (2005)). Thus, empirical studies often provide mixed results. Some studies have used corruption as a measure of political risk. The early study of Wheeler and Mody (1992) used the first principal component of 13 risk factors as a proxy for good institutions, and find no evidence of a significant positive relationship between good quality of institutions and location of US foreign affiliates.[3] On the contrary, Wei (2000) uses bilateral data on FDI stocks to test the effect of corruption on FDI and finds support for a negative relationship between weak institutions and FDI. Specifically, Wei (2000) shows that a rise in the level of corruption in the host country reduces inward FDI. However, studies such as Egger and Winner (2005) and Adam and Filippaios (2007) found that corruption has a positive impact on FDI, whilst Harms and Ursprung (2002) and Busse and Hefeker (2007) confirm Wei (2000) results in that MNEs are rather attracted by countries in which civil and political freedom is respected. In line with these studies, Asiedu (2006) provides evidence that not only corruption but also political instability deter FDI. Financial institutions also seem to play a significant role on FDI. Di Giovanni (2005) found that the size of financial markets, as measured by the stock market capitalization to GDP ratio, has a strong positive association with domestic firms investing abroad. A number of studies focus on the different dimensions of governance that can affect the investment decisions of MNEs. Governance infrastructure is usually measured by the six governance indicators estimated by Kaufmann et al. (1999).[4] Empirical evidence confirms that the quality of institutions has positive effects on FDI. However, other studies show that the impact of institutions depends on the specific dimension considered. According to Daude and Stein (2007), Regulatory Quality and Government Effectiveness have a positive and significant impact on the volume of FDI while Voice and Accountability, Political Stability, Rule of Law and Control of Corruption do not. Globerman and Shapiro (2002) shows that good governance impacts positively both FDI inflows and outflows. In the same line, Bénassy-Quéré et al. (2007), relying on a gravity model for bilateral FDI stocks, shows that institutional distance tends to reduce bilateral FDI. More recently, Blonigen and Piger (2014), applying a Bayesian approach to variable selection, find no robust evidence on the role of institutions in bilateral FDI.

[3]Risk factors included indices of political stability, inequality, corruption, red tape, quality of the legal system, cultural compatibility, attitude toward foreign capital, and general expatriate comfort.

[4]The six governance indicators are: Voice and Accountability, Political Stability and Lack of Violence, Government Effectiveness, Regulatory Quality, Rule of Law and Control of Corruption. In all cases, larger values indicate better institutions.

3.3.3 New Trade Theory

The New Trade Theory has proposed a number of key FDI determinants for the different motivations of overseas investment. In terms of HFDI (market seeking), the most important factor is the size and growth of the host country (Horstmann and Markusen (1987), Brainard (1997)). Empirical studies show that market size, usually proxied by GDP or GDP per capita, have a significant positive influence on FDI (Eaton and Tamura (1994), Carr et al. (2001), Chakrabarti (2001), Bergstrand and Egger (2007), Head and Ries (2008)). Indeed, it is considered the most robust FDI determinant in empirical studies. Trade openness has been other of the factors the New Trade Theory has focused its attention on. Trade openness is identified as a driver of export-platform FDI.[5] Indeed, a range of studies show a significant positive influence of openness (measured mostly by the ratio of exports plus imports to GDP) on FDI (Agiomirgianakis et al. (2003), Talamo (2007), Demekas et al. (2007)). However, the relationship between FDI and trade depends on the type of investment. The general prediction of the economic models is the existence of complementarity between trade and vertical FDI, while substitutability should prevail between trade and horizontal FDI (Markusen and Maskus (2002)). However, these results can be somewhat at odds with those found by Franco (2013), in the sense that market seeking FDI can positively influence export intensity, while resources-seeking or export platform types of FDI are not relevant in enhancing export intensity. Internationalization and firm performance can be jointly rationalized within the "new trade theory" and the "off-shoring of international firms" literature signaling that both country-level and firm-level factors can play a determinant role in the internationalization strategies followed by firms. In a recent paper, Yang and Mallick (2014) uncover the impact of several macroeconomic and firm-specific factors explaining the variation in the learning effect across studies beyond the self-selection effect. Overall, they find that the returns from outward FDI tend to reduce the learning effect.

[5] See Ekholm et al. (2007) and Bergstrand and Egger (2007)

The implementation of RTAs has also been considered to have an impact on FDI, due to its effect through trade liberalization. However, the empirical results are inconclusive. Some studies such as Bergstrand and Egger (2007) found a negative relationship between trade agreements and FDI, whilst others (for example Berger et al. (2013)) found a positive one. According to VFDI (resource seeking), resource endowments are one of the most important location factors for overseas investments. The reason is that resource-seeking investors want to gain access to locally available natural resources (such as raw materials) or that are available at a lower cost (such as cheaper unskilled labor (Dunning (1977, 1979, 2000))). The ratio of population to land area is often used in the literature to capture agglomeration effects, as confirmed by Eaton and Tamura (1994), that obtained a positive impact on FDI. At the same time, there are mixed results on the impact of country skill abundance on FDI. Some empirical studies used average years of schooling as a proxy for human capital and obtained a positive relationship (see, for instance, Eaton and Tamura (1994) and Razin and Sadka (2007)). The seminal work of Carr et al. (2001) uses a measure of skilled labour to total employment and also finds a positive sign.[6] However, this proxy has been questioned by Blonigen et al. (2003) because it is affected by the differences in statistics on skill classifications across countries. They propose instead using as a proxy for skilled labour the absolute difference in average years of schooling between source and host country and get opposite results. Finally, concerning the availability of natural resources, oil, for example, has been identified as an important factor for the decision of MNEs to invest in a particular market (Harms and Ursprung (2002), Asiedu et al. (2015)).

[6]Specifically, their measure is the percentage of total employment that is employed in categories 0/1 (professional, technical, and kindred workers) and 2 (administrative workers).

3.3.4 Other

Additional variables, difficult to classify, have also been included in empirical studies for FDI determination. Some of these studies, related to the international finance literature, find that host country economic and financial stability are relevant for FDI. Inflation, exchange rates and volatility of exchange rates are often used as proxies for financial stability. Empirical studies like Blonigen (1997) predict that exchange rate depreciation increases FDI inflows into the host country. A negative relationship is predicted, among others, by Di Giovanni (2005) whereas Cavallari and D'Addona (2013) finds no significant effect. Wei (2000) and Razin and Sadka (2007) argue that corporate tax rates are an important determinant of FDI, and provide evidence of a negative effect, whereas for Wheeler and Mody (1992), for example, this variable is not significant.

3.4 Econometric methodology

3.4.1 Bayesian Variable Selection

In Sections 2 and 3 we have argued that there is a lack of consensus in both, the theoretical and the empirical FDI literature, regarding the empirical specification and the potential covariates of bilateral FDI determination. This model uncertainty problem, where the entertained models differ in which explanatory variables, from a given set, should be included in order to explain the response, is known in statistics as *variable selection* problem. In this paper, we apply a Bayesian Model Averaging (BMA) approach to deal with the variable selection problem in linear models.

More concisely, following Garcia-Donato and Forte (2018), we define the following econometric specification for each model M_γ :

$$M_\gamma : y = \alpha 1_n + X_\gamma \beta_\gamma + \varepsilon, \qquad \varepsilon \sim N_n(0, \sigma^2 I) \tag{3.1}$$

where y is the n dimensional vector of observations for the response variable (i.e. German outward FDI); X_γ is the $n \times p_\gamma$ matrix of potential FDI determinants in X; β_γ is the vector of linear regressors and finally, ε is a white error noise. Following the traditionally specific notation for variable selection we use a p dimensional binary vector $\gamma = (\gamma_1, \ldots, \gamma_p)$ to identify the models. In Eq.3.1, each competing model M_γ for $\gamma = 0, \ldots, 2^p - 1$ relates the response variable y to a subset of p covariates. Considering p possible potential FDI determinants grouped in the $n \times p$ matrix $X = (x_1, ..., x_p)$, the variable selection problem has a model space of 2^p competing models. The set of all competing models is called the model space and is denoted as M.

In order to compare alternative models, the Bayesian variable selection problem is based on the posterior probability that M_γ is the true model that generated the data. The posterior probability is formally defined by the Bayes theorem:

$$Pr(M_\gamma \mid \mathrm{y}) = \frac{m_\gamma(y)\, Pr(M_\gamma)}{\sum_\gamma m_\gamma(y) Pr(M_\gamma)}. \tag{3.2}$$

where $Pr(M_\gamma)$ is the researcher's prior probability that M_γ is the true model and m_γ is the integrated likelihood with respect to the prior π_γ:

$$m_\gamma(y) = \int f_\gamma(y \mid \beta_\gamma, \alpha, \sigma)\, \pi_\gamma(\beta_\gamma, \alpha, \sigma^2)\, d\beta_\gamma d\alpha\, d\sigma^2, \tag{3.3}$$

In Eq.3.3, π_γ is the prior distribution for the model-specific parameters of M_γ and the most controversial element in the BMA analysis.

The posterior probability in Eq.3.2 is the end product of the Bayesian approach and shows how the model probability is distributed across different specifications of the FDI drivers.

In order to summarize the information contained in the posterior distribution, we will make use of the inclusion probabilities for each competing variable.

$$p(x_\gamma \mid y) = \sum_{\{M_l:\, x_\gamma \in M_l\}} P(M_l \mid y), \;\; \gamma = 1, 2, \ldots p \tag{3.4}$$

The inclusion probabilities should be interpreted as evidence (in a probabilistic scale) that x_γ explains the response variable. These probabilities are useful summaries of the posterior distribution and have interesting theoretical properties as shown in Barbieri and Berger (2004). Nevertheless, inclusion probabilities do not provide any information regarding the magnitude and sign of the coefficient β_γ. Such type of information can be obtained by averaging over all entertained models using the posterior probabilities in Eq.3.2 as weights. In the next section we report the model averaged estimated coefficients (namely, the posterior mean) alongside the inclusion probabilities. However, model averaged estimated coefficients should be interpreted with caution since, as highlighted by Garcia-Donato and Forte (2018), the simulations obtained from the model average estimation can be highly multimodal and thus providing default summaries of it (such as the mean or standard deviation) might be misleading. Thereby, as argued in Camarero et al. (2015), we will focus primarily on the inclusion of the probability for each variable from a large group of potential explanatory variables as the outcome of this methodology.

3.4.2 The Prior distribution

From a Bayesian perspective, the analysis of the posterior distribution provides the answer of the study under consideration. To obtain the posterior distribution in Eq.3.2 we need to determine the prior distributions for the parameters within each model π_γ and the prior distribution over the model space $Pr(M_\gamma)$. The choice of the prior will affect the posterior that we get. The subjectivity of this choice when $m_\gamma(y)$ can not be derived analytically given the large number of entertained models 2^p poses a significant challenge. In light of this, substantial research has been conducted on this topic. In this study, we will adopt the Robust prior for the regression parameters following the methodology proposed by Bayarri et al. (2012). The Robust prior for M_γ can be specified hierarchically as

$$\pi_\gamma^R(\alpha, \beta_\gamma, \sigma) = \sigma^{-1} N_{p_\gamma}(\beta_\gamma | 0, g\Sigma_\gamma), \tag{3.5}$$

where $\Sigma_\gamma = \sigma^2 (V_\gamma^\top V_\gamma)^{-1}$, with

$$V_\gamma = (I_n - X_0(X_0^\top X_0)^{-1} X_0^\top) X_\gamma, \tag{3.6}$$

and

$$g \sim p_\gamma^R(g) = \frac{1}{2}\sqrt{\frac{1+n}{p_\gamma + p_0}}(g+1)^{-3/2}, \quad g > \frac{1+n}{p_\gamma + p_0} - 1. \tag{3.7}$$

The theoretical properties of this prior (see Bayarri et al. (2012) for the details) as well as the computational advantage that arises from the fact that it provides marginal densities in an analytic way (i.e., integral in Eq. 3.3 can be solved algebraically), make it an attractive choice for this kind of analyses. Concerning the prior distribution over the model space $Pr(M_\gamma)$, we adopt the ScottBerger prior. Multiplicity issues are a concern in variable selection problems and particularly, spurious evidence is high for models with moderate to large p. Thus, we control for multiplicity with the prior probabilities $Pr(M_\gamma)$ as proposed by Scott and Berger (2006):

$$Pr(H_\gamma) = \left((p+1)\begin{pmatrix} p \\ p_\gamma \end{pmatrix}\right)^{-1} \tag{3.8}$$

The ScottBerger prior assigns the same probability to models of the same dimension (the dimension of M_γ *is* $p_\gamma + p_0$) and it must be inversely proportional to the number of models of that dimension.

The described variable selection approach is implemented in R using the package BayesVarSel described in Garcia-Donato and Forte (2018) and that we apply to the case of German FDI in the next section. In particular, we use the function GibbsBvs to obtain approximations to the posterior inclusion probability of covariates based on the methodology by García-Donato and Martínez-Beneito (2013).

GibbsBvs approximates computation of summaries of the posterior distribution using a Gibbs sampling algorithm to explore the model space and frequency of "visits" to construct the estimates. The Gibbs sampling scheme is a simple (yet very efficient) algorithm proposed by George and McCulloch (1997) and later studied in García-Donato and Martínez-Beneito (2013) in the context of large model spaces (with p > 25).

3.5 Data and results

3.5.1 Data

In this Section we apply BMA techniques to select the determinants of German outward FDI.[7] For this purpose, we have assembled a large panel dataset covering the period 1996-2012.

[7]Our study ties in with the literature on partial equilibrium analysis of FDI determinants. Restricting ourselves to outward FDI data from a common parent country, Germany, allows us to focus on the investors' motives (i.e. the supply side) in a particular location. However, this approach is not free from limitations: by leaving third country characteristics outside the analysis, our study neglects the interdependence of FDI decisions, what may cause some bias in the results obtained. Following the developments in spatial econometric techniques, general equilibrium theories of FDI have recognized the role of the potential interdependence in FDI across locations, suggesting that FDI should be modelled in a multilateral context, that is, considering home, host, and third country characteristics (see, for instance (Baltagi et al. (2007) and Blonigen et al. (2007))). Bearing these limitations in mind, we consider our analyses informative for a given country's policy options to attract investment.

The data on FDI stocks were obtained from the UNCTAD's Bilateral FDI Statistics. Our full sample has 1105 total observations, of which 0.90% are zero. Given that we consider a log-linear regression for the BMA analysis, these zero values create a selection issue as the logarithm of zero is undefined. We have dropped these observations from the sample. This solution might provide misleading coefficient estimates as highlighted by Silva and Tenreyro (2006) who recommend the use of non-linear estimators. More recently, Eicher et al. (2012) provide a methodology to address selection bias and model uncertainty simultaneously. However, despite the potential limitation of our approach, it facilitates the comparison of our results with the majority of previous studies, that usually use a logarithmic transformation to address skewness in the FDI variable.[8]

The FDI dataset was then augmented with 61 additional explanatory variables, although those included in the different country-group analysis may vary depending on the specific countries considered as well as for statistical reasons (perfect collinearity, for example). The dataset includes first of all, standard gravity variables, and then a series of complementary variables that we have assembled taking into account the survey made in Sections 2 and 3: other related GDP and population measures, factor endowments and productivity, cultural/historical factors, economic risk and exchange rate variables, trade openness measures, infrastructure and political environment and institutions. Table3.A.1 in the Appendix provides a full list of the variables included, their definition and sources. The dataset contains information about 59 destination countries of FDI stock - 38 developed and 21 developing. See Table 3.A.2 for a detailed presentation of the destination countries included in our sample.

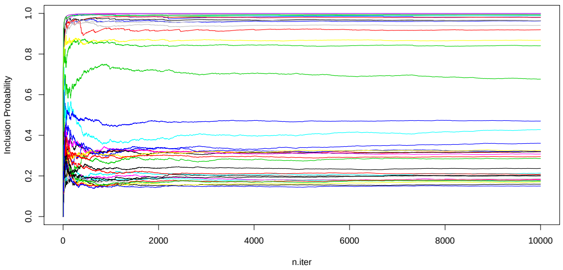
Due to the heterogeneity of the destinations of German FDI, we have considered several country-groups. We distinguish between developed and developing countries to get a first insight of the main motives of German FDI. Yet, we also distinguish between Latin American

[8]A Shapiro and Wilk (1965) test clearly rejects the hypothesis that FDI is normally distributed.

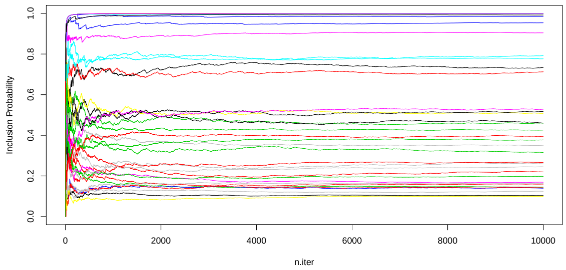
and Asian countries within developing countries, as well as “core” and “periphery” countries within Europe. We have decided to focus on the groups of countries because, even if there are not statistical limitations to apply this method to the whole dataset, the average effects obtained from this aggregation would omit information concerning the diversity of reasons to invest in different country-groups (see, Mitchell et al. (2011) for the so-called aggregation bias).

We run the BMA analyses with 100.000 iterations. In order to informally asses the convergence of the posterior inclusion probabilities, Figure 3.3 provides trace plots for each of the corresponding subsamples and periods considered in our analyses. The figure shows that the posterior inclusion probabilities in all subsamples stabilize after 8.000 iterations.

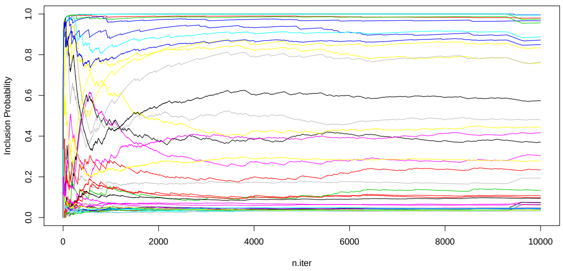
Fig. 3.3 Evolution of the inclusion probabilities with the iterations in Gibbs sampling, 1996-2012.



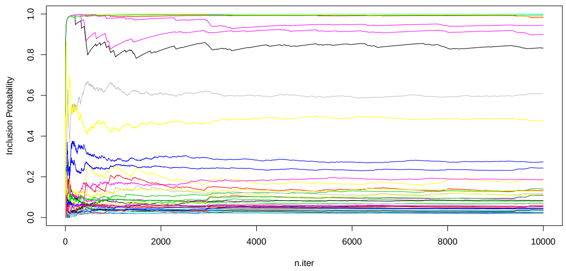
(a) Developed



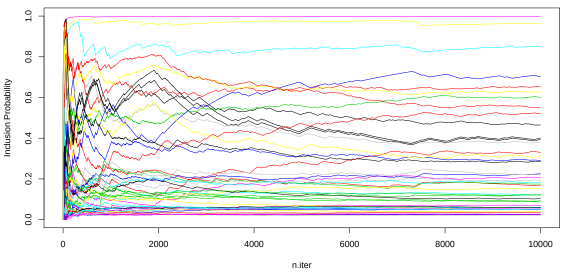
(b) Developing



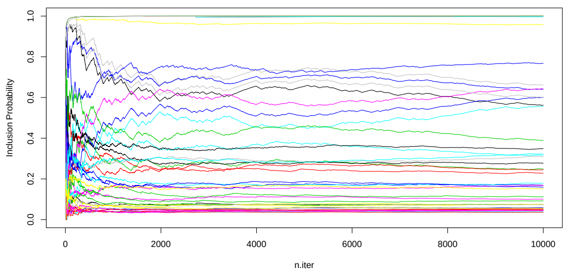
(c) Latin America



(d) Asia



(e) EU Core



(f) EU Periphery

3.5.2 Results

In what follows, we present the results for the BMA analysis on the determinants of German outward FDI across regions.[9] We report the inclusion probabilities for each competing variable as a summary of the posterior distribution. These probabilities should be interpreted as the evidence shown by the data that a variable explains German outward FDI once the potential control variables have been taken into account. They provide a summary measure of the marginal importance of different drivers of FDI. We consider robust FDI determinants those variables having a posterior inclusion probability (PIP) above the recommended threshold of 0.50. Following the criterion posit by Raftery (1995), evidence for a regressor with a posterior inclusion probability from 50 to 75% is called weak, from 75 to 95 % positive, from 95 to 99 % strong, and >99 % very strong. Furthermore, the model defined by those variables with an inclusion probability greater than 0.5 is called in the literature a Median Probability Model (MPM) and the theory in Barbieri and Berger (2004) shows that, under general conditions, the MPM is optimal for prediction purposes.

The inclusion probabilities of each sample considered in this paper are presented in Figures 3.4, 3.5 and 3.6. Given the high number of variables considered we have opted to present the results on a visual basis and we refer the interested reader to Tables 3.B.2, 3.B.3 and 3.B.4 in the Appendix for a detailed report of the estimated probabilities and posterior means. With a few exceptions, the magnitude and sign of the coefficients are in line with the theoretical predictions and previous empirical studies. Nevertheless, one should bear in mind that we consider a log-linear regression for the BMA analysis and thus our results may suffer

[9] We have also conducted the BMA analysis for all host countries, both developed and developing or emerging. Results are reported in Table 3.B.1 in Appendix B. We have found a large number of variables with large inclusion probabilities mainly due to the heterogeneity of the destination countries. The implicit assumption is that a homogeneous set of FDI determinants governs FDI across all countries. There exists, however, ample evidence that subsamples of countries follow distinctly different FDI patterns (Eicher et al. (2012)).

from selection bias as highlighted by Eicher et al. (2012). Thereby model average estimated coefficients need to be interpreted with caution.[10]

FDI determinants in developed and developing countries

Taking into account the world patterns of German FDI, a first approach to identify investment strategies consists of dividing our sample into developed and developing countries. The results for the BMA analysis are presented in Figure 3.4. We find that 21 variables out of 44 and 22 out of 43 have an inclusion probability over 0.5 for developed and developing countries, respectively. These findings point to a specification more parsimonious (as half of the covariates are dropped out) than previously suggested by the literature, in line with the seminal paper of Blonigen and Piger (2014). Overall, although the determinants identified highlight that the decision to invest abroad is based on a mixture of FDI theories (as argued by Faeth (2009) in her survey), we can infer from the results that the motivations for German FDI in developed and developing countries are different.

Concerning the developed countries (left-hand side graph in Figure 3.4), the inclusion probabilities of GDP measures point towards the relevance of market size or market potential, suggesting thereby horizontal FDI (HFDI) motivations in developed countries. GDP per capita differences, urban population and population have high inclusion probabilities, that is consistent with the idea that German MNEs invest largely in developed countries, such as the US, due to its market size. Figure 3.1 reports the importance of the US as destination of German FDI once the “core” and peripheral EU countries are considered separately. In addition, traditional variables for the gravity model are found to be robust determinants of German FDI in developed countries. Shared Regional Trade Agreements (RTAs) and established trade relationships present also large statistical support. Another important reason

[10] A more accurate estimation of the coefficients of the robust determinants obtained from the BMA analysis it is beyond the scope of this paper. We leave this estimation to future research.

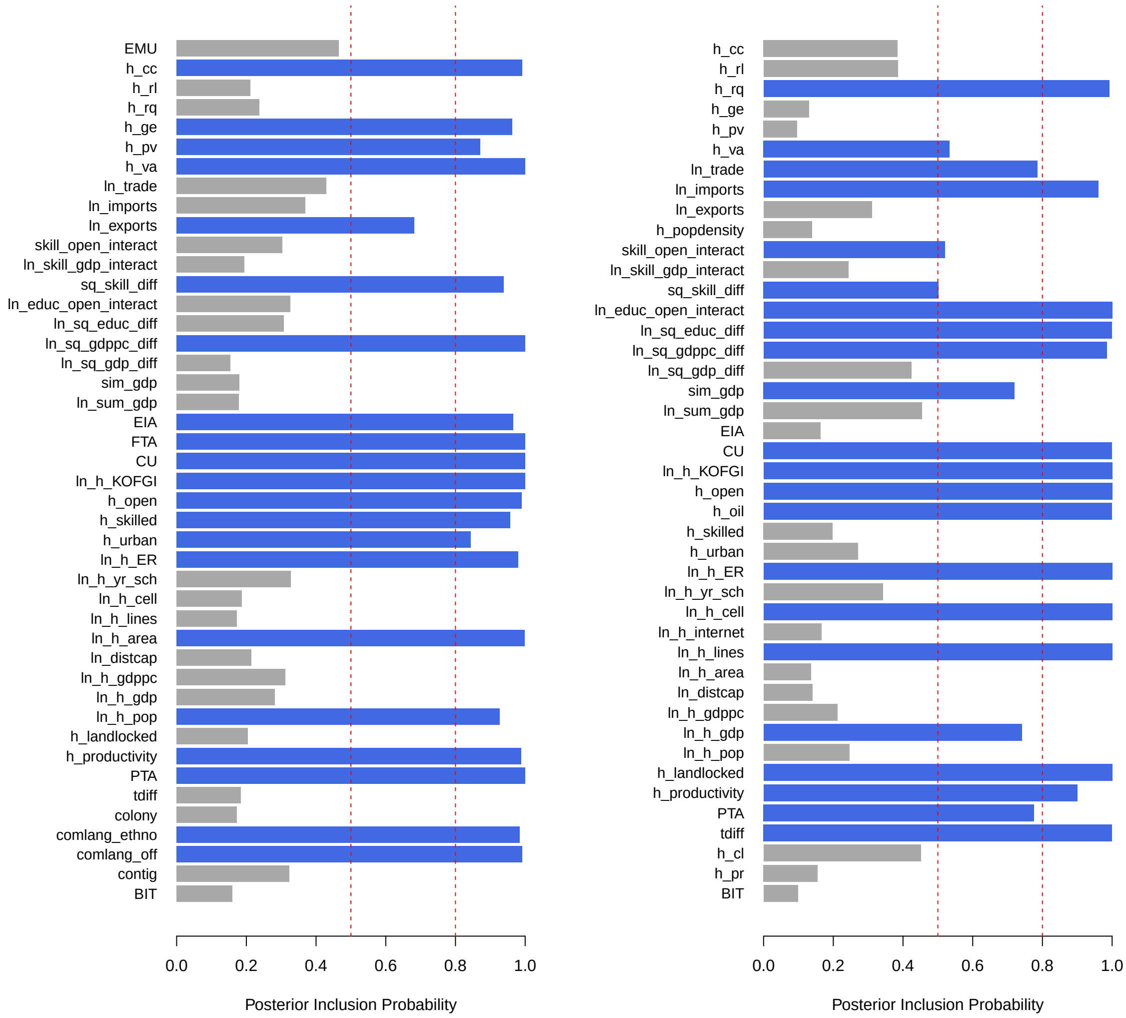
for German overseas investment is to serve as export-platform; MNEs would invest in a host country in order to serve a third one with exports of final goods as argued by Ekholm et al. (2007)). The high PIP of the host country's degree of openness to international trade, exports and the KOF Globalisation index, also support this argument. These findings are also in line with the idea of vertical specialization in the current process of internationalization of production, where integration strategies into GVC imply intra-firm trade and, thus, MNEs are attracted to relatively open economies. Other vertical motives for FDI are also statistically relevant: the acquisition of advanced technologies and skilled labour (skilled labour, skill differences and productivity) are found to be robust determinants. Covariates related to cultural and historical factors are also selected. The ability to communicate seems to play a major role for overseas investment, as speaking the same language (or a significant proportion of the population speaking it) have the highest inclusion probabilities. Competitiveness (the real exchange rate), as well as the quality of institutions (including voice and accountability, political stability, government effectiveness and control of corruption) have high inclusion probabilities, as in Antonakakis and Tondl (2015), that also found that institutional quality is a German FDI determinant. Finally, we have found no support for geography measures and infrastructure variables.

Regarding German investment in developing countries (see the right-hand side of Figure 3.4), we find strong statistical support for trade openness and treaties variables, suggesting an export-platform FDI strategy. RTAs, established trade relationships, the host country's openness to international trade and the KOF Globalisation index present high PIP. This supports a complementary relationship between trade and FDI. Other vertically-oriented determinants, such as education and skill differences as well as productivity have high inclusion probabilities. At the same time, the interactions among key variables (trade openness and skilled labour endowments and trade openness and education differences) are also found to be robust determinants. Horizontal FDI indicators, such as similarity index, GDP per capita difference as well as the geography measures have high inclusion probabilities. Thus, these findings are consistent with the idea that traditional gravity variables play a key role in overseas investment. Furthermore, access to natural resources is found to be important for German investment in these countries, as the oil rents indicator has high inclusion probability. Other covariates, such as competitiveness (measured as the bilateral real exchange rate) and telecommunications infrastructure (fixed telephone and mobile cellular subscriptions) present an inclusion probability of 1. Institutional covariates (voice and accountability and regulatory quality) are also selected. Finally, there is no statistical support for cultural/historical factors. All in all, our results provide evidence in favor of both market-related (HFDI) as well as vertical motives for FDI (VFDI), although the latter are stronger.

In summary, we can conclude from the BMA analysis that MNEs engage in complex integration strategies which involve a mixture of FDI theories, including horizontal and vertical motives. Indeed, our results for both country groups support the knowledge-capital model, where vertical and horizontal motives may play an important role alike. However, some FDI strategies receive a stronger statistical support than others.

However, the analysis of these two large groups of countries is not completely satisfying, as the mixed strategy that prevails in both developed and emerging countries may be the result of aggregation, masking different strategies for different geographical areas or income groups. For this reason, taking into account the distribution of German FDI, we divide the developing countries into two large areas: Latin America and Asia. For developed countries, we focus on the European continent and separate "core" from "peripheral" European countries. We devote the next two subsections to this disaggregated analysis.

Fig. 3.4 Inclusion probabilities for each of the potential covariates considered for developed (left) and developing countries (right), 1996-2012.



FDI determinants in Latin American and Asian countries

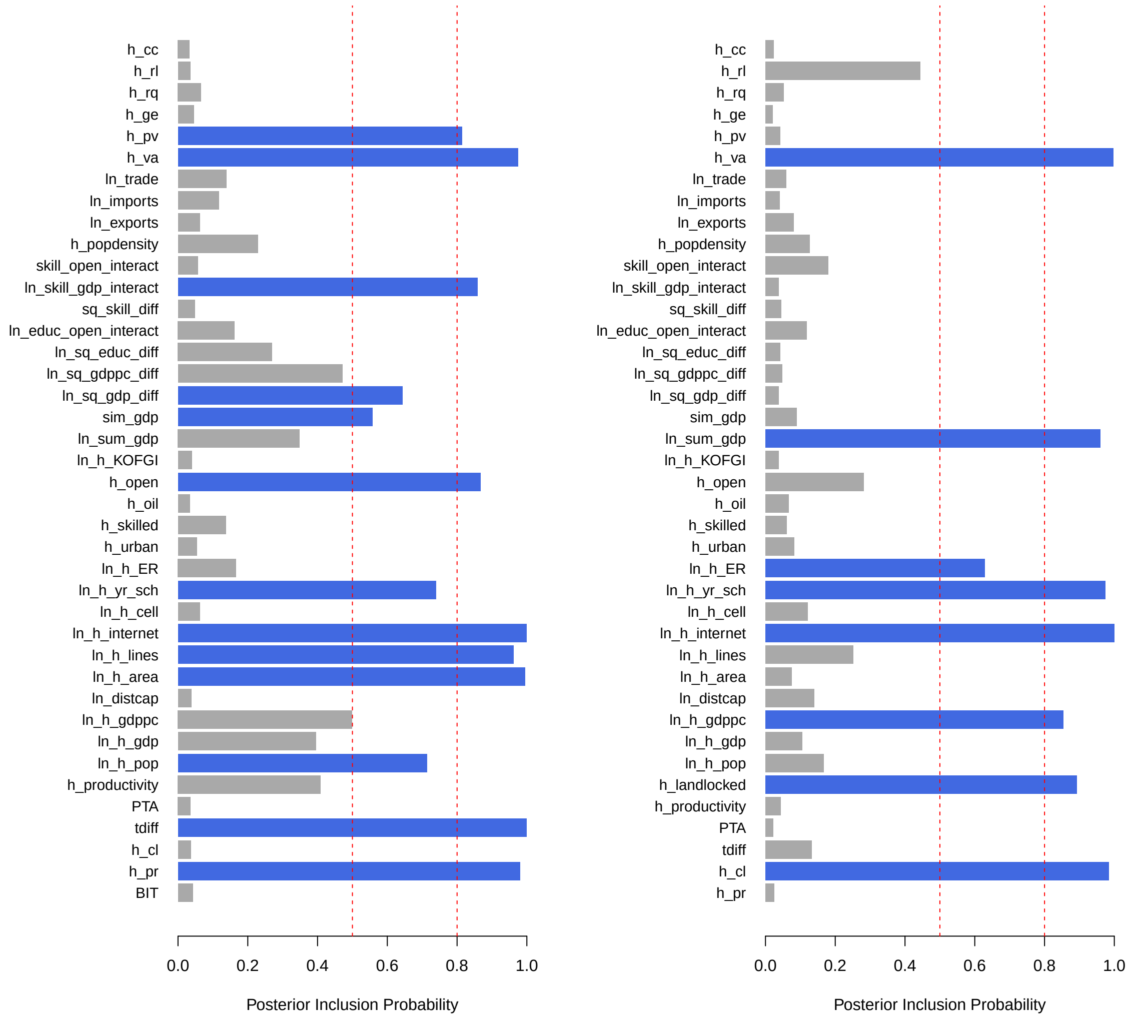
Within developing countries, Asian countries hold the 41.92% of German OFDI along the period 1996-2012, while Latin American countries represent the 35.15%. We hypothesize that German FDI strategies towards these regions might be different and, thereby we split our sample of developing countries into Latin American and Asian countries. The BMA results are presented in Figure 3.5. In this case, 13 variables out of 40 and 8 out of 40 have an inclusion probability over 0.5 for Latin American and Asian countries, respectively.

Concerning the Latin American countries (see the left-hand side panel of Figure 3.5), there is some evidence of HFDI indicators. Population, the similarity index, GDP per capita difference as well as the geography measures (such as time zone difference) have an inclusion probability above 0.5. Stronger is the support for vertical FDI motives, as factor endowment variables and other vertical variables, such as the level of education and trade openness, as well as the interaction of GDP differences with skill differences, present high inclusion probabilities. Moreover, telecommunications infrastructure, as well as institutions (in this case, the political rights, Voice and Accountability and political stability indices with a PIP above 0.8) are statistically relevant. Finally, there is no support for cultural and historical factors, exchange rate/monetary policy and trade and investment treaties.

Similar to the Latin American case, the market-driven (HFDI) motive seems to be relevant for German investment in Asian countries (right-hand side of Figure 3.5). GDP covariates receive more statistical support: the sum of host and parent GDP and GDP per capita have high inclusion probabilities. Access to the sea is also a robust FDI determinant. Taken together, the gravity model receives support in Asian countries. The evidence is consistent with the idea that German MNEs seek to access large-sized markets in order to expand their production. This might be the case for German investments in China whose high GDP growth and enormous population has boosted its attractiveness as an investment location. In addition, factor endowment covariates, such as the education level, have high PIP, suggesting also vertical FDI motivations. Competitiveness, telecommunications infrastructure (internet users) and institutions (the civil liberties and voice and accountability indices) have all very large inclusion probabilities. Finally, we found no support for other broad categories of variables, specifically those related to cultural/historical factors.

In summary, both vertical and horizontal FDI determinants are present in the two regions, together with a strong importance of the quality of institutions indices to explain FDI patterns in the two country-groups. However, vertical FDI motives are more relevant for Latin America, while horizontal FDI seems to prevail in Asian countries. Finally, as expected in the case of Germany, we find no support for cultural nor historical factors.

Fig. 3.5 Inclusion probabilities for each of the potential covariates considered for Latin American (left) and Asian countries (right), 1996-2012.



FDI determinants in EU regions

As previously shown in Figure 3.1, the European Union is the single largest regional location for German investments. In order to examine the internationalization strategies of German MNEs within Europe we divide our sample into Core and Peripheral countries following the pattern of German investment shown in Figure 3.2. The "core" region includes the founding countries plus the first enlargement in 1973 and the Northern countries, excepting Italy, Ireland and Finland. In parallel, the "periphery" comprises the enlargement to the south and east plus Italy, Ireland and Finland. Figure 3.6 provides the results of the BMA analysis, whereas the specific inclusion probabilities and the posterior means are reported in the Appendix (Table 3.B.4). For the core countries, 9 variables out of 47 are found to be robust determinants, while for peripheral countries, 16 variables out of 48 have an inclusion probability higher than 0.5.

In the case of the core countries, the GDP measures receive some statistical support: similarity index and GDP per capita have inclusion probabilities between 0.5 and 0.6. Distance and other geography measures, namely landlocked host country, are robust determinants for investment in core countries as well, in line with the gravity model. However, there is also a strong statistical support for factor endowment variables, suggesting thus vertical FDI motivations (land area and wages have high inclusion probabilities). This latter finding is in line with those provided by Antonakakis and Tondl (2015) who state that German MNEs are particularly attracted by advanced markets with relatively high wages and productivity. Probably, the reason for the relevance of wages has more to do in this case with it being a proxy for income than with the cost of production. Furthermore, trade openness variables are found to be robust determinants: exports have a high inclusion probability. Taking into account the positive sign of the posterior mean, FDI and exports would be complementary rather than substitutes. In fact, after all the years of deep economic integration in Europe, German's FDI is concentrated in the core countries, as it represents around 40% of its total stock. Finally, fixed telephone subscriptions and internet users (that represent telecommunications infrastructure) have a high inclusion probability, while there is no support for cultural and historical factors, exchange rate/monetary policy, trade and investment treaties and institutions. This is an expected outcome, as these countries are EU members and most of them are in the euro area.

On the other hand, in the case of German investment in peripheral countries, our results point to market-seeking FDI motivations. Similarity index and urban population have high inclusion probabilities. Distance and other geography measures, namely contiguity, receive also statistical support, in line with the gravity model. Productivity has also a high PIP, suggesting VFDI motives, and competitiveness of host country is a robust determinant in this period too. Regional Trade Agreements (RTAs) and all the covariates that capture the integration process, such as CU, PTA and BIT membership, have high inclusion probabilities, as some of the countries in the sample joined the EU during the sample period.[11] We also find statistical support for trade openness variables: exports, imports and trade present high inclusion probabilities, and provide evidence of a complementarity relationship between FDI and trade (as the associated posterior mean of the trade variable is positive). Furthermore, these findings are consistent with vertical FDI and complex integration strategies into GVCs. Indeed, Germany was the most important of the three interconnected production hubs in the global trading system in 2000, although since then China has been increasing its weight in the international production networks (Shin (2019)).[12] In addition, there appears to be more statistical support for transport and telecommunications infrastructure for these countries: kilometres of rail lines and fixed and mobile telephone subscriptions have an inclusion probability of 1. Furthermore, the quality of institutions, political stability and absence of

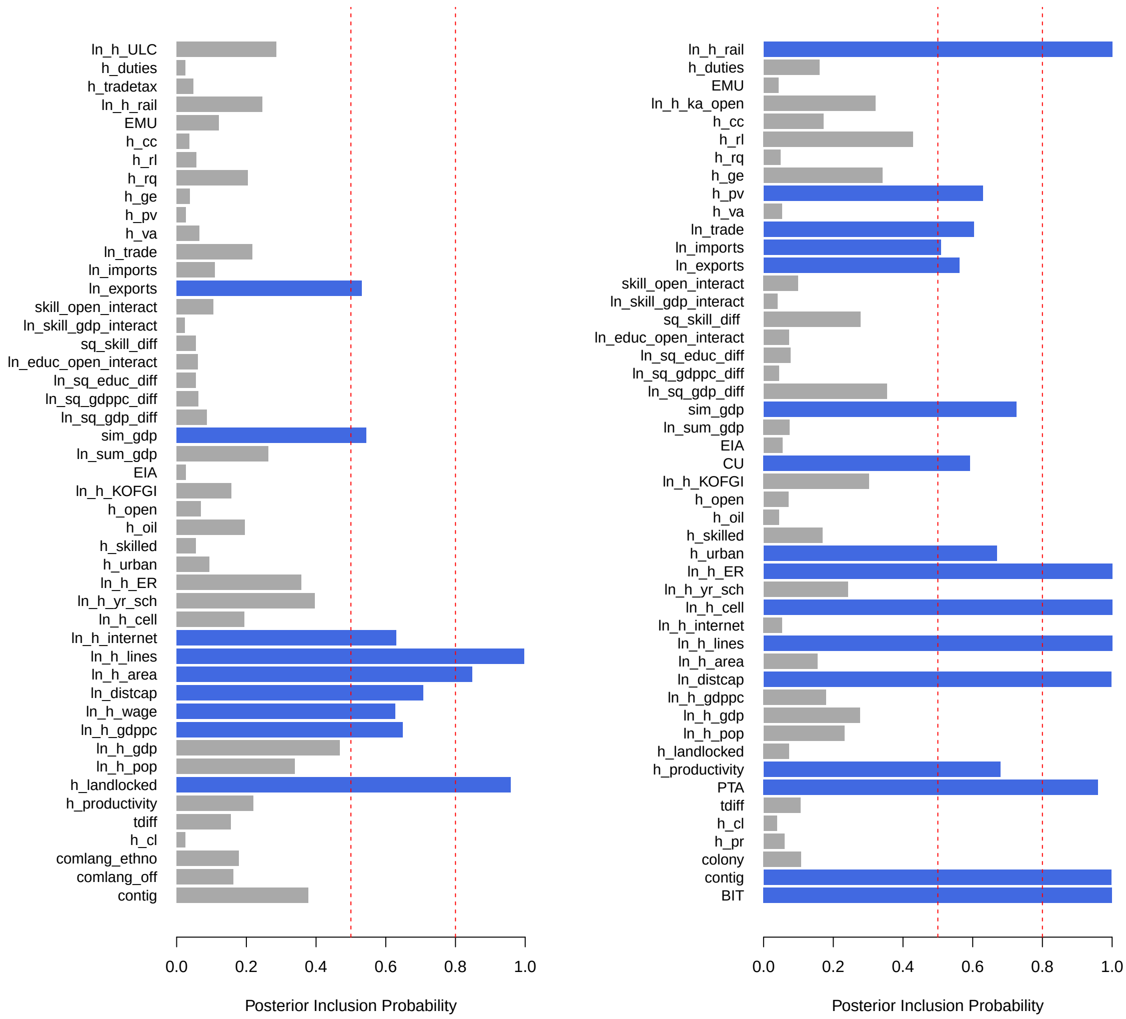
[11]As already stated, for this same reason, that they were already members, these variables were not relevant in the group of core countries above.

[12]There appear to be three interconnected production hubs of GVCs in the world: North America (with its epicenter in the United States), East Asia (i.e. China, Japan and the Republic of Korea) and Europe (being Germany the core country).

violence have all high inclusion probability. Finally, there is also no statistical support for cultural and historical factors.

Overall, the results show that drivers for horizontal FDI to large euro area markets are relatively more important for German investment in "core" EU countries than is the case of the peripheral ones, where vertical motives appear to play a major role.

Fig. 3.6 Inclusion probabilities for each of the potential covariates considered for core (left) and peripheral EU countries (right), 1996-2012.



3.6 Concluding remarks

Despite of the importance of German Outward FDI, the studies analyzing its determinants are relatively scarce. Moreover, from a theoretical point of view, there is not a unified model to explain FDI determinants, but several theories that postulate different hypotheses. This lack of consensus is extended to the empirical literature, as not only the baseline econometric specification, but also the proxy variables selected to account for the alternative theoretical channels, are heterogeneous. We claim, that in some areas, where there is uncertainty about the true or most adequate theoretical model, model selection should be implemented prior to any causality or regression analysis.

Our contributions to this literature about model uncertainty and variable selection in FDI models are twofold. First, from a methodological point of view, we adopt a Bayesian Model Averaging approach for the selection of the best model, providing an "objective" statistical procedure to model selection in contrast to the subjective-driven approach commonly used in the literature. Accordingly, we do not condition on a single model and instead attach probabilities to different models obtaining a robust identification of the FDI determinants. Using this method, we can select those variables with the highest explanatory power from a large group of 61 potential candidates. Second, we take into account the heterogeneity of the destinations of German FDI and consider different recipient-country groups in the analysis.

The results obtained for the country-groups show that only a small proportion of the variables proposed in the literature have high posterior inclusion probabilities. Consequently, we find that a robust FDI specification is rather parsimonious and many variables included in other studies seem to have a relatively low explanatory power. Overall, our results confirm that the decision to invest abroad involves a mixture of FDI theories and that the main internationalization strategies are different across regions. In summary, we find that determinants that are associated with market-seeking or HFDI motives are relatively more important for developed countries. In the case of developing countries, due to the increasing fragmentation of the production process, the factor endowments and thus, cost savings, play a greater role. The disaggregation of our sample of developing countries into Latin American and Asian countries highlights two features. First, HFDI and VFDI motives coexist for German MNEs investment in both, Latin American and Asian countries, together with a strong relevance of the quality of institutions. Second, HFDI seems to play a major role in Asian countries, while German multinational firms access Latin American markets appear looking for lower production costs, in line with VFDI. Concerning German FDI in its European neighbours, a further disaggregation into “core” and “peripheral" countries reveals that market access is the main motivation. However, the growing importance of GVC, where Germany plays a key regional role in Europe, is in line with vertically integrated multinational firms that are more relevant in developing countries as well as in peripheral European countries. Finally, our conclusions concerning the specific variables that are relevant in the vertical or horizontal strategies can provide some hints for policymakers to develop programs oriented to attract German investment.

Appendix A

Table 3.A.1 Data description and source

Variable name	Description	Source
ln_ofdistock	Log of bilateral outward FDI stock in millions (constant 2010 US$)	UNCTAD Bilateral FDI database
	GDP and Population Measures	
ln_sum_gdp	Log of sum of HOST and PARENT real GDP	World Development Indicators, World Bank
sim_gdp	Log of share of HOST real GDP in the sum of HOST and PARENT GDP ∗ Share of PARENT real GDP in the sum of HOST and PARENT GDP	World Development Indicators, World Bank
ln_sq_gdp_diff	Log of squared real GDP difference between HOST and PARENT country	World Development Indicators, World Bank
ln_sq_gdppc_diff	Log of squared real GDP per capita difference between HOST and PARENT country	World Development Indicators, World Bank
h_urban	Urban population (% of total) in HOST country	World Development Indicators, World Bank
ln_h_pop	Log of HOST population, total in mn	Gravity database from CEPII
ln_h_gdp	Log of HOST GDP in trillions (constant 2010 US$)	World Development Indicators, World Bank
ln_h_gdppc	Log of HOST GDP per capita in trillions (constant 2010 US$)	World Development Indicators, World Bank
	Distance and other geography measures	
contig	Dummy variable indicating PARENT and HOST countries are geographically contiguous	Gravity database from CEPII
ln_distcap	Log of distance between the capitol cities in the PARENT and HOST country	GeoDist database from CEPII
tdiff	nb of hours difference between PARENT and HOST	Gravity database from CEPII
h_landlocked	1 if HOST is landlocked	GeoDist database from CEPII
	Factor endowments/productivity	
h_skilled	Percent of employment by skilled labor in HOST country	ILOSTAT
ln_h_area	Log of land area (sq. km) in HOST country	Gravity database from CEPII
ln_sq_educ_diff	Log of squared difference in average education years between PARENT and HOST country	PWT 9.0
sq_skill_diff	Squared difference in percent of employment by skilled labor between PARENT and HOST country	ILOSTAT
ln_educ_gdp_interact	Log(sq_gdp_diff ∗ sq_educ_diff)	PWT 9.0, World Development Indicators
ln_skill_gdp_interact	Log(sq_gdp_diff ∗ sq_skill_diff)	ILOSTAT, World Development Indicators
ln_h_popdensity	Log of population divided by land area in HOST country	Gravity database from CEPII
ln_h_yr_sch	Log of average years of schooling in the population aged 25 years and older, HOST country	PWT 9.0
h_oil	Oil rents (% of GDP)	World Development Indicators, World Bank
h_productivity	Share of labour compensation in GDP at current national prices in host country	PWT 9.0
ULC	Log of the average cost of labour per unit of output, index(2010=100)	OECD
ln_h_wage	Log of the average annual wages (In 2016 constant prices at 2016 USD PPPs)	OECD

(Continued)

Table 3.A.1 Data description and source *(Continued)*

Variable name	Description	Source
	Cultural/Historical factors	
comlang_off	Dummy variable indicating PARENT and HOST countries share a common official language	Gravity database from CEPII
comlang_ethno	Dummy variable indicating PARENT and HOST countries share a language which at least 9% speak in each country	Gravity database from CEPII
colony	Dummy variable indicating PARENT and HOST countries have had (or do have) a colonial link	Gravity database from CEPII
comcol	Dummy variable indicating PARENT and HOST countries have had a common colonizer since 1945	Gravity database from CEPII
curcol	Dummy variable indicating PARENT and HOST countries are currently in a colonial relationship	Gravity database from CEPII
col45	Dummy variable indicating PARENT and HOST countries have had a colonial link since 1945	Gravity database from CEPII
smctry	Dummy variable indicating PARENT and HOST countries were (or are) the same country	Gravity database from CEPII
	Trade and investment treaties	
PTA	1 if preferential trade agreement signature	Hofmann et al. (2017)
CU	1 if CU or CU & EIA	Hofmann et al. (2017)
FTA	1 if FTA or FTA & EIA	Hofmann et al. (2017)
EIA	1 if EIA or FTA & EIA or CU & EIA	Hofmann et al. (2017)
BIT	1 if BIT was signed in this year or earlier	Neumayer(2017); IIA Unctad
	Exchange Rate/Monetary policy	
ln_h_ER	Log of real exchange rate in host country, national currency/USD	PWT 9.0
EMU	1 if both belong to the euro area in year t	Own elaboration
	Trade openness	
h_tradetax	Taxes on international trade (% of revenue) in HOST country	World Development Indicators, World Bank
h_duties	Customs and other import duties (% of tax revenue) in HOST country	World Development Indicators, World Bank
ln_educ_open_interact	Log(sq_educ_diff ∗ h_open)	PWT 9.0, World Development Indicators
skill_open_interact	sq_skill_diff ∗ h_open	ILOSTAT, World Development Indicators
ln_exports	Log of 5-years lag of bilateral exports (of i to j) in mn (constant 2010 US$)	CHELEM database from CEPII
ln_imports	Log of 5-years lag of bilateral imports (of i from j) in mn (constant 2010 US$)	CHELEM database from CEPII
ln_trade	Log of 5-years lag of bilateral exports plus imports in mn(constant 2010 US$)	CHELEM database from CEPII
h_open	Trade (% of GDP)	World Development Indicators, World Bank
ln_h_KOFGI	Log of HOST KOF Globalization Index	Gygli et al.(2018)
ln_h_ka_open	Log of HOST Chinn-Ito index, normalized (0-1)	Chinn et al. (2006)

(Continued)

Table 3.A.1 Data description and source *(Continued)*

Variable name	Description	Source
	Infrastructure	
ln_h_rail	Log of kilometers of rail lines per sq km of land area in HOST country	World Development Indicators, Gravity database from CEPII
ln_h_phone	Log of telephone lines (per 100 people) in HOST country	World Development Indicators, World Bank
ln_h_internet	Log of internet users (per 100 people) in HOST country	World Development Indicators, World Bank
ln_h_lines	Log of fixed telephone subscriptions (per 100 people) in HOST country	World Development Indicators, World Bank
ln_h_cell	Log of mobile cellular subscriptions (per 100 people) in HOST country	World Development Indicators, World Bank
	Institutions	
h_pr	Political rights index for HOST country (Ranges from 1 to 7 with highest score indicating the lowest level of freedom)	Freedom House
h_cl	Civil liberties index for HOST country (Ranges from 1 to 7 with highest score indicating the lowest level of freedom)	Freedom House
h_va	Voice and Accountability, in percentile rank (Ranges from 0 (lowest) to 100 (highest))	World Governance Indicators (WGI), World Bank
h_pv	Political Stability and Absence of Violence/Terrorism, in percentile rank (Ranges from 0 (lowest) to 100 (highest))	World Governance Indicators (WGI), World Bank
h_ge	Government Effectiveness, in percentile rank (Ranges from 0 (lowest) to 100 (highest))	World Governance Indicators (WGI), World Bank
h_rq	Regulatory Quality, in percentile rank (Ranges from 0 (lowest) to 100 (highest))	World Governance Indicators (WGI), World Bank
h_rl	Rule of Law, in percentile rank (Ranges from 0 (lowest) to 100 (highest))	World Governance Indicators (WGI), World Bank
h_cc	Control of Corruption, in percentile rank (Ranges from 0 (lowest) to 100 (highest))	World Governance Indicators (WGI), World Bank

Table 3.A.2 Countries included in the study disaggregated by country-groups.

Destination countries				
Developed				
Australia	Iceland	New Zealand	Romania[a]	Switzerland
Canada	Israel	Norway	Singapore	United States
Hong Kong	Japan			
	EU Core			
	Austria	Denmark	Luxembourg	Sweden
	Belgium	France	Netherlands	United Kingdom[b]
	EU Peripheral			
	Bulgaria	Finland	Latvia	Portugal
	Croatia	Greece	Lituania	Slovak Republic
	Cyprus	Hungary	Malta	Slovenia
	Czech Republic	Ireland	Poland	Spain
	Estonia	Italy		
Developing				
Argentina[c]	Morocco	Saudi Arabia	Turkey	Ukraine
Egypt	Russian Federation			
	Latin American			
	Brazil	Colombia	Mexico	
	Chile	Ecuador	Uruguay	Venezuela
	Asian			
	China	Indonesia	Korea, Republic of	Thailand
	India	Kazhastan	Malaysia	

[a]Although Romania is a EU member since 2007, it is not included in the group of peripheral EU because it is classified as an outlier concerning the level of skilled labour endowments.

[b]Since the 2016 referendum vote to leave the EU, the UK is on course to leave the EU.

[c]Argentina is not included in the Latin American countries' group because German FDI shrank sharply in the year 2000 due to the economic depression that hit the country.

Appendix B

Table 3.B.1 Determinants of FDI in all recipient countries

	1996-2012	
Variable	*incl prob*	*post mean*
ln_sum_gdp	0.2257	0.000
sim_gdp	0.2281	0.000
ln_sq_gdp_diff	**0.7472**	-0.139
ln_sq_gdppc_diff	**1.0000**	0.287
h_urban	0.2697	0.000
ln_h_pop	**0.8351**	1.095
ln_h_gdp	0.4343	0.000
ln_h_gdppc	**0.7310**	0.674
contig	0.1540	0.000
ln_distcap	**0.7654**	-0.230
h_landlocked	**0.7879**	0.470
tdiff	0.2180	0.000
h_skilled	**0.8428**	-3.649
ln_sq_educ_diff	**0.9507**	1.069
sq_skill_diff	**0.7434**	-10.594
ln_educ_gdp_interact	LD	LD
ln_skill_gdp_interact	0.1450	0.000
ln_h_wage	NA	NA
ln_h_popdensity	LD	LD
ln_h_yr_sch	0.4084	0.000
ln_h_area	**0.8563**	-0.139
h_oil	0.1600	0.000
ln_ULC	NA	NA
h_productivity	**0.8262**	-1.697
comlang_off	0.2700	0.000
comlang_ethno	**0.8511**	1.084
colony	0.1238	0.000
comcol	C	C
curcol	C	C
col45	C	C
smctry	C	C
ln_h_ER	**0.8297**	-0.071
EMU	0.1440	0.000

(Continued)

Table 3.B.1 Determinants of FDI in all recipient countries *(Continued)*

	1996-2012	
Variable	*incl prob*	*post mean*
PTA	**1.0000**	-2.819
CU	**1.0000**	2.982
FTA	**1.0000**	2.919
EIA	0.1892	0.000
BIT	0.4509	0.000
h_tradetax	NA	NA
h_duties	NA	NA
ln_educ_open_interact	**0.9765**	-1.170
skill_open_interact	0.3204	0.000
ln_exports	0.4894	0.000
ln_imports	0.2216	0.000
ln_trade	0.3590	0.000
h_open	**1.0000**	2.057
ln_h_KOFGI	**0.9856**	3.085
ln_h_ka_open	NA	NA
ln_h_rail	NA	NA
ln_h_phone	LD	LD
ln_h_internet	NA	NA
ln_h_lines	0.1323	0.000
ln_h_cell	**0.6130**	0.071
h_pr	**0.9635**	0.363
h_cl	**0.9777**	-0.356
h_va	**0.9753**	0.040
h_pv	0.3301	0.000
h_ge	0.4888	0.000
h_rq	0.2383	0.000
h_rl	**0.9993**	-0.0293
h_cc	0.1811	0.000

Notes: posterior inclusion probabilities larger than 0.5. LD and C stand for variables dropped for being linear dependent and constant, respectively. NA stands for variables not included because of data availability.

Table 3.B.2 Determinants of FDI in developed and developing countries

	1996-2012			
	Developed		Developing	
Variable	*incl prob*	*post mean*	*incl prob*	*post mean*
ln_sum_gdp	0.1783	0.000	0.4547	0.000
sim_gdp	0.1802	0.000	**0.7201**	-0.656
ln_sq_gdp_diff	0.1537	0.000	0.4237	0.000
ln_sq_gdppc_diff	**1.0000**	0.175	**0.9850**	0.840
h_urban	**0.8439**	-2.070	0.2710	0.000
ln_h_pop	**0.9269**	1.347	0.2465	0.000
ln_h_gdp	0.2816	0.000	**0.7405**	0.950
ln_h_gdppc	0.3114	0.000	0.2116	0.000
contig	0.3228	0.000	C	C
ln_distcap	0.2144	0.000	0.1404	0.000
h_landlocked	0.2036	0.000	**1.0000**	-1.624
tdiff	0.1840	0.000	**0.9996**	0.172
h_skilled	**0.9571**	-6.679	0.1982	0.000
ln_sq_educ_diff	0.3068	0.000	**0.9986**	1.031
sq_skill_diff	**0.9390**	-33.636	**0.5023**	0.000
ln_educ_gdp_interact	LD	LD	LD	LD
ln_skill_gdp_interact	0.1949	0.000	0.2438	0.000
ln_h_wage	NA	NA	NA	NA
ln_h_popdensity	LD	LD	0.1393	0.000
ln_h_yr_sch	0.3281	0.000	0.3424	0.000
ln_h_area	**0.9984**	-0.340	0.1358	0.000
h_oil	NA	NA	**0.9991**	-2.554
ln_ULC	NA	NA	NA	NA
h_productivity	**0.9880**	-4.971	**0.9000**	-1.313
comlang_off	**0.9922**	2.867	C	C
comlang_ethno	**0.9846**	-2.401	C	C
colony	0.1732	0.000	C	C
comcol	C	C	C	C
curcol	C	C	C	C
col45	C	C	C	C
smctry	C	C	C	C
ln_h_ER	**0.9799**	-0.239	**1.0000**	-0.112
EMU	0.4651	0.000	C	C

(Continued)

Table 3.B.2 Determinants of FDI in developed and developing countries *(Continued)*

	1996-2012			
	Developed		Developing	
Variable	*incl prob*	*post mean*	*incl prob*	*post mean*
PTA	**1.0000**	-2.784	**0.7758**	-0.204
CU	**1.0000**	1.742	**0.9998**	1.362
FTA	**1.0000**	2.596	LD	LD
EIA	**0.9657**	0.651	0.1636	0.000
BIT	0.1596	0.000	0.0986	0.000
h_tradetax	NA	NA	NA	NA
h_duties	NA	NA	NA	NA
ln_educ_open_interact	0.3261	0.000	**1.0000**	-1.554
skill_open_interact	0.3036	0.000	**0.5200**	-1.016
ln_exports	**0.6820**	0.360	0.3103	0.000
ln_imports	0.3699	0.000	**0.9598**	-0.394
ln_trade	0.4290	0.000	**0.7853**	0.585
h_open	**0.9899**	0.826	**1.0000**	1.949
ln_h_KOFGI	**1.0000**	8.874	**1.0000**	-2.809
ln_h_ka_open	NA	NA	NA	NA
ln_h_rail	NA	NA	NA	NA
ln_h_phone	LD	LD	LD	LD
ln_h_internet	NA	NA	0.1660	0.000
ln_h_lines	0.1732	0.000	**1.0000**	-0.430
ln_h_cell	0.1873	0.000	**1.0000**	0.370
h_pr	NA	NA	0.1550	0.000
h_cl	NA	NA	0.4513	0.000
h_va	**0.9999**	0.054	**0.5338**	0.003
h_pv	**0.8715**	0.013	0.0961	0.000
h_ge	**0.9636**	0.058	0.1303	0.000
h_rq	0.2378	0.000	**0.9920**	0.012
h_rl	0.2115	0.000	0.3849	0.000
h_cc	**0.9918**	-0.061	0.3844	0.000

Notes: posterior inclusion probabilities larger than 0.5. LD and C stand for variables dropped for being linear dependent and constant, respectively. NA stands for variables not included because of data availability.

Table 3.B.3 Determinants of FDI in Latin American and Asian countries

	1996-2012			
	Latin America		Asia	
Variable	*incl prob*	*post mean*	*incl prob*	*post mean*
ln_sum_gdp	0.3488	0.000	**0.9595**	2.885
sim_gdp	**0.5574**	0.000	0.0902	0.000
ln_sq_gdp_diff	**0.6432**	-1.140	0.0378	0.000
ln_sq_gdppc_diff	0.4709	0.000	0.0486	0.000
h_urban	0.0541	0.000	0.0819	0.000
ln_h_pop	**0.7135**	-6.100	0.1665	0.000
ln_h_gdp	0.3950	0.000	0.1054	0.000
ln_h_gdppc	0.4991	0.000	**0.8545**	-1.070
contig	C	C	C	C
ln_distcap	0.0383	0.000	0.1400	0.000
h_landlocked	C	C	**0.8926**	-2.480
tdiff	**0.9989**	1.871	0.1330	0.000
h_skilled	0.1369	0.000	0.0614	0.000
ln_sq_educ_diff	0.2698	0.000	0.0424	0.000
sq_skill_diff	0.0484	0.000	0.0447	0.000
ln_educ_gdp_interact	LD	LD	LD	LD
ln_skill_gdp_interact	**0.8580**	0.520	0.0385	0.000
ln_h_wage	NA	NA	NA	NA
ln_h_popdensity	0.2285	0.000	0.1269	0.000
ln_h_yr_sch	**0.7398**	-2.310	**0.9745**	4.034
ln_h_area	**0.9949**	5.101	0.0753	0.000
h_oil	0.0343	0.000	0.0664	0.000
ln_ULC	NA	NA	NA	NA
h_productivity	0.4088	0.000	0.0436	0.000
comlang_off	C	C	C	C
comlang_ethno	C	C	C	C
colony	C	C	C	C
comcol	C	C	C	C
curcol	C	C	C	C
col45	C	C	C	C
smctry	C	C	C	C
ln_h_ER	0.1667	0.000	**0.6292**	-0.090
EMU	C	C	C	C

(Continued)

Table 3.B.3 Determinants of FDI in Latin American and Asian countries *(Continued)*

	1996-2012			
	Latin America		Asia	
Variable	*incl prob*	*post mean*	*incl prob*	*post mean*
PTA	0.0361	0.000	0.0226	0.000
CU	C	C	C	C
FTA	LD	LD	LD	LD
EIA	LD	LD	LD	LD
BIT	0.0427	0.000	C	C
h_tradetax	NA	NA	NA	NA
h_duties	NA	NA	NA	NA
ln_educ_open_interact	0.1621	0.000	0.1190	0.000
skill_open_interact	0.0568	0.000	0.1793	0.000
ln_exports	0.0624	0.000	0.0808	0.000
ln_imports	0.1173	0.000	0.0407	0.000
ln_trade	0.1389	0.000	0.0598	0.000
h_open	**0.8668**	-2.685	0.2812	0.000
ln_h_KOFGI	0.0393	0.000	0.0377	0.000
ln_h_ka_open	NA	NA	NA	NA
ln_h_rail	NA	NA	NA	NA
ln_h_phone	LD	LD	LD	LD
ln_h_internet	**0.9991**	0.391	**0.9995**	0.240
ln_h_lines	**0.9620**	-0.733	0.2519	0.000
ln_h_cell	0.0627	0.000	0.1213	0.000
h_pr	**0.9804**	0.171	0.0247	0.000
h_cl	0.0368	0.000	**0.9847**	0.398
h_va	**0.9749**	-0.028	**0.9966**	0.027
h_pv	**0.8135**	0.010	0.0427	0.000
h_ge	0.0450	0.000	0.0210	0.000
h_rq	0.0660	0.000	0.0522	0.000
h_rl	0.0349	0.000	0.4436	0.000
h_cc	0.0331	0.000	0.0235	0.000

Notes: posterior inclusion probabilities larger than 0.5. LD and C stand for variables dropped for being linear dependent and constant, respectively. NA stands for variables not included because of data availability.

Table 3.B.4 Determinants of FDI in EU regions

	1996-2012			
	Core		Periphery	
Variable	*incl prob*	*post mean*	*incl prob*	*post mean*
ln_sum_gdp	0.2626	0.000	0.0750	0.000
sim_gdp	**0.5445**	0.617	**0.7252**	0.554
ln_sq_gdp_diff	0.0862	0.000	0.3535	0.000
ln_sq_gdppc_diff	0.0617	0.000	0.0441	0.000
h_urban	0.0943	0.000	**0.6696**	-2.130
ln_h_pop	0.3385	0.000	0.2322	0.000
ln_h_gdp	0.4688	0.000	0.2760	0.000
ln_h_gdppc	**0.6483**	1.742	0.1797	0.000
contig	0.3782	0.000	**0.9975**	1.345
ln_distcap	**0.7074**	2.790	**0.9970**	1.563
h_landlocked	**0.9584**	1.012	0.0735	0.000
tdiff	0.1562	0.000	0.1054	0.000
h_skilled	0.0551	0.000	0.1683	0.000
ln_sq_educ_diff	0.0548	0.000	0.0768	0.000
sq_skill_diff	0.0556	0.000	0.2780	0.000
ln_educ_gdp_interact	LD	LD	LD	LD
ln_skill_gdp_interact	0.0231	0.000	0.0399	0.000
ln_h_wage	**0.6277**	1.455	NA	NA
ln_h_popdensity	LD	LD	LD	LD
ln_h_yr_sch	0.3969	0.000	0.2420	0.000
ln_h_area	**0.8482**	-0.570	0.1544	0.000
h_oil	0.1952	0.000	0.0447	0.000
ln_ULC	0.2857	0.000	NA	NA
h_productivity	0.2200	0.000	**0.6793**	-2.795
comlang_off	0.1626	0.000	C	C
comlang_ethno	0.1784	0.000	C	C
colony	C	C	0.1077	0.000
comcol	C	C	C	C
curcol	C	C	C	C
col45	C	C	C	C
smctry	C	C	C	C
ln_h_ER	0.3583	0.000	**1.000**	0.331
EMU	0.1216	0.000	0.0420	0.000

(Continued)

Table 3.B.4 Determinants of FDI in EU regions *(Continued)*

	1996-2012			
	Core		Periphery	
Variable	*incl prob*	*post mean*	*incl prob*	*post mean*
PTA	C	C	**0.9595**	0.674
CU	C	C	**0.5923**	-0.400
FTA	C	C	LD	LD
EIA	0.0264	0.000	0.0544	0.000
BIT	C	C	**0.9997**	-0.989
h_tradetax	0.0480	0.000	NA	NA
h_duties	0.0247	0.000	0.1601	0.000
ln_educ_open_interact	0.0607	0.000	0.0724	0.000
skill_open_interact	0.1061	0.000	0.0990	0.000
ln_exports	**0.5320**	0.156	**0.5620**	0.000
ln_imports	0.1099	0.000	**0.5089**	0.000
ln_trade	0.2178	0.000	**0.6035**	0.255
h_open	0.0691	0.000	0.0716	0.000
ln_h_KOFGI	0.1574	0.000	0.3026	0.000
ln_h_ka_open	NA	NA	0.3208	0.000
ln_h_rail	0.2462	0.000	**1.000**	1.162
ln_h_phone	LD	LD	LD	LD
ln_h_internet	**0.6296**	0.122	0.0532	0.000
ln_h_lines	**0.9977**	-1.174	**1.000**	-1.072
ln_h_cell	0.1939	0.000	**1.000**	0.360
h_pr	C	C	0.0601	0.000
h_cl	0.0247	0.000	0.0380	0.000
h_va	0.0653	0.000	0.0527	0.000
h_pv	0.0263	0.000	**0.6289**	-0.008
h_ge	0.0376	0.000	0.3407	0.000
h_rq	0.2044	0.000	0.0483	0.000
h_rl	0.0563	0.000	0.4290	0.000
h_cc	0.0359	0.000	0.1713	0.000

Notes: posterior inclusion probabilities larger than 0.5. LD and C stand for variables dropped for being linear dependent and constant, respectively. NA stands for variables not included because of data availability.

Chapter 4

Alternative estimators for the FDI gravity model: an application to German outward FDI

4.1 Introduction and motivation

Since Bergstrand and Egger (2007) and Head and Ries (2008) established a theoretical foundation for the gravity equations for foreign direct investment (FDI), a popular and empirically successful strand of research has used a gravity approach to investigate the cross-country pattern of FDI. Even though the gravity model has proved to be a useful tool to approximate bilateral FDI flows in most empirical studies (see Blonigen and Piger (2014) for extensive overview), a consensus on its empirical application is still missing. Indeed, new developments in the literature (such as new theoretical approaches, the use of panel data and other econometric improvements) have highlighted several empirical problems in estimating the gravity equation and generated a debate with divergent opinions about the best performing estimator. Empirical analyses of the factors determining FDI across countries have employed a variety of econometric specifications and estimation methods. A primary concern is related to the econometric problems encountered by estimating the gravity equation in its additive form (i.e. log-log form). Silva and Tenreyro (2011, 2006) argued that the conventional practice in the literature of log-linearizing the gravity model and subsequent estimation in its additive form through Ordinary Least Squares (OLS) could not deal with zero-valued bilateral FDI observations and heteroskedasticity in the data and thereby, it led to misleading estimates. Consequently, they propose to estimate the gravity model in its multiplicative form.

Another concern involves the choice of the most suitable estimation method that allows to deal with zero-valued bilateral FDI observations. Zero values are frequent in FDI data and neglecting them might provide inconsistent estimates. Thereby, several alternatives on how to address this issue have been proposed in the literature. The most succesful and frequently used has been the Poisson Pseudo Maximum Likelihood (PPML) estimator, a special case of the Generalized Linear Model (GLM) framework, posit by Silva and Tenreyro (2006). Silva and Tenreyro (2006) argue that the PPML estimator naturally deals with zero FDI observations and is consistent in the presence of heteroskedasticity.

Nevertheless, the recent literature has highlighted that the PPML is not without its cons by comparing the performance of alternative estimators. The studies of Martin and Pham (2008), Burger et al. (2009), Siliverstovs and Schumacher (2009), Martínez-Zarzoso (2013), Westerlund and Wilhelmsson (2011), Gómez-Herrera (2013), Head and Mayer (2014) and Egger and Staub (2016) have followed a model selection approach in order to identify the best performing estimator for the gravity model of trade. Yet, the results obtained are still controversial.

In this paper, we aim to contribute to the debate by analyzing the performance of different estimators in a GLM framework. We compare several methods estimating gravity models in their multiplicative form via GLMs with a log-link using different distribution families:

Poisson Pseudo Maximum Likelihood (PPML), Gamma Pseudo Maximum Likelihood (GPML), Negative Binomal Pseudo Maximum Likelihood (NBPML) and Gaussian GLM. As an empirical application, we examine the determinants of German outward FDI using a three-dimensional (i,j,t) FDI dataset covering the period 1996-2012. We undertake a careful consideration of the robustness of the estimators and conclude that NBPML is the best performing estimator for this application, followed by GPML.

Our study makes two important contributions to the existing literature. Firstly, we provide a comprehensive empirical evidence of the determinants of German outward FDI. Despite Germany is among the largest investors worldwide, the studies analyzing its determinants are relatively scarce. Second, to the best of our knowledge, there exist no study within the gravity model literature that compares the performance of different estimators for the FDI gravity model as the extant literature so far is dealing with trade.

The reminder of the paper proceeds as follows. Section 2 discusses the main estimation techniques and econometric problems in the empirical application of the gravity model. Section 3 then lays out the alternative estimators considered together with the data. Section 4 reports the results. Robustness checks are contained in Section 5, and Section 6 concludes.

4.2 Model uncertainty in gravity model estimation

The Gravity approach to FDI describes the volume of bilateral FDI between two countries as positively related to their economic sizes and negatively to the distance between them. During the last decade some of the literature on FDI tried to generalize the use of the gravity approach to analyze FDI patterns (Brainard (1997), Eaton and Tamura (1994)). Nonetheless, there was a lack of theoretical foundation for the gravity equations for FDI. Since Bergstrand and Egger (2007) such a theoretical foundation does exist. They extend the 2x2x2 knowledge-capital model in Markusen and Maskus (2002), by adding an extra factor and country, and derive a specification for the FDI gravity equation that explains its empirical fit to the data. This paper, together with the one by Head and Ries (2008) are considered the only two formal general equilibrium theories for FDI. Subsequently, more research followed and the theoretical justification of the gravity model for FDI is not longer questioned. Kleinert and Toubal (2010) illustrate how an aggregate FDI equation can be derived from different theoretical models. In particular, we adopt here the Kleinert and Toubal (2010) horizontal model where firms can serve the foreign market j either by producing abroad or by exporting. The gravity equation estimated by Kleinert and Toubal (2010) is as follows:

$$AS_{ij} = s_i(\tau D_{ij}^{\eta 1})^{(1-\sigma)(1-\in)} m_j \tag{4.1}$$

where AS_{ij} are aggregate sales of foreign affiliates from firm i in j; s_i and m_j denote home and host country's market capacity, respectively and $\tau D_{ij}^{\eta 1}$ stands for geographical distance between i and j where τ represents the unit distance costs and $\eta_1 > 0$.

Equation 4.1 can be log-linearized as

$$ln(AS_{ij}) = \alpha_1 + \zeta_1 ln(s_i) - \beta_1 ln(D_{ij}) + \xi_i ln(m_j) \tag{4.2}$$

Despite the theoretical foundations of FDI gravity models and its popularity in the empirical literature, the problems that arise in its application has raised a debate on the best performing estimator. Heteroskedasticity in the data and how to deal with zero values in the dependent variable are the two most common specific problems often encountered in gravity model estimation (see Matyas (2017)). Since the amount of estimation methods proposed in the literature to tackle these issues is rather large, we split this section in two subsections. In Sect.4.2.1 we present estimation methods that estimate the gravity model in its additive form (i.e. log-log form), while in Sect.4.2.2 we report estimation methods estimating the gravity model in its multiplicative form.

4.2.1 Additive functional form estimators

Traditionally, gravity models have been applied to cross-sectional data and estimated in its additive form through OLS or pooled OLS (Brainard (1997), Brenton and Di Mauro (1999), Buch et al. (2003)). However, the OLS estimation of the log-linearized gravity equation has been argued to yield biased and inconsistent estimates due to (1) the violation of the homoskedasticity assumption, (2) the bias that results from the log-linearization and (3) the failure to model zero FDI flows.

Later on, the availability of panel data allowed to estimate gravity models by panel econometric methods. The panel framework permits to control for the unobserved heterogeneity, that is common in this data, as well as the multilateral resistance by the introduction of exporter-and-year and importer-and-year fixed effects in addition to dyadic fixed effects (Baldwin and Taglioni (2007), Martínez-Zarzoso et al. (2009)). The two most frequently used panel data techniques have been the fixed effects and random effects estimators. The rationale behind the fixed effects estimator is to control for the unobserved time-invariant idiosyncratic characteristics correlated with the independent variables that capture heterogeneities across individuals. Alternatively, the unobserved individual effect is assumed to be random and uncorrelated with the regressors in the random effects estimator. Under the null hypothesis of no correlation, even though both estimators yield consistent parameters, the random efffects is the most efficient; whereas, under the alternative hypothesis, only the fixed effects is consistent. A drawback of fixed effects estimators is that by including dyadic fixed effects, the coefficients of time-invariant explanatory variables (such as distance, common language or common borders, among others) can no longer be estimated as they are perfectly collinear with the fixed effects. Thereby, this is an important limitation when the researcher wants to estimate the effect of these variables. Furthermore, the number of fixed effects included may lead to computational difficulties. These issues together with missing FDI observations and thus, unbalanced panels, are the main econometric problems for panel data estimation of the gravity model. A thorough discussion of the econometric issues and estimation techniques for gravity models when using panel data is provided by Baltagi et al. (2014).

Another estimation method for the gravity model in its additive form, is the Tobit model initially suggested by Eaton and Tamura (1994) (ET-Tobit) for FDI gravity models and then implemented by Wei (2000). An advantage of the Tobit estimator is that it deals with the problem of zero-valued FDI observations by replacing the zero values by a constant for the sake of the logarithmic convenience. However, it could be too restrictive because it assumes that the same mechanism generates both the selection and the outcome equation; in the sense that the same variables would determined the decision to invest and the amount of investment (Gómez-Herrera (2013)). Consequently, several studies recommend the use of the Heckman two-step estimator as it presents a better fit by assuming independence among the selection and outcome equations. The Heckman model is presented in the next subsection concerning nonlinear estimators.

All these estimators, except for the Tobit methods, can not deal with the problem of excessive zeros in the dependent variable. A solution proposed by the literature has been to add a small constant to the dependent variable before the logarithmic transformation. Nonetheless, this approach has been critized as it lacks a theoretical foundation and results strongly depends on the magnitude of the constant (Head and Mayer (2014)).

4.2.2 Multiplicative functional form estimators

Based on Jensen's inequality, that states that $E[ln(\varepsilon_{ij})] \neq [lnE(\varepsilon_{ij})]$, Silva and Tenreyro (2006) demonstrate that the OLS estimation of models, like the gravity model in Equation 4.2, in its additive functional form provides inconsistent estimates. This is because, under heteroskedastic data, the expected values of the log-linearized error term ($E[ln(\varepsilon_{ij})]$) will depend on the regressors, thus leading to fallacious inferences. Consequently, Silva and

Tenreyro (2006) recommend estimating constant-elasticity models (such as the gravity model) in its original multiplicative form and propose the use of the PPML estimator rather than OLS. Following Silva and Tenreyro (2006), the recent literature has turned towards multiplicative functional form estimators, which include GLMs as well as two part models, such as the Heckman sample selection model. Among GLMs, the PPML estimator has been considered the “workhorse” estimator of the gravity equation, as it has been shown to be adequate in the presence of heteroskedasticity and zero values in the dependent variable (Silva and Tenreyro (2006)). Nevertheless, the recent literature has highlighted some drawbacks of the PPML estimator and divergent opinions have arose about which is the best performing estimator for the gravity equation.

Martin and Pham (2008) show through Monte-Carlo simulations that the PPML estimator could potentially result in limited-dependent variable bias under the presence of excessive zeros in the dependent variable. Furthermore, they posit that the ET-Tobit estimator outperforms PPML as long as heteroskedasticity is properly controlled for. A second argument that has been made is that the PPML estimator might yield inconsistent estimates in the presence of overdispersion due to a misspecification of the mean and thus, different Poisson-family alternatives to PPML have been recommended. Burger et al. (2009) recommend the use of the NBPML to allow for overdispersion; whereas, when there is a large share of zero-values in the dependent variable, they support the use of the zero-inflated Negative binomial (ZINBPML) and zero-inflated Poisson model (ZIPPML). Martínez-Zarzoso (2013) questions also the performance of the PPML estimator and shows through Monte-Carlo simulations that alternative estimation methods outperform the PPML. They found that, under an unknown form of heteroskedasticity, Feasible Generalised Least Squares (FGLS) is the preferred estimator; whereas, in the absence of zeros in the dependent variable, GPML performs better.

In light of these concerns raised by the literature, Silva and Tenreyro (2011) extend their simulation study in Silva and Tenreyro (2006) and demonstrate that their results validate the use of the PPML estimator even under overdispersion or excessive zeros in the dependent variable. Similarly, Head and Mayer (2014) show that PPML and GPML are consistent in the presence of overdispersion. Nonetheless, they posit that GPML performs better for certain empirical applications than PPML.

A comprehensive survey of alternative estimation techniques for the trade gravity model can be found in Gómez-Herrera (2013). The study suggests, based on an empirical exercise, that the Heckman sample selection model is the best performing estimator under the existence of heteroskedasticity and excess zeros in the dependent variable. Focusing on the performance of GLM estimators, Egger and Staub (2016) conduct a set of Monte Carlo simulations together with an empirical application, and found that the NBPML is the prefered estimator for the chosen specification of the trade gravity equation.

Overall, even though the estimation of the gravity equation in its multiplicative form is no longer in doubt, it is not certain which is the best performing estimator. Bearing in mind this uncertainty, several studies recommend to use the PPML in comparison to alternative estimators in order to select the appropriate estimator for a particular application (Martínez-Zarzoso (2013), Head and Mayer (2014)).

4.3 Econometric methodology and data

While the literature points towards the multiplicative functional specification of the gravity model, there is uncertainty about the optimal nonlinear estimator. Alternatively to the PPML, recent studies recommend other exponential-family models; see Martínez-Zarzoso (2013), Head and Mayer (2014) and Egger and Staub (2016). Furthermore, these studies all claim that the proper estimator for the gravity model largely depends on the data, thereby there is a need for additional empirical analysis. To contribute to this strand of the literature, which has attracted the interest of many researchers, we compare several estimators in a GLM framework. GLMs estimate the gravity models in their multiplicative form as:

$$y_i = exp(x_i\beta_i)\varepsilon_i \tag{4.3}$$

where $\mathbb{E}(\varepsilon_i|x) = 1$, y_i is the dependent variable, x_i are the explanatory variables and β are the parameters to be estimated.

GLMs estimators are maximum likelihood estimators that are based on an assumed linear exponential family (LEF) density, a linear predictor and a link function - which provides the relationship between the linear predictor and the mean (McCullagh and Nelder (1989); Nelder and Wedderburn (1972)). Our modeling framework includes GLMs with a logarithmic link function and four exponential family distributions, the key attributes of which is the assumption on the functional form of $V[y_i|x]$.

Table 4.1 Conditional mean-variance relationships of GLM estimators

Estimator	Assumptions on $V[y_i|x]$
PPML	$V[y_i|x] \propto \mathbb{E}[y_i|x]$
GPML	$V[y_i|x] \propto \mathbb{E}[y_i|x]^2$
NBPML	$V[y_i|x] \propto \mathbb{E}[y_i|x] + kE[y_i|x]^2$
Gaussian GML	$V[y_i|x] = 1$

Source: Own elaboration.

Table 4.1 shows the conditional mean-variance relationships of each of the LEF of distributions studied here. We obtain the PPML estimator under the assumption that the conditional variance is proportional to the conditional mean. GPML and NBPML, in turn, are obtained when the variance is a function of higher powers of the mean; whereas the Gaussian GLM is obtained when the variance equals 1. In the following subsections, we briefly present the alternative estimators and highlight its merits and drawbacks. Then, we describe the data.

4.3.1 Poisson pseudo maximum likelihood

Considering the gravity model in Equation 4.3, the PPML estimator computes β, the vector of parameters of interest, from the following first-order conditions:

$$\sum_{i=1}^{n}[y_i - exp(x_i\tilde{\beta})]x_i = 0 \tag{4.4}$$

The PPML estimator is a special case of the GLM framework, in which the variance is assumed to be proportional to the mean, $V[y_i|x] \propto \mathbb{E}[y_i|x]$. The proportionality assumption implies that the PPML estimator equally weights all observations.

Silva and Tenreyro (2006) argue that PPML estimator has a number of interesting properties. First, it provides a natural way of dealing with zero-valued FDI observations as the functional form allows to include the dependent variable in levels. Second, even though the proportionality assumption does not usually holds, it provides consistent estimates in the presence of heteroskedasticity once a robust covariance matrix is considered. Nevertheless, some studies claim that it performs relatively poorly in the presence of overdispersion and excess zeros in the dependent variable (Burger et al. (2009)). More recently, Pfaffermayr (2019) proves that the standard errors of the PPML estimated parameters are downward biased in cross-section data.

4.3.2 Gamma pseudo maximum likelihood

The GPML estimator defines the following set of first-order conditions:

$$\sum_{i=1}^{n}[y_i - exp(x_i\check{\beta})]exp(-x_i\check{\beta})x_i = 0 \tag{4.5}$$

It is based on the assumption that the variance is a function of higher powers of the mean, $V[y_i|x] \propto \mathbb{E}[y_i|x]^2$ and thereby, this estimator down-weights observations with larger means. This assumption has been considered a drawback by some researchers and a merit by others. In particular, Silva and Tenreyro (2006) points out that GPML might give excessive weight to the observations that are more prone to measurement errors.[1] Nevertheless, Egger and Staub (2016) states that this could lead to efficiency gains whenever those observations with larger means exhibit also a larger variance (i.e. noisier observations). Several empirical studies show that GPML outperforms alternative estimators (see, Manning and Mullahy (2001) or Martínez-Zarzoso (2013), among others). GPML has also been widely used to address the zero-valued observations problem.

4.3.3 Negative binomial pseudo maximum likelihood

The NBPML estimator is defined by:

$$Pr[I_{ij}] = \frac{\Gamma(I_{ij}+\alpha^{-1})}{I_{ij}!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_{ij}}\right)^{\alpha^{-1}}\left(\frac{\mu_{ij}}{\alpha^{-1}+\mu_{ij}}\right)^{I_{ij}} \tag{4.6}$$

where $\mu_{ij} = exp(\alpha_0 + \beta' X_{ij} + \eta_i + \gamma_j)$, Γ is the gamma function and α is a parameter that determines the degree of dispersion in predictions.

[1] Silva and Tenreyro (2006) argue that GPML might not be desirable for trade data, as data from larger countries (measured in terms of GDP) tend to be of higher quality and thereby, country pairs with little bilateral trade are more prone to measurement errors than the observations with large bilateral trade.

It assumes that the variance is a specific quadratic function of the mean, $V[y_i|x] \propto \mathbb{E}[y_i|x] + kE[y_i|x]^2$. Thus, it has been frequently used to allow for overdispersion in the data. Again, it down-weights observations with larger means what can be seen either as a merit or a drawback as previoulsy explained for the case of GPML. The primary reason for applying NBPML is to improve efficiency as it comprises both the PPML and GPML assumptions (Bosquet and Boulhol (2014)). Nevertheless, an important limitation is that it strongly depends on the units of measurement for the dependent variable (Bosquet and Boulhol (2014), Head and Mayer (2014)). Furthermore, Burger et al. (2009) shows that NBPML is inconsistent when the dependent variable exhibits a substantial amount of zeros.

4.3.4 Gaussian GLM

Gaussian GLM provides the estimates of β by solving the following first-order conditions:

$$\sum_{i=1}^{n}[y_i - exp(x_i\check{\beta})]exp(x_i\check{\beta})x_i = 0 \tag{4.7}$$

Notice Gaussian GLM estimator is equivalent to the Nonlinear least-squares by GLM for y with a log link, and an additive homoscedastic error term (Manning and Mullahy (2001)), thereby we will refer to Gaussian GLM or NLS indistinctly. Gaussian GLM assumes the variance equals 1, $V[y_i|x] = 1$. It assigns more weight to noisier observations (in the sense of a larger variance) and thus, leads to a reduction in efficiency. It has been found to perform very badly under heteroskedasticity and presents sample selection bias Silva and Tenreyro (2006). Despite these limitations, it has also been applied in the gravity model empirical literature and frequently used alongside alternative estimators for the sake of comparison (see Silva and Tenreyro (2006), Gómez-Herrera (2013) or Egger and Staub (2016), among others).

4.3.5 Data

The analysis makes use of the data set described in Chapter 3, which provides information on German outward FDI stock over the period 1996-2012 in 59 destination countries (38 developed and 21 developing). We refer the reader to Tables 3.A.1 and 3.A.2 in Chapter 3 for a detailed description of the data sources and countries included.

A short note should be made regarding our FDI measure. We rely on FDI stocks extracted from the Bilateral UNCTAD FDI Statistics. Nevertheless, we bear in mind that this FDI measure may be somewhat distortive due to corporate accounting practices and valuation methods across countries and hence, results should be interpreted with caution. Despite UNCTAD FDI statistics do not report FDI to Special purpose entities (SPEs), it may still be capturing statistical artefacts, such as round tripping.[2] In 2014 the IMF's Balance of Payments and International Investment Position Manual (BPM6) and the fourth edition of OECD's Benchmark Definition of Foreign Direct Investment (BD4) provide new guidelines for FDI compilation in order to improve the quality of the data. However, Blanchard and Acalin (2016) posit that these practices do not completely remove the uncertainty surrounding the quality of FDI data. They examine the correlation between FDI inflows and outflows for the US, as well as the correlation between outflows and the US policy rate, and provide evidence of the speculative nature of FDI measures. More recently, Dellis et al. (2017) using a new OECD database on FDI statistics (OECD BMD4) that filters out the distortive effects from the data together with the approach proposed by Blanchard and Acalin (2016), found that their results were robust to the use of the "non-cleaned" FDI data set from UNCTAD. Thereby, even though our data might not be completely filtered out, it allows us to provide insights on the long-run behavior of investment decisions.

[2]UNCTAD FDI statistics do not report FDI to Special purpose entities (SPEs) for Austria, Cyprus, Hungary, Luxembourg, the Netherlands and Portugal.

Explanatory variables have been selected according to the results provided in Chapter 3. We remind the reader that in the previous chapter we applied a BMA analysis to identify the main determinants of German outward FDI. To address the heterogeneity in the destinations, we considered different recipient-country groups: developed, developing, Latin American, Asian, EU core and EU peripheral. Our results provided the inclusion probabilities (PIP) of each potential FDI determinants; those variables with a PIP above 0.50 were considered robust determinants. Table 4.A.1 reports a summary of the variables that exhibited a PIP greater than 0.50 for each country-group. A drawback of this methodology was that it only computed model averaged estimated coefficients. Thereby, this study aims to provide a more accurate estimation of the coefficients. In doing so, we depart from the variables obtained by BMA and compare alternative GLM estimators applying a backward elimination (BE) procedure.[3]

4.4 Results

In this Section, we report the results obtained using the alternative GLM estimators. In order to asses the performance of the different GLM estimators, we rely on different measures of goodness-of-fit (see Silva and Tenreyro (2006) or Martínez-Zarzoso (2013), among others). First, the Ramsey (1969) Regression Equation Specification Error Test (RESET) is computed to asses the general misspecification of the estimators. If the test rejects the null hypothesis of a good specification, it would mean either that the model is inappropriate due to its functional form or that some relevant information is missing.[4] We also provide the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). In both cases, a smaller value generally indicates a better model fit. We compare also the deviance and dispersion of the

[3]We have also use a stepwise backward selection procedure and results remain stable.

[4]Ramsey's Reset test is essentially a test for the correct specification of the conditional expectation, by testing the significance of an additional regressor constructed as $(x'b)^2$, where b denotes the vector of estimated parameters (Silva and Tenreyro (2006)).

different GLM families. The lowest values indicate a better model fit. Finally, we compute three goodness-of-fit functions: the bias, the mean squared error (MSE) and the absolute error loss. The latter is considered more appropriate than the bias as shown in Martínez-Zarzoso (2013).

Tables 4.2 to 4.8 report the estimated coefficients as well as the goodness-of-fit statistics for the whole dataset and for the different country-groups considered. Furthermore, we discuss also graphical techniques to assess the validity of the models. We provide plots of the residuals most widely used in model selection for GLMs, the Pearson and deviance residuals (McCullagh and Nelder (1989)). To informally check the validity of the assumed variance function, we examine the scatterplots of the Pearson residuals in Figures 4.1 to 4.13. A wrongly specified variance function will result in a trend in the mean. Thereby, we should expect (mean-)independence of the Pearson residuals of the conditional mean (i.e. a horizontal line) for a proper especification of the variance function. Nevertheless, the deviance residuals are generally preferred to the Pearson residuals as pointed out by McCullagh and Nelder (1989). Thus, we plot the density of deviance residuals for the different GLM estimators in order to gain further insights on the adequacy of the variance function in Figures 4.2 to 4.14. The deviance residuals are approximately normally distributed if the model is correctly specified. Following Egger and Staub (2016), we plot the kernel density of deviance residuals illustrated by the black dashed curve together with a normal density plot based on the same variance for readibility.

Comparing the results of the different estimators, we observe that for all samples, GPML and NBPML yield the same results with similar estimated coefficients and signs. Something similar happens for the PPML and Gaussian GLM estimators. Because we consider GLMs with a logarithmic link function, the estimated coefficients can be interpreted as semi-elasticities (Cameron and Trivedi (2009)).[5]

[5]For a continuous variable in levels or a dummy variable, semi-elasticity is equal to $[exp(\beta)-1]*100$. Note that for continuous variables this is roughly equivalent to $(\beta*100)$.

Overall, among the alternative estimators, the Gaussian GLM fails to pass the RESET test in most of the samples suggesting some sort of misspecification, either because of an inappropriate specification of the model due to its functional form or the omission of relevant information. The NBPML presents the lowest AIC, whereas GPML exhibits the lowest BIC. Concerning the overall deviance and the dispersion of the deviance, the GPML estimator presents the better fit. As regards the goodness-of-fit functions, our results show that all the estimators exhibit a bias, variance and error loss of similar magnitudes. However, PPML and GPML present the lowest bias; whereas, GPML shows the smallest variance; and NBPML exhibits the least error loss. The bias in the estimators could be due to the omission of relevant variables. Recently, Basu (2019) has shown, for OLS estimators, that a bias can be caused by either the omission of relevant variables or the inclusion of irrelevant variables.

Acccording to the graphical tecniques, the Pearson residuals suggest that GPML and NBPML perform better than PPML and Gaussian GLM. Likewise, the deviance residuals provides further evidence for NBPML. Taken together, the goodness-of-fit statistics as well as the visual inspection of the residuals suggest that NBPML and GPML perform the best. Yet, taking into account that the deviance residuals are recomended by many researchers (see McCullagh and Nelder (1989), among others) for model selection of GLMs, NBPML appears to be the best estimator. Our findings are in line with previous empirical studies that question the ad hoc estimation of the gravity model by PPML and recommend the use of alternative estimators. In particular, our results are in line with Egger and Staub (2016) who compare several estimators for the gravity model of trade through a Monte Carlo experiment and concludes that NBPML appears to be the best estimator for their application. Likewise, Martínez-Zarzoso (2013) shows also through Monte Carlo simulations that GPML outperforms PPML. In the next subsections, we describe the results obtained by our preferred models: NBML and GPML. Notice that we do not distinguish among the estimators as, even though they slightly differ in the magnitude of the coefficients, they yield the same results with only a few exceptions. The exceptions are the robust evidence reported by GPML for distance (*ln_distcap*) in core EU countries and competitiveness (*ln_h_ER*) in peripheral EU countries.

Following up on Chapter 3 in which we proposed to tackle the heterogenity in FDI destinations by disaggregating our sample in different country-groups, we decide to present our results in two subsections. In Subsection 4.4.1, we report the results for the whole dataset. Then, in Subsection 4.4.2, we report the results for each of the recipient-country groups separately.

Our findings show that, consistent with previous literature (Faeth (2009)), the determinants of German outward FDI are not driven by one specific FDI theory. On the contrary, a set of variables associated with different theoretical approaches are found to be robust FDI determinants. The variables that are found most frequently statistically significant across our different subsamples are: Custom Union (*CU*), the number of fixed telephone subscriptions (*ln_h_lines*), the voice and accountability index (*h_va*) and Preferential Trade Agreement (*PTA*).

4.4.1 Worldwide German FDI determinants

Table 4.2 reports the estimated coefficients for all recipient countries in our data set. Only 11 variables remained as robust FDI determinants out of the 24 singled out by the BMA analysis in Chapter 3. In what follows, we briefly present the estimated coefficients obtained while a more detailed interpretation of the variables will be provided in the country-groups subsections.

Concerning GDP and Population Measures, we find a parameter estimate of 1.025 for GDP per capita (*ln_h_gdppc*) consistent with gravity model predictions and HFDI motivations of German multinational firms.

Distance (*ln_distcap*) presents a negative coefficient of -0.394, in line with previous empirical studies (see, for instance, Bénassy-Quéré et al. (2005) or Basile et al. (2008)). This concerns the fact that the longer the distance, the higher the information costs resulting in a reduction in FDI as predicted by the gravity model. Another explanation might be that firms engaged in VFDI would prefer closer locations due to the intra-firm trade involved in the fragmentation of production.

In relation to factor endowment variables, our results show statistical significance for those variables capturing the quality of labor. Particularly, the parameter estimate of skilled labour endowment (*h_skilled*) of -4.253 presents the opposite sign than predicted by the literature. To some extent, however, this variable could be interpreted as a proxy for wages and hence, the negative parameter would suggest VFDI motivations.

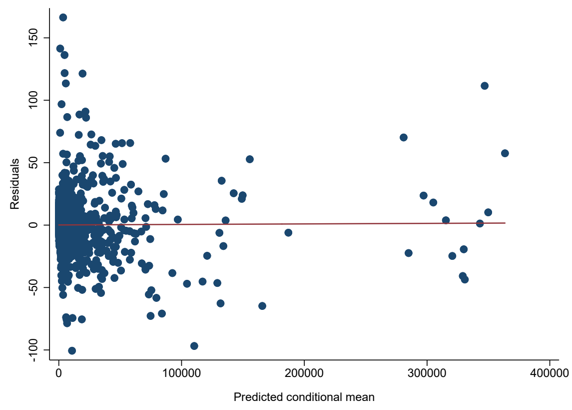
Differences in education levels (*ln_sq_educ_diff*) between Germany and host countries presents an estimated parameter of 1.569. This finding is in line with the knowledge-capital model developed by Carr et al. (2001) and VFDI, implying that differences in education are an attractive factor for German MNEs seeking to minimize costs. Similarly, we obtain a coefficient for squared skill differences (*sq_skill_diff*) of -12.949, in line with the results drawn by Markusen and Maskus (2001) and Blonigen et al. (2003). Concerning trade agreements, we found that the presence of a Preferential Trade Agreement (*PTA*) decreases German outward FDI by 22.04%; whereas a Custom Union (*CU*) increases FDI by 96.99%. As regards trade openness variables, the sign and magnitude of the coefficients are aligned with the literature. The interaction of squared education differences with host trade openness (*ln_educ_open_interact*) presents a parameter estimate of -1.597. The coefficient for trade openness, in turn, is positive and with a parameter value of 1.299.

Mobile cellular subscriptions (*ln_h_cell*) are also a significant FDI determinant and present a positive estimated coefficient of 0.148. Robust evidence is also found for institutional variables in line with recent contributions in the literature (see, for instance Berden et al. (2012), among others). Concretely, the voice and accountability index (*h_va*) exerts and estimated parameter of 0.015.

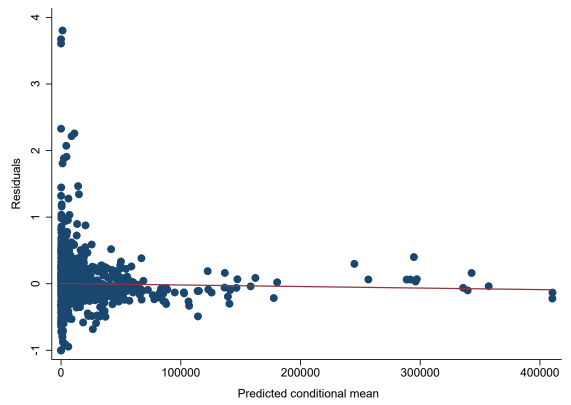
Finally, remember that we argued in Chapter 3 that aggregate estimations might omit information regarding the heterogenity of German FDI recipient countries and suffer for the so-called aggregation bias (Mitchell et al. (2011)). Accordingly, in the next subsections we take into account the heterogenity of the destination countries by disaggregating the analysis in different country-groups. This allows us to provide further insights into the key FDI determinants in each region.

Fig. 4.1 GLMs estimators for all host countries: Predictions and Pearson residuals.

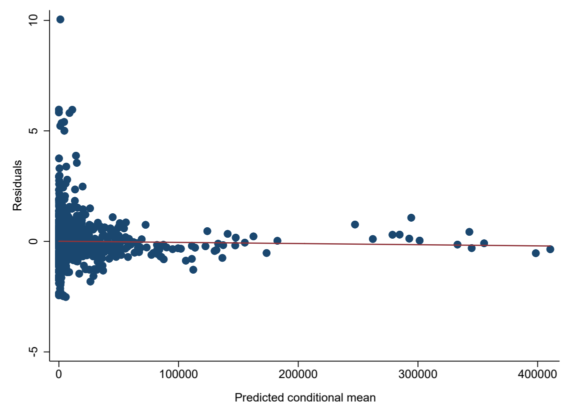
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

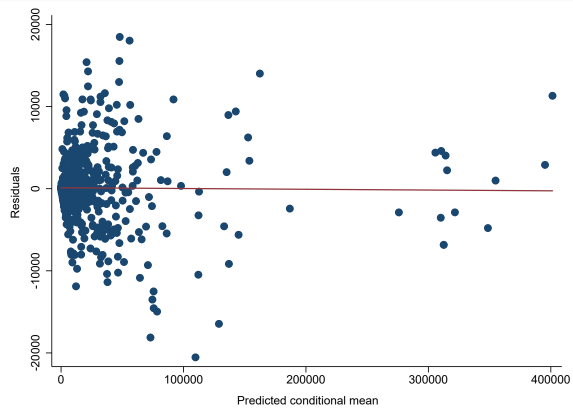
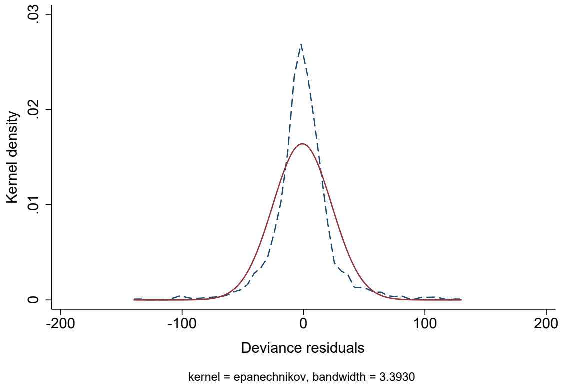
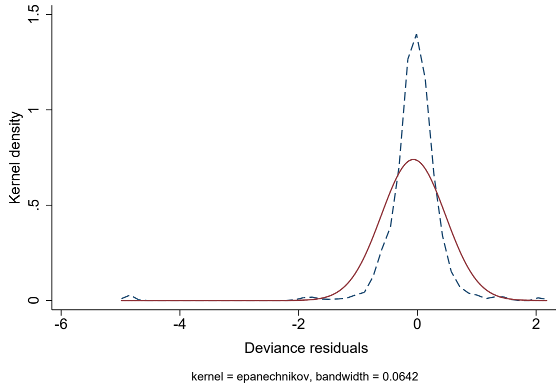


Fig. 4.2 GLMs estimators for all host countries: Density of deviance residuals.

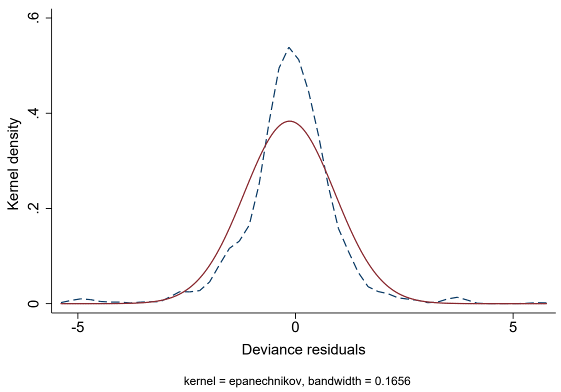
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

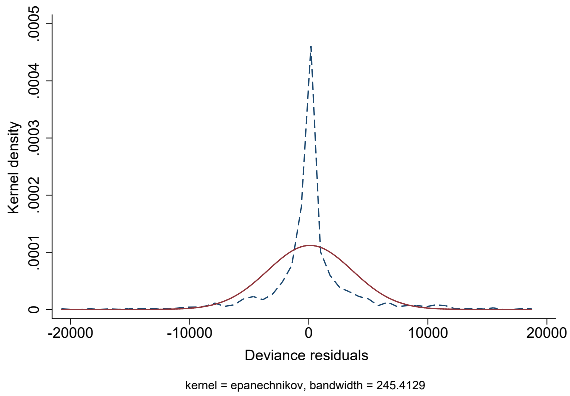


Table 4.2 Determinants of FDI in all recipient countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_sq_gdp_diff				-0.084***
				(0.03)
ln_sq_gdppc_diff				
ln_h_pop				
ln_h_gdppc	1.192***	1.082**	1.025**	2.308***
	(0.31)	(0.46)	(0.46)	(0.34)
ln_distcap	0.759***	-0.410***	-0.394***	-0.819***
	(0.14)	(0.15)	(0.15)	(0.15)
h_landlocked	1.325***			-1.538***
	(0.46)			(0.28)
h_skilled		-4.890**	-4.253**	
		(2.01)	(1.81)	
ln_h_area				0.283***
				(0.11)
ln_sq_educ_diff	1.325***	1.625***	1.569***	1.201***
	(0.26)	(0.53)	(0.53)	(0.27)
sq_skill_diff	-6.639**	-14.374**	-12.949**	
	(2.63)	(6.80)	(6.19)	
h_productivity	1.701**			2.358***
	(0.86)			(0.76)
comlang_ethno	2.338***			-1.948***
	(0.18)			(0.40)
PTA		-0.251*	-0.249*	
		(0.14)	(0.14)	
CU	0.264*	0.676***	0.678***	0.446*
	(0.15)	(0.24)	(0.23)	(0.23)
FTA	-0.213**			
	(0.10)			

(Continued)

Table 4.2 Determinants of FDI in all recipient countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_h_ER				
ln_educ_open_interact	-1.363***	-1.653***	-1.597***	-1.240***
	(0.26)	(0.52)	(0.51)	(0.26)
h_open	0.837***	1.330**	1.299**	0.833***
	(0.20)	(0.63)	(0.63)	(0.09)
ln_h_KOFGI	3.345**			
	(1.53)			
ln_h_cell		0.137**	0.148**	-0.290***
		(0.07)	(0.07)	(0.10)
h_pr	0.131**			0.216**
	(0.06)			(0.09)
h_cl				
h_va	0.012**	0.014**	0.015**	0.018***
	(0.01)	(0.01)	(0.01)	(0.01)
h_rl				
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	986	986	986	986
$RESET$ test $p-values$	0.2931	0.1287	0.2685	0.0000
AIC	602.217	18.30744	17.16185	19.293
BIC	577396.6	-6328.724	-5532.727	1.25e+10
Deviance	584007.6515	289.1862419	1085.182638	1.25049e+10
Dispersion	608.9757	0.3012357	1.130399	1.33e+07
Bias	0.1384227	0.1414216	0.1418073	0.1664739
MSE	1.253814	1.174959	1.185416	1.372065
ErrorLoss	0.3675128	0.3516237	0.3494596	0.4241226

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

4.4.2 Region-specific German FDI determinants

German FDI in developed countries

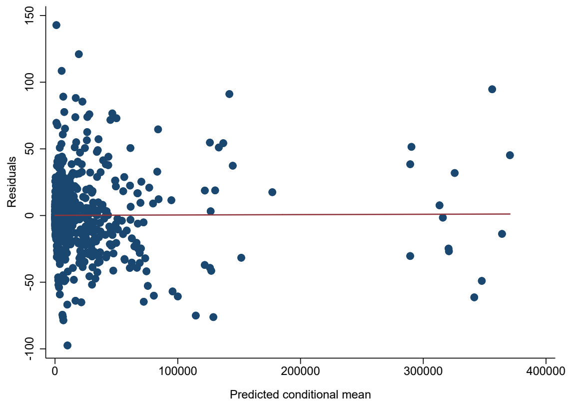
Table 4.3 reports the estimated coefficients for German outward FDI in developed countries. We identify 4 robust FDI determinants out of the 21 obtained in the BMA analysis. In particular, we have found that the variables associated to factor endowments, trade treaties and institutions are statistically significant. The estimated coefficient of skilled labour endowment (*h_skilled*) is negative with a parameter value of -2.647. This finding might be somewhat counterintuitive as developed countries, which hold the largest stock of German FDI, provide the required inputs that MNEs need for production, such as a highly qualified workforce. A plausible explanation might be, as in the case of the worldwide sample, that the variable is acting as a proxy for wages and thus, the negative relationship obtained is consistent with the increased participation in GVC and "outsourcing". Recently, Gunnella et al. (2019) analyzes the effects of the increasing participation of the euro area in GVCs and found that engagement in GVCs is associated with a rise in compensation per hour, in line with previous empirical literature that shows a rise in productivity of those firms that import inputs for production. Martínez-Galán and Fontoura (2019) also presents evidence of the role of country embeddedness to GVCs as an FDI determinant for a sample of 40 OECD countries.

Furthermore, Custom Union membership (*CU*) exerts a positive and statistically significant impact on German outward FDI. Concretely, this variable is capturing the relevance of the European Union as the largest regional location for German multinational firms. The coefficient estimate of 0.776 implies that FDI in EU member countries would be 117.276% larger than in non-members (i.e., $117.276 = [e^{0.776} - 1] * 100$).

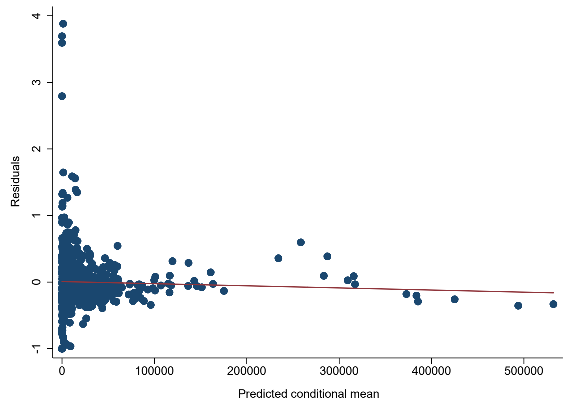
Finally, we find that the quality of institutions, such as government effectiveness (*h_ge*) and control of corruption (*h_cc*), has a positive impact on FDI, in accordance with recent empirical literature that highlights the role of economic structures in developed countries (see Antonakakis and Tondl (2015) and Dellis et al. (2017)). The sign and magnitude of the coefficients (0.022 and 0.019, respectively) are in line with the findings in the literature (Berden et al. (2012)).

Fig. 4.3 GLMs estimators for developed countries: Predictions and Pearson residuals.

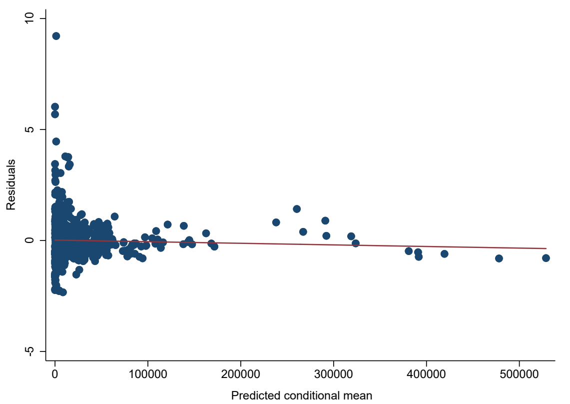
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

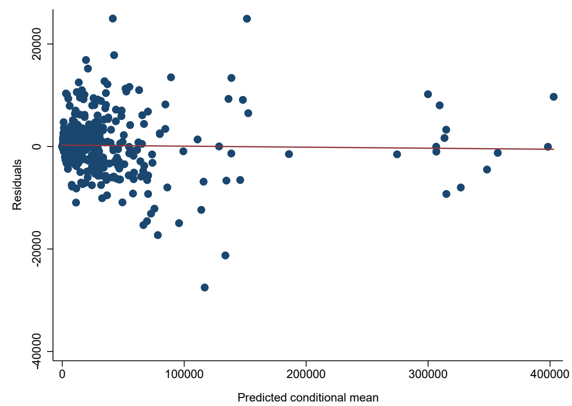
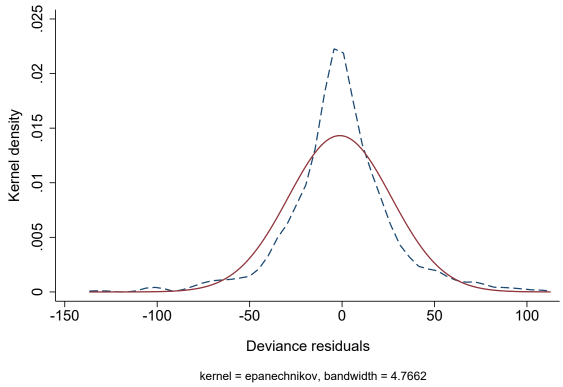
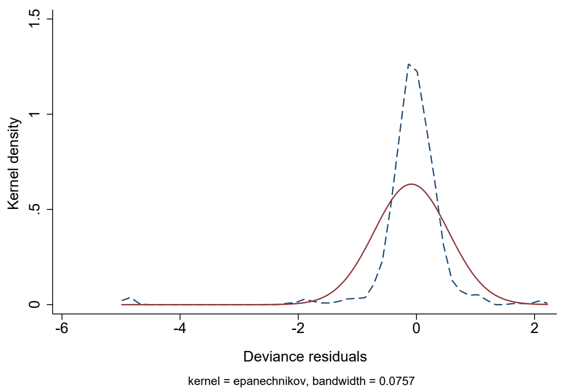


Fig. 4.4 GLMs estimators for developed countries: Density of deviance residuals.

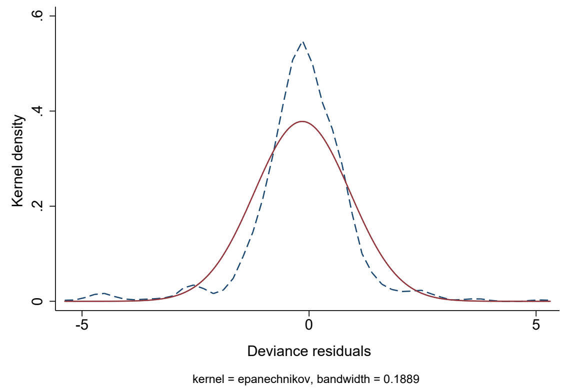
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

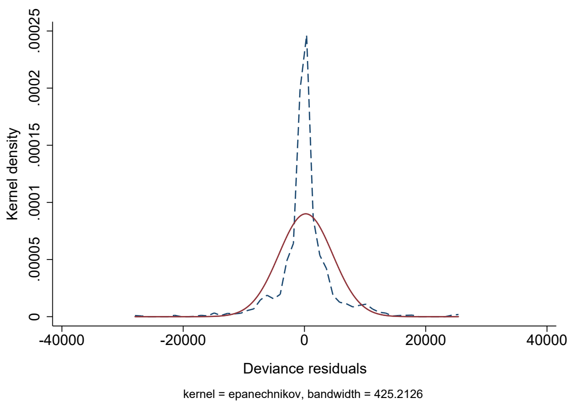


Table 4.3 Determinants of FDI in developed countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_sq_gdppc_diff				
ln_h_urban				
ln_h_pop				
h_skilled		-3.289**	-2.647*	
		(1.67)	(1.43)	
ln_h_area				
sq_skill_diff				
h_productivity	1.702*			2.280**
	(1.02)			(0.96)
comlang_off				
comlang_ethno				
PTA	-1.065***			-1.530***
	(0.34)			(0.24)
CU	1.345***	0.773***	0.776***	1.915***
	(0.34)	(0.22)	(0.22)	(0.14)
FTA	1.103*			1.867***
	(0.61)			(0.26)
EIA				0.160**
				(0.07)
ln_h_ER	-1.025***			-0.903***
	(0.25)			(0.11)
ln_exports				
h_open	0.446**			0.612***
	(0.18)			(0.12)

(Continued)

Table 4.3 Determinants of FDI in developed countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_h_KOFGI	3.619*			
	(2.10)			
h_va				0.012*
				(0.01)
h_pv	-0.004**			-0.004***
	(0.00)			(0.00)
h_ge	0.012**	0.023*	0.022**	0.012***
	(0.01)	(0.01)	(0.01)	(0.00)
h_cc		0.020*	0.019*	
		(0.01)	(0.01)	
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	646	646	646	646
RESET test $p-values$	0.1030	0.0272	0.0760	0.0001
AIC	786.9512	19.20668	18.24703	19.74661
BIC	497658.7	-3791.979	-3318.494	1.27e+10
Deviance	501683.5579	258.7413291	732.2262884	1.27015e+10
Dispersion	806.5652	0.4133248	1.169691	2.09e+07
Bias	0.2007483	0.1927667	0.1928793	0.1930551
MSE	1.857006	1.769485	1.783027	1.945276
ErrorLoss	0.4140162	0.4184945	0.4151283	0.4407476

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

German FDI in developing countries

For developing countries, in turn, the estimated coefficients are reported in Table 4.4. Out of the 22 variables singled out by the BMA analysis in Chapter 3, only 15 have remained robust FDI determinants. All the trade and investment treaties and trade openness variables are significant and have the expected sign with a few exceptions. Host country's openness to international trade has been dropped out by the BE procedure. We find that the presence of a Preferential Trade Agreement (*PTA*) results in a decrease of 31.81% of German outward FDI to developing countries, implying that when a PTA is in force MNEs would rather trade than engage in FDI. On the contrary, as in the case of developed countries, Custom Union membership (*CU*) positively impacts German FDI. Particularly, for this country-group, this variable may be capturing the Custom Union agreement between the European Union and Turkey in force since 1995. In this respect, the estimated parameter implies that the EU-Turkey custom union results in three times more German FDI (in Turkey) than with the rest of the countries in this group. Indeed, the EU and, in particular, Germany was one of the main investors in Turkey (Hadjit and Moxon-Browne (2005)).

We also found that a 1% increase in the 5-years lag of bilateral imports (*ln_imports*) is associated with a decrease of 0.311% of German FDI in developing countries, suggesting a substitutive relationship between FDI and imports. Compared to Economou and Hassapis (2015), the estimated coefficient is quantitatively smaller. Yet, they examine FDI into four European countries applying a dynamic panel data approach. Established trade relationships (*ln_trade*) between Germany and a developing economy result in an increase of FDI of 0.467%. This finding is consistent with the idea that vertical MNEs (VFDI) generate trade of intermediate goods within firms in different stages of the production chain, thereby acting as a complement of trade. Despite these findings might be seen as contradictory, they are consistent with previous empirical evidence. Concretely, Economou and Hassapis (2015) argues that the impact of exports and imports on FDI might differ and recommend to examine them separately. They found a positive impact for exports, while a negative effect for imports. The KOF Globalization Index (*ln_h_KOFGI*) is found to reduce German FDI by 3.142%. A plausible explanation might be that the more developing countries embrace globalization MNEs find easier to trade than undertake FDI. Another explanation might be that the KOF Globalization Index diminishes for some of the countries in our sample: Chile, Egypt, Saudi Arabia and Venezuela.

As regards interactions with host trade openness, the variables are statistically significant and of the expected sign. The estimated parameter of our *ln_educ_open_interact* variable (-1.181) is in line with Carr et al. (2001) who predict a coefficient of -1.264 for a Tobit specification using pool inward and outward U.S. FAS data. However, our variable is based on data on average educational attainment whereas they use data on annual surveys conducted by the International Labour Organization. The coefficient of *ln_educ_open_interact* (13.528), in turn, is also of the same sign as in Blonigen et al. (2003) but the magnitude is quantitatively larger.[6]

Concerning other vertically-oriented variables, skill differences (*sq_skill_diff*) and (*productivity*) are not robust FDI determinants after the BE procedure. Differences in education levels (*ln_sq_educ_diff*) is found to be statistically significant and its estimated coefficient implies an increase of FDI of 0.609%. The positive sign of the parameter is consistent with the knowledge-capital model of (Carr et al. (2001)) and VFDI, but the magnitude is smaller.

We also found statistical significance and robust evidence of all the variables related to GDP measures with the exception of squared GDP per capita differences (*ln_sq_gdppc_diff*). The similarity index (*sim_gdp*) exerts a negative coefficient of -1.873, suggesting VFDI motivations. As predicted by the gravity model, the estimated coefficient of host GDP

[6]These interaction variables where considered in the analysis as a measure of relative labour endowments following Blonigen and Piger (2014). Nevertheless, one must bear in mind that skill differences proxied by percent of employment by skilled labour extracted from ILOSTAT might not be an accurate measure, as concerns have been raised on comparability of classification schemes across countries (Blonigen et al. (2003))

(*ln_h_gdp*) is positive as expected, but the magnitude is larger than the findings in the literature, 1.552. Similarly, the coefficient of landlocked (*ln_h_landlocked*) exerts the correct sign but is larger than predicted by the literature, -2.028. These findings support HFDI motivations and hence, the fact that German MNEs seek to open up new sales markets in developing countries.

Unlike Chapter 3, we do not find statistical significance for access to natural resources (*h_oil*) and competitiveness (*ln_h_ER*). Nevertheless, telecommunication infrastructure (*ln_h_lines* and *ln_h_cell*) remain robust. The negative coefficient of the number of fixed telephone subscriptions (*ln_h_lines*) and positive of mobile cellular subscriptions (*ln_h_cell*) are plausible, supporting the idea that in recent times mobile cellular subscriptions have increased in detriment of the fixed ones.[7]

Finally, we have found robust evidence of institutional covariates. The estimated parameters for voice and accountability (*h_va*) and regulatory quality (*h_rq*) indeces are 0.019 and 0.013, respectively. These finding are similar to those obtained by Berden et al. (2012) although the size of the parameters is slightly larger.

[7]Bear in mind that fixed telephone subscriptions (*ln_h_lines*) as a measure of telecommunications infrastructure is somewhat misleading as it does not capture the reliability of the infrastructure as pointed out by Asiedu (2002) and thereby, results might be interpreted with caution.

Fig. 4.5 GLMs estimators for developing countries: Predictions and Pearson residuals.

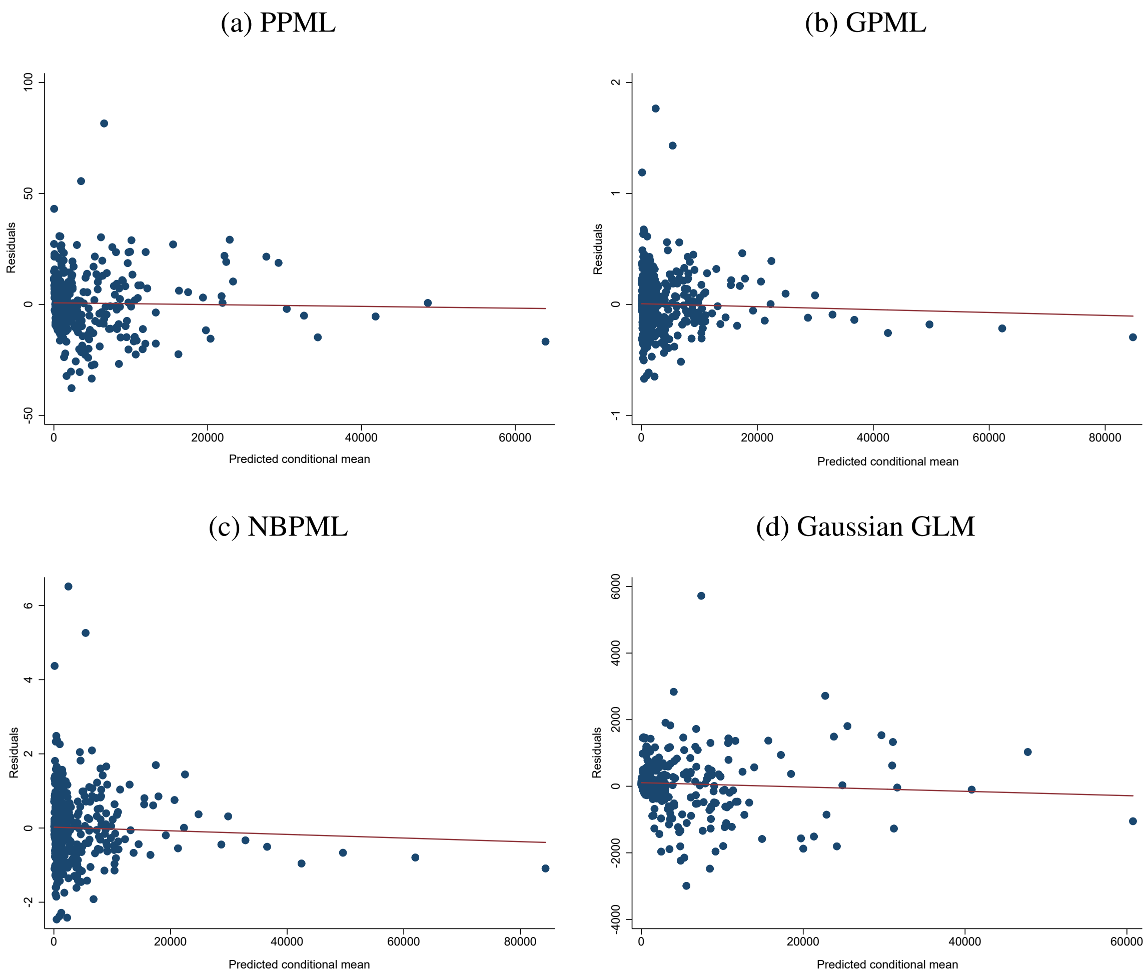
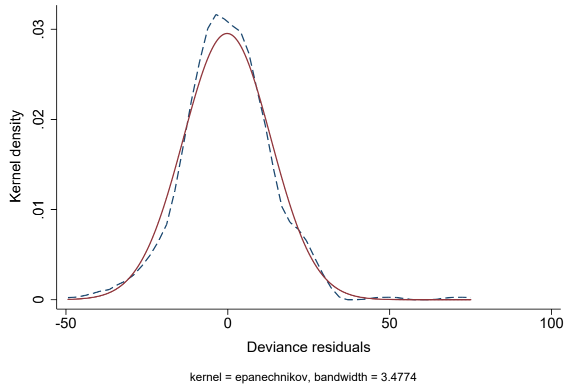
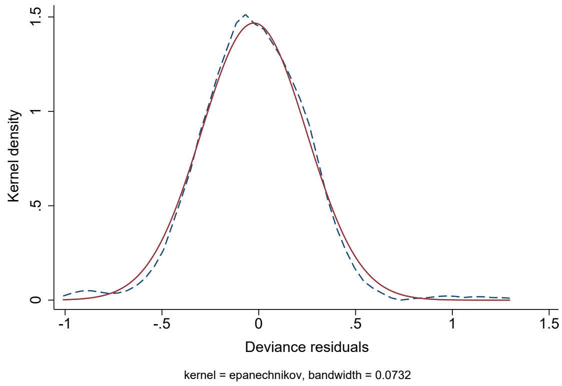


Fig. 4.6 GLMs estimators for developing countries: Density of deviance residuals.

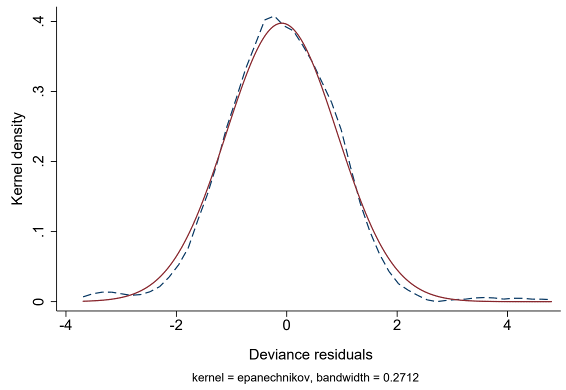
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

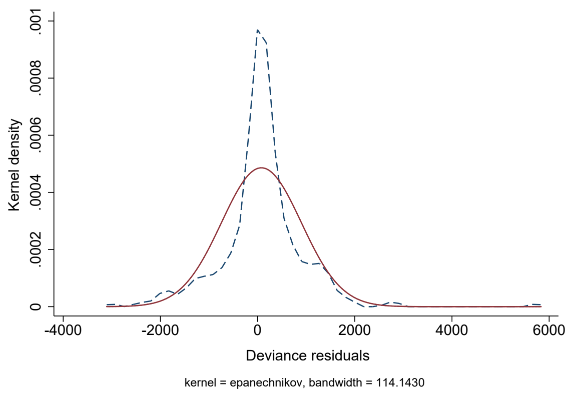


Table 4.4 Determinants of FDI in developing countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp		-1.896***	-1.873***	
		(0.48)	(0.49)	
ln_sq_gdppc_diff	-1.987**			-2.176***
	(0.83)			(0.61)
ln_h_gdp	1.220***	1.554***	1.552***	1.203***
	(0.28)	(0.28)	(0.28)	(0.20)
tdiff				
h_landlocked		-2.047***	-2.028***	
		(0.57)	(0.57)	
ln_sq_educ_diff	0.886***	0.606**	0.609**	2.217***
	(0.26)	(0.26)	(0.26)	(0.32)
sq_skill_diff				-14.802***
				(4.37)
h_oil				
h_productivity				
PTA	-0.157**	-0.388**	-0.383***	
	(0.07)	(0.10)	(0.10)	
CU		1.505***	1.486***	0.932***
		(0.39)	(0.39)	(0.19)
ln_h_ER				
ln_educ_open_interact	-0.942***	-1.175***	-1.181***	-2.003***
	(0.24)	(0.31)	(0.31)	(0.38)
skill_open_interact		13.499**	13.528**	29.448***
		(6.43)	(6.43)	(6.01)
ln_imports		-0.305***	-0.311***	
		(0.09)	(0.09)	
ln_trade		0.455***	0.467***	-0.197**
		(0.15)	(0.15)	(0.10)

(Continued)

Table 4.4 Determinants of FDI in developing countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
h_open				
ln_h_KOFGI		-3.141***	-3.142***	
		(1.03)	(1.03)	
ln_h_lines		-0.484***	-0.481***	
		(0.12)	(0.12)	
ln_h_cell	0.359***	0.420***	0.421***	0.336***
	(0.06)	(0.07)	(0.07)	(0.07)
h_va		0.019***	0.019***	
		(0.01)	(0.01)	
h_rq		0.013***	0.013***	
		(0.00)	(0.00)	
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	340	340	340	340
$RESET$ test $p-values$	0.2250	0.5385	0.5446	0.2165
AIC	191.2929	16.98826	15.16469	16.37623
BIC	59989.14	-1845.847	-1527.279	2.30e+08
$Deviance$	61860.23541	25.24448718	343.8129588	230448064.5
$Dispersion$	192.711	0.0786433	1.071068	717906.7
$Bias$	-0.0018126	0.0371242	0.0366329	-0.1573099
MSE	0.176307	0.0770302	0.0770313	0.329175
$ErrorLoss$	0.2845177	0.2114025	0.2113197	0.3677197

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

German FDI in Latin American countries

Table 4.5 reports the estimated parameters for German outward FDI in Latin American countries. For this country-group, the robust FDI determinants identified are those variables posited by the BMA analysis.

In this respect, most GDP measures appeared to be statistically significant. Particularly, the variable that accounts for the similarity index (*sim_gdp*) has a positive coefficient of 0.915, indicating horizontal multinational activity (HFDI). This finding is in line with those drawn by Baltagi et al. (2007) for US outward FDI stock who predicts an estimated coefficient around 1. The parameter obtained for squared GDP differences (*ln_sq_gdp_diff*) is -2.010, consistent with the knowledge-capital model and HFDI (Carr et al. (2001), Blonigen et al. (2003)). Thereby, this finding confirms that FDI should be larger among similar countries. Martínez-San Román et al. (2016) also get the same results although their estimated parameter ranges from -0.259 to -0.371.

As expected, we find a negative effect for host country population (*ln_h_pop*), in line with previous empirical literature (see Brenton and Di Mauro (1999) or Gutiérrez-Portilla et al. (2019), among others). This finding is consistent with the gravity model and the idea that an increase in population reduces the GDP per capita of a country and hence, deters FDI.

The estimated coefficient of time zone difference (*tdiff*) is 1.662. This finding is opposed to those drawn by Stein and Daude (2007) who predicts the reverse sign and a coefficient of -0.303 for a Tobit specification. A plausible explanation is that the variable is acting as a proxy for transportation costs, that is, for distance. If transport costs are high, a firm might rather produce in both countries (HFDI) than serve them through exports, thus the setup of horizontal MNEs is seen as substitutive of trade. Land area (*ln_h_area*) has a positive impact on FDI in line with the gravity model, but the magnitude of the estimated coefficient (4.553) is larger than in the literature. Camarero et al. (2018) reports a parameter than ranges from 0.087 to 0.724 for the gravity model of trade.

We obtain a coefficient of 0.485 for the interaction of skill differences with GDP differences (*ln_skill_gdp_interact*). This result is consistent with the results drawn by Markusen and Maskus (2001) and Blonigen et al. (2003), but the opposite of that found by Carr et al. (2001). Blonigen et al. (2003) explain that the negative sign predicted by Carr et al. (2001) is due to a misspecification of the skill differences variable due to ignoring the issue of whether the skill differences are positive or negative. Once corrected, they predicted a positive association between the interaction of skill differences with GDP differences and FDI. Education level (*ln_h_yr_sch*) presents a negative coefficient of -2.406, as opposed to what we should expect. The rationale behind this negative association might be that the variable is acting as a proxy for wages and thereby, it is providing further evidence of vertical MNEs.

Trade openness (*h_open*) presents a negative and significant coefficient, -2.402. This finding might suggest that the gradually increase in the degree of openness of Latin American economies to trade have encouraged MNEs to serve these markets through exports rather than engaging in HFDI.

As regards telecommunications infrastructure, our findings are similar to the ones reported above, as fixed subscriptions (*ln_h_lines*) reduce FDI by 0.711%, whereas internet users (*ln_h_internet*) increases FDI by 0.374%. Finally, concerning institutional variables, the coefficient of the political rights index (*h_pr*) is positive and statistically significant (0.127), suggesting that lower levels of democratic rights may be seen as an attractive factor for German MNEs in Latin American countries. Voice and accountability index (*h_va*) exerts a negative impact on FDI (-0.029) in line with Berden et al. (2012), whereas the political stability index (*h_pv*) positively impacts FDI (0.007) in line with the notion that MNEs prefer a stable host government.

Fig. 4.7 GLMs estimators for Latin American countries: Predictions and Pearson residuals.

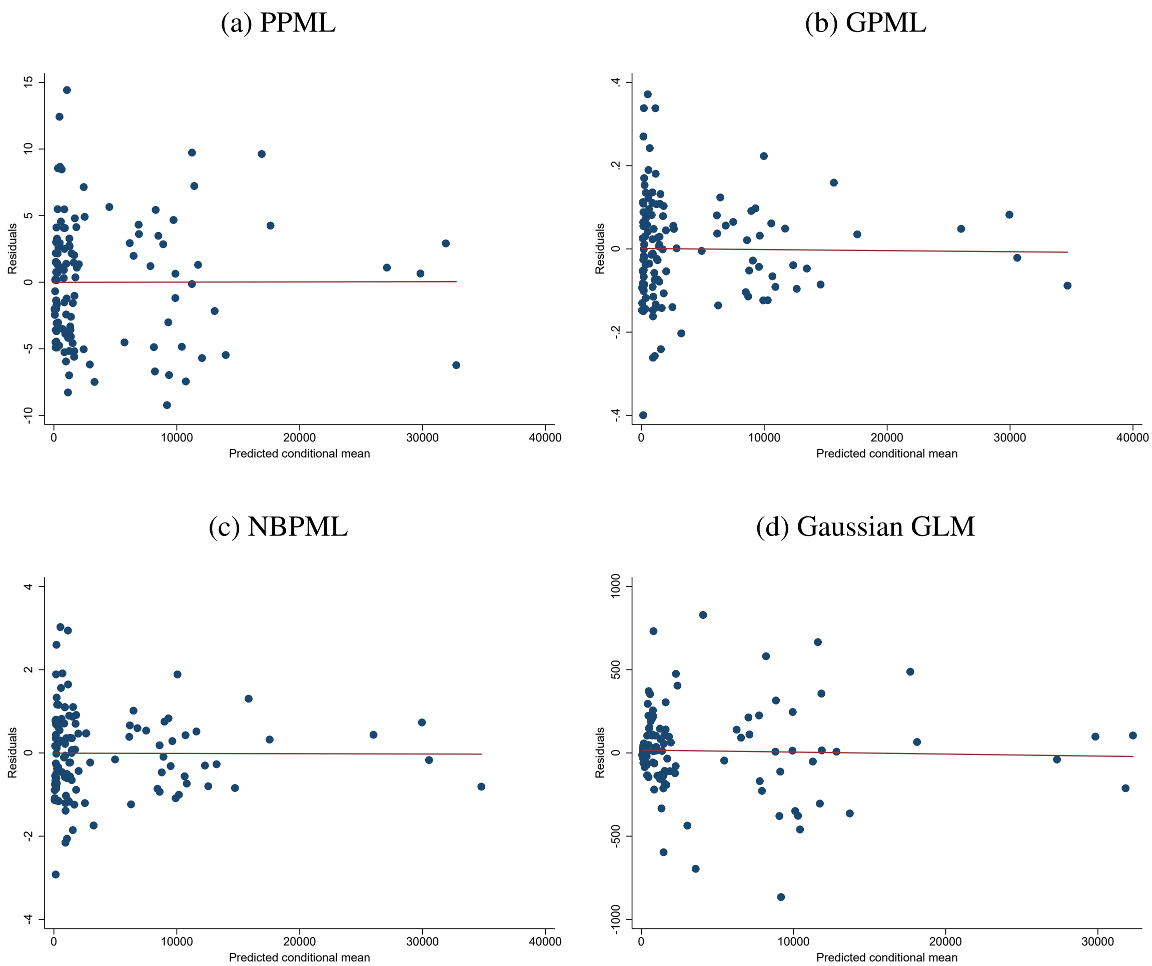
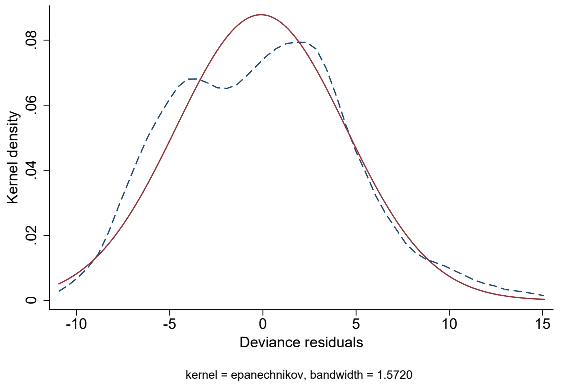
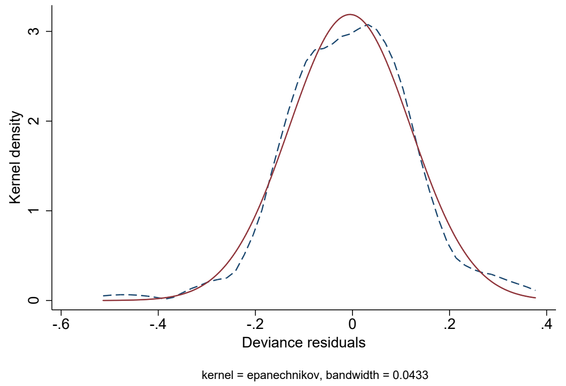


Fig. 4.8 GLMs estimators for Latin American countries: Density of deviance residuals.

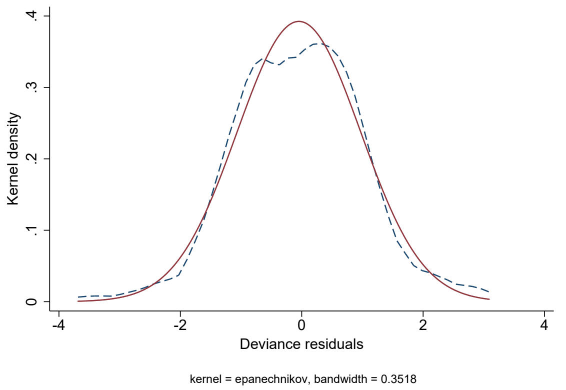
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

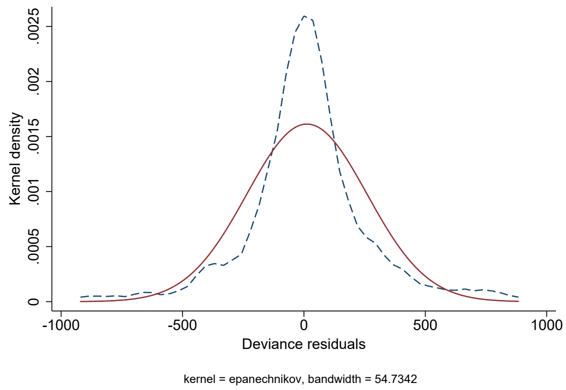


Table 4.5 Determinants of FDI in Latin American countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp	1.017***	0.970***	0.915***	2.241***
	(0.42)	(0.37)	(0.34)	(0.60)
ln_sq_gdp_diff	-2.277***	-2.076***	-2.010***	-2.501***
	(0.17)	(0.30)	(0.28)	(0.15)
ln_h_pop		-5.424***	-5.310***	
		(1.00)	(1.06)	
tdiff		1.685***	1.662***	
		(0.29)	(0.29)	
ln_h_area		4.582***	4.553***	-2.019***
		(1.33)	(1.33)	(0.54)
ln_skill_gdp_interact	0.445***	0.488***	0.485***	0.346***
	(0.10)	(0.05)	(0.04)	(0.13)
ln_h_yr_sch	-2.868***	-2.607***	-2.406***	-3.662***
	(0.42)	(0.83)	(0.77)	(0.37)
h_open	-2.413***	-2.394***	-2.402***	-2.220***
	(0.26)	(0.24)	(0.26)	(0.21)
ln_h_lines	-0.677***	-0.713***	-0.711***	-0.462***
	(0.16)	(0.06)	(0.07)	(0.13)
ln_h_internet	0.237***	0.390***	0.374***	0.321***
	(0.07)	(0.08)	(0.08)	(0.05)

(Continued)

Table 4.5 Determinants of FDI in Latin American countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
h_pr	0.159***	0.132***	0.127***	0.246***
	(0.03)	(0.02)	(0.02)	(0.03)
h_va	-0.015***	-0.030***	-0.029***	
	(0.00)	(0.00)	(0.00)	
h_pv	0.007***	0.007***	0.007***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	119	119	119	119
$RESET$ test $p-values$	0.6125	0.0040	0.0152	0.0000
AIC	29.54138	16.37211	13.00016	13.954
BIC	1896.413	-538.1903	-417.8305	7234199
Deviance	2436.454031	1.85062599	122.2104901	7234738.776
Dispersion	21.56154	0.0163772	1.081509	64024.24
Bias	0.0185131	0.0077757	0.010164	-0.0065372
MSE	0.0274735	0.0158241	0.01605	0.0415093
ErrorLoss	0.122962	0.0977254	0.0996833	0.1420331

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

German FDI in Asian countries

For Asian countries, in turn, the estimated coefficients are reported in Table 4.6. We identify 6 robust FDI determinats out of the 8 obtained in the BMA analysis. The variables that account for GDP per capita (*ln_h_gdppc*) and Competitiveness (*ln_h_ER*) have been dropped out by the BE procedure.

The estimated coefficient of the sum of the two countries' real GDPs (*ln_sum_gdp*) implies an increase in FDI of 1.757%, suggesting that German investors are strongly attracted to large markets in Asian economies. This finding is aligned with Martínez-San Román et al. (2016) who reports a parameter of around 1.5. Similarly, the coefficient estimate for landlocked (*ln_h_landlocked*) of -4.266 is aligned with market-seeking motivations (HFDI) of German MNEs in this region.

The education level (*ln_h_yr_sch*) positively impacts FDI, consistent with the idea that a highly educated workforce fosters productivity and hence, leads to an increase in profitability for MNEs activities. The magnitude of the estimated parameter (2.848) is larger than the findings in previous empirical studies (see, for instance, Basile et al. (2008)).

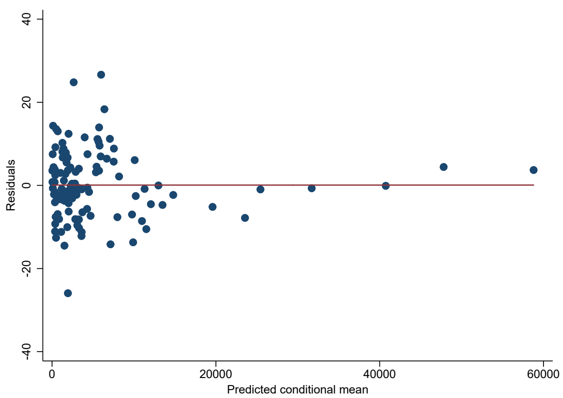
Concerning the coefficient for the number of internet users (*ln_h_internet*) is also positive in this case and implies that a one percent increase in the number of internet users is associated with a 0.080% increase of FDI.

Finally, many institutional variables are significant for this group of countries. The civil liberties index (*h_cl*) exerts a coefficient estimate of 0.277, consistent with the results of Adam and Filippaios (2007) who argues that the effect of civil liberties on FDI is non-linear and hence, a high degree of repression may deter FDI, while a low level of repression is attractive for MNEs when they seek to minimize costs. This might be the case of German FDI in China. Nevertheless, the magnitud of the coefficient is larger than the findings of Adam and Filippaios (2007). Last, the coefficient for the voice and accountability index

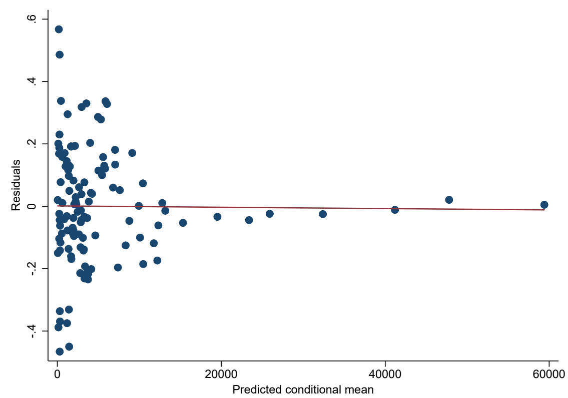
(h_va) is 0.027, consistent with Berden et al. (2012) and the idea that democratic institutions promote a good investment climate for MNEs.

Fig. 4.9 GLMs estimators for Asian countries: Predictions and Pearson residuals.

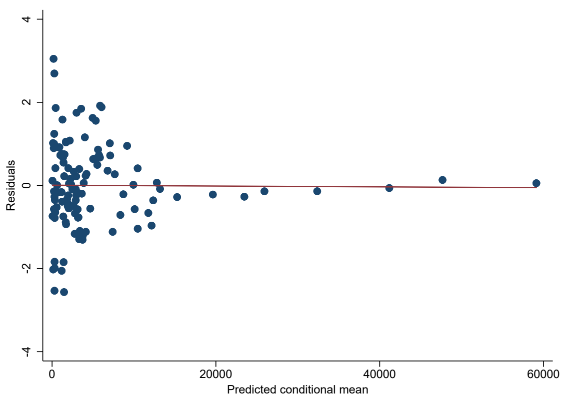
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

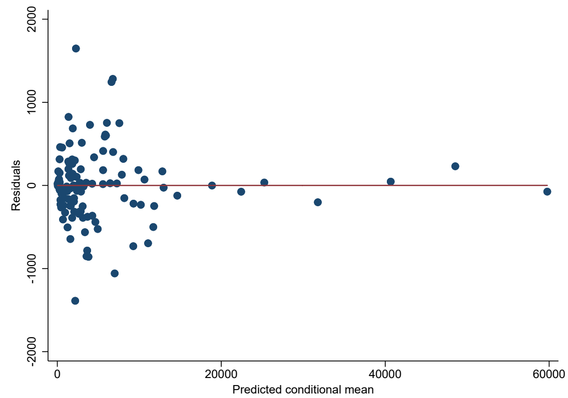


Fig. 4.10 GLMs estimators for Asian countries: Density of deviance residuals.

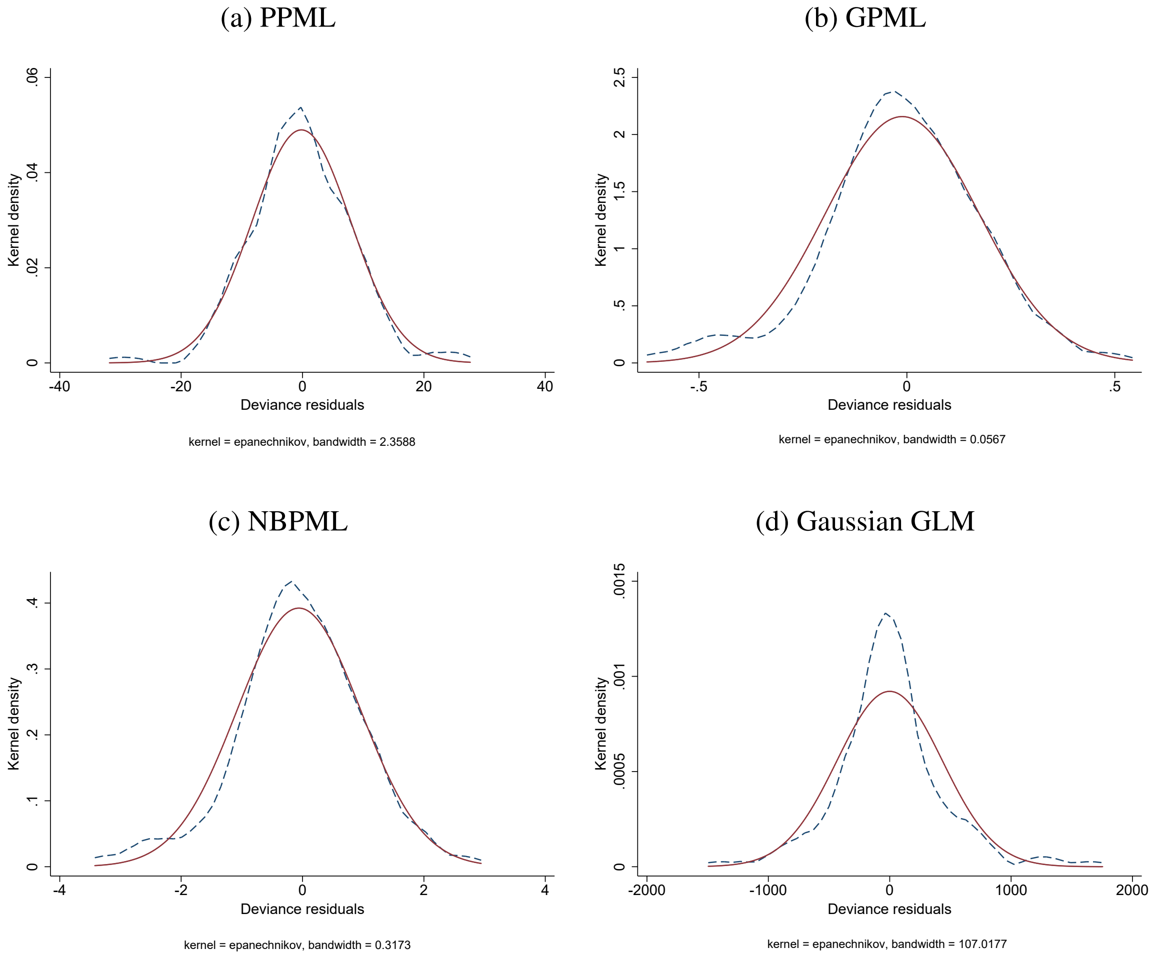


Table 4.6 Determinants of FDI in Asian countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_sum_gdp		1.763***	1.757***	
		(0.08)	(0.08)	
ln_h_gdppc	1.758***			1.928***
	(0.17)			(0.14)
h_landlocked	-8.091***	-4.247***	-4.266***	-8.024***
	(0.39)	(0.35)	(0.33)	(0.29)
ln_yr_sch_d	4.360***	2.820***	2.848***	6.618***
	(0.50)	(0.69)	(0.67)	(0.78)
ln_h_ER				-0.502***
				(0.10)
ln_h_internet		0.079***	0.080***	
		(0.03)	(0.03)	
h_cl		0.280**	0.277**	
		(0.11)	(0.12)	
h_va	0.017***	0.027***	0.027***	0.022***
	(0.00)	(0.01)	(0.01)	(0.01)
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	119	119	119	119
$RESET$ test $p-values$	0.1819	0.9395	0.9422	0.9166
AIC	75.4968	17.59368	14.99792	15.07193
BIC	7293.547	-535.9904	-417.5759	2.21e+07
$Deviance$	7833.587578	4.050531479	122.4650283	22127650.53
$Dispersion$	69.32378	0.0358454	1.083761	195819.9
$Bias$	0.0119449	0.0170187	0.0170236	0.0123652
MSE	0.0586087	0.0355448	0.0355855	0.0804531
$ErrorLoss$	0.1591156	0.1410872	0.1405214	0.1735322

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

German FDI in core EU countries

Table 4.7 reports the estimated coefficients for core EU countries. In this case, 6 out of the 9 variables identified by the BMA analysis in Chapter 3 are maintained as robust FDI determinants.

Starting with the gravity variables, the coefficient estimate for GDP per capita (*ln_h_gdppc*) is 3.214, impliying that a one percent increase in host countries' GDP per capita increases FDI by 3.214%. This finding is capturing the HFDI strategy of German FDI in core EU countries. Moreover, resulting from the GPML specification, we obtain a coefficient for distance (*ln_distcap*) of 2.836. This finding may point towards a substitutive relationship between FDI and trade and thus, HFDI (Helpman (2006)). Egger and Pfaffermayr (2004a) had similar results, as he found an estimate of 1.188 for distance on German outward FDI using a Hausman-Taylor SUR approach. Similarly, the positive coefficient of Landlocked (*ln_h_landlocked*), 0.466, is in line with the market-seeking motive of German MNEs in this region and consistent with a substitutive relationship between FDI and trade.

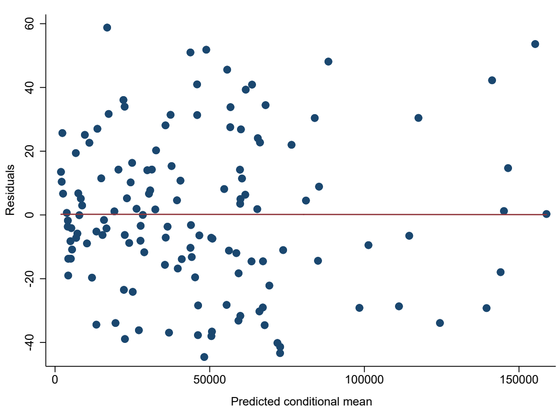
The coefficient for wages (*ln_h_wage*) of 1.111, is also compatible with a prevalence of HFDI and the idea that MNEs seek for qualified labour and perceive wages as a signal of higher qualification or productivity. Recent contributions for the euro area suggest that this would be the case of MNEs involved in GVCs (Gunnella et al. (2019)). The magnitude of the parameter is aligned with the findings of Mitze et al. (2010) for German FDI in the European Union (1.22), Basile et al. (2008) for non-European firms in EU countries (0.550) and Head et al. (1999) for Japanese FDI in the US (1.941).

The positive coefficient estimate of the 5-year lag of bilateral exports (*ln_h_exports*) implies a complementarity relationship between exports and FDI (an elasticity of 0.614%). Economou and Hassapis (2015) also found this positive relationship in the case of FDI inflows into European countries, although they report a larger coefficient (2.513).

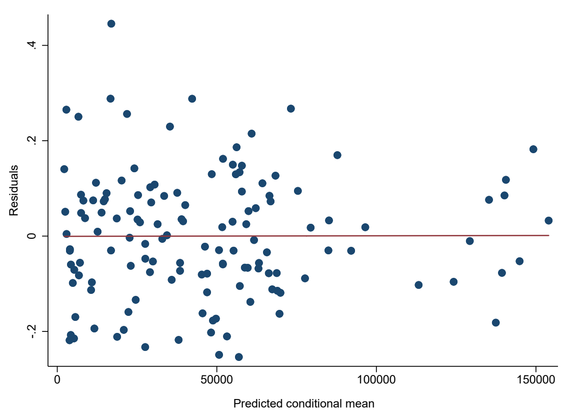
Finally, similarly to the previous country-groups we find a negative and statistically significant coefficient for the number of fixed telephone subscriptions (*ln_h_lines*), -1.045.

Fig. 4.11 GLMs estimators for Core EU countries: Predictions and Pearson residuals.

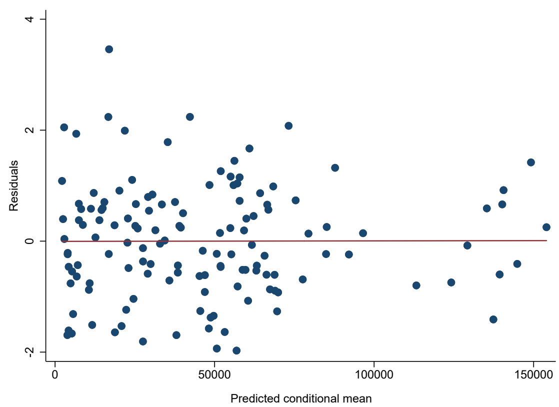
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

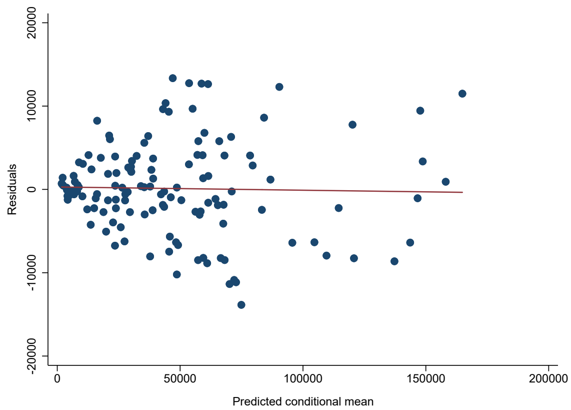
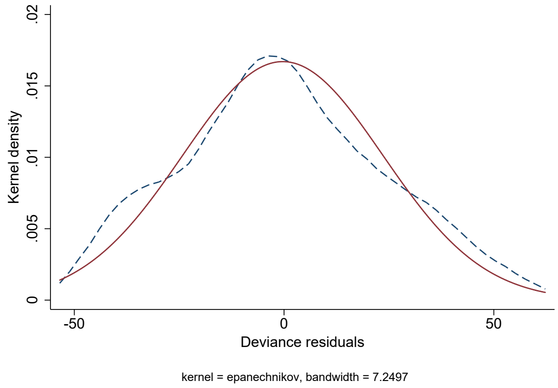
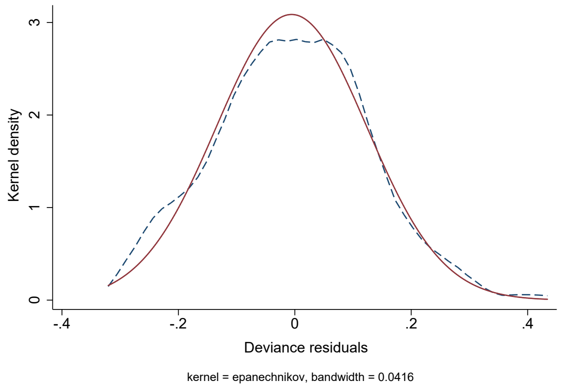


Fig. 4.12 GLMs estimators for Core EU countries: Density of deviance residuals.

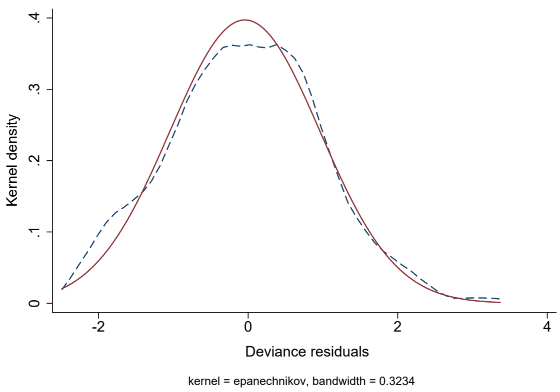
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

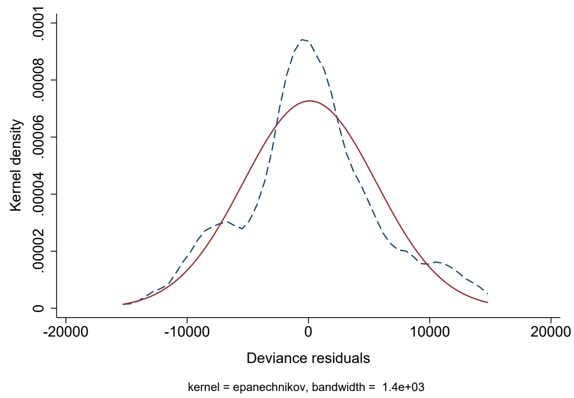


Table 4.7 Determinants of FDI in core EU countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp	1.264***			0.805***
	(0.22)			(0.27)
ln_h_gdppc		3.224***	3.214***	-3.099***
		(0.70)	(0.70)	(0.59)
ln_distcap		2.836***		
		(0.56)		
h_landlocked	-2.841***	0.466***	0.466***	
	(0.56)	(0.09)	(0.09)	
ln_h_area	-1.372***			-1.322***
	(0.22)			(0.24)
ln_h_wage	1.153**	1.114**	1.111**	1.082**
	(0.57)	(0.52)	(0.52)	(0.48)
ln_exports	0.845***	0.615**	0.614**	1.102***
	(0.31)	(0.31)	(0.31)	(0.30)
ln_h_internet	-0.217**			-0.258*
	(0.11)			(0.14)
ln_h_lines	-0.681***	-1.046***	-1.045***	-0.649*
	(0.22)	(0.10)	(0.10)	(0.34)
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	136	136	136	136
$RESET$ test $p-values$	0.1440	0.9685	0.9671	0.0102
AIC	578.8406	22.83443	19.56037	20.15496
BIC	76416.02	-631.4732	-497.2952	4.07e+09
Deviance	77049.75539	2.259265837	136.4372673	4069538014
Dispersion	597.2849	0.0175137	1.057653	3.15e+07
Bias	0.0040716	0.0083061	0.0082994	-0.0048906
MSE	0.0186355	0.016693	0.016694	0.0223414
ErrorLoss	0.1075208	0.1046733	0.1046702	0.1148918

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

German FDI in peripheral EU countries

For peripheral EU countries, in turn, the estimated coefficients are reported in Table 4.8. We identify 6 variables out of the 16 obtained in the BMA analysis. The variables eliminated in the BE procedure are: GDP measures (*sim_gdp*, *h_urban*), geography variables (*contig*, *ln_distcap*), productivity (*h_productivity*), trade openness variables (*ln_imports*, *ln_trade*), infrastructure (*ln_h_rail*, *ln_h_cell*) and institutions (*h_pv*).

The coefficient for the presence of a Preferential Trade Agreement (*PTA*) implies that a PTA in force have a large effect on German FDI (143.27%) compared with other countries in the group. This finding is in line with previous empirical literature Medvedev (2012) and the notion that a PTA implies the availability of a larger market due to a deeper integration between PTA members. In this particular case it refers to Croatia that before becoming full EU member signed a preferential agreement with the EU.

On the contrary, Custom Union membership (*CU*) has a negative coefficient, suggesting a reduction in FDI by 39.95%. A plausible explanation might be that FDI was horizontal for central and eastern european countries before they become members of the EU with the 2004 enlargment. Accordingly, regional integration might have had a negative impact on HFDI and lead to the dismantling of factories in those countries.

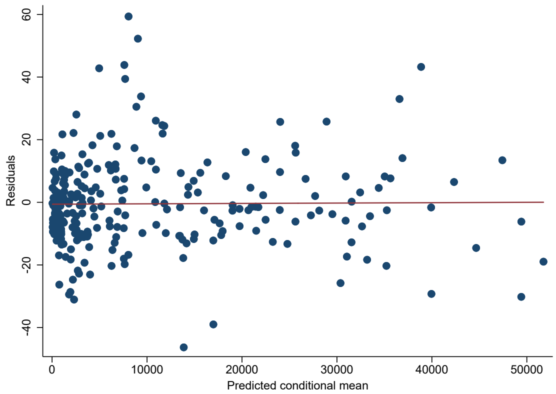
The coefficient for Bilateral Investment Treaties (*BIT*) of 1.158 is consistent with the idea that BITs reduce the risk to foreign investment. This finding is confirmed by previous empirical literature (Neumayer and Spess (2005); Egger and Merlo (2007); Egger and Pfaffermayr (2004b)).

From the GPML specification, we also obtain a parameter estimate of 0.646 for the real exchange rate (*ln_h_ER*), in line with VFDI and MNEs seeking for cost competitiveness in peripheral EU countries. Blonigen (1997) and Buch and Kleinert (2008) also found that an appreciation has a positive effect on outward FDI.

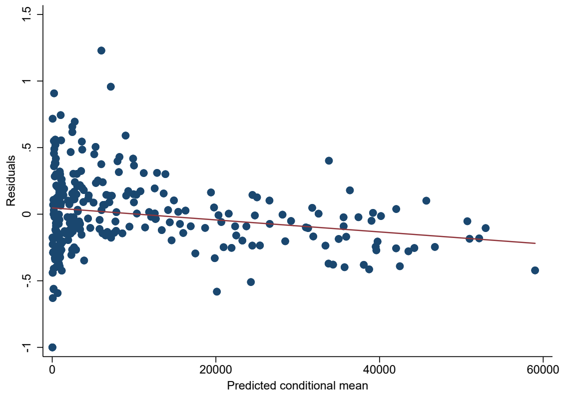
Finally, the coefficients for the 5-year lag of bilateral exports (*ln_exports*) and the number of fixed telephone subscriptions (*ln_h_lines*) exert the same sign and similar magnitude as those reported for the country-group of core EU countries.

Fig. 4.13 GLMs estimators for Peripheral EU countries: Predictions and Pearson residuals.

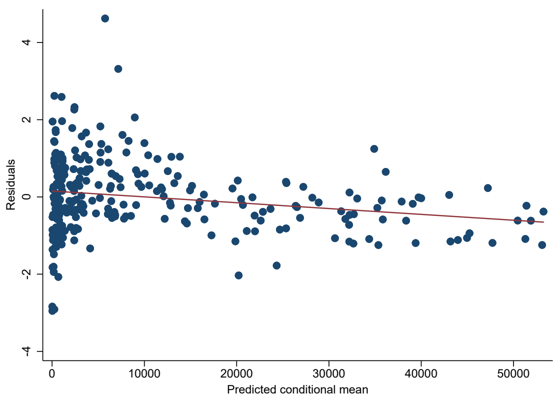
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

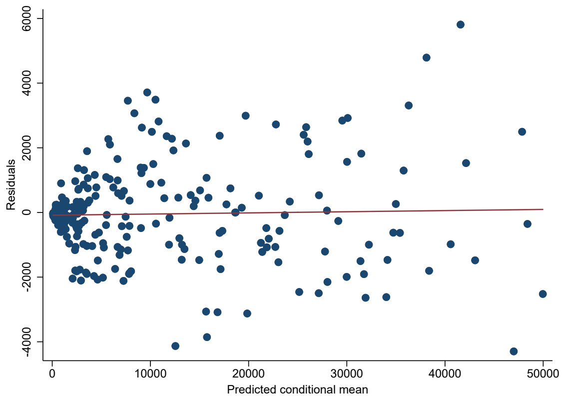
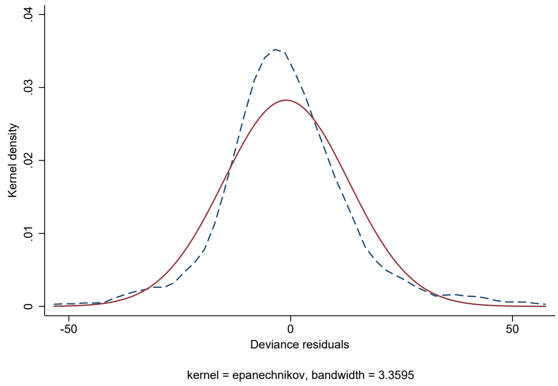
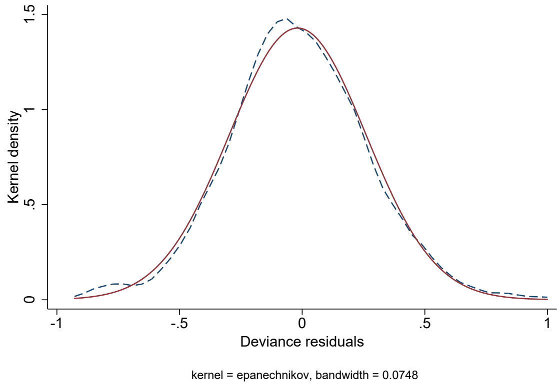


Fig. 4.14 GLMs estimators for Peripheral EU countries: Density of deviance residuals.

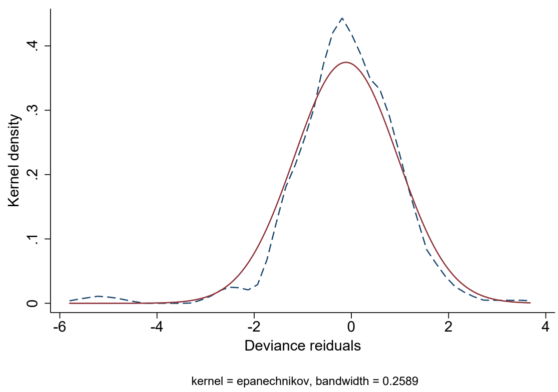
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

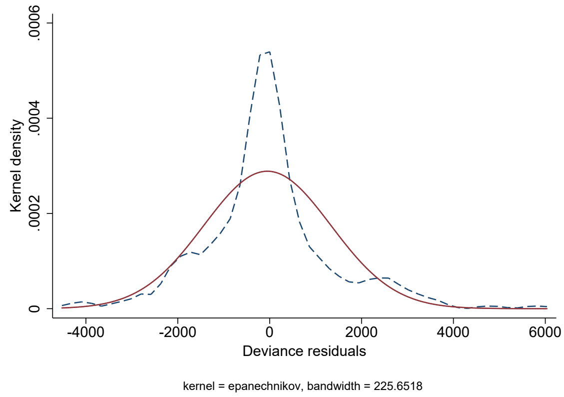


Table 4.8 Determinants of FDI in peripheral EU countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp				-1.121*
				(0.57)
h_urban				
contig				
ln_distcap	-2.306***			-3.626***
	(0.61)			(0.59)
h_productivity				-1.647***
				(0.40)
PTA	0.342***	0.938***	0.889***	0.310***
	(0.09)	(0.14)	(0.15)	(0.08)
CU		-0.587***	-0.510***	
		(0.13)	(0.15)	
BIT	1.088***	1.189***	1.158***	1.429***
	(0.17)	(0.16)	(0.16)	(0.17)
ln_h_ER	-0.560*	0.646*		-0.874***
	(0.31)	(0.38)		(0.22)
ln_exports	0.700***	0.839***	0.721***	0.717***
	(0.20)	(0.23)	(0.19)	(0.24)
ln_imports	-0.482**			
	(0.19)			
ln_trade				-0.456**
				(0.23)

(Continued)

Table 4.8 Determinants of FDI in peripheral EU countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_h_rail				
ln_h_lines	-0.836***	-1.078***	-0.959**	-0.507***
	(0.28)	(0.34)	(0.37)	(0.14)
ln_h_cell	0.306***			0.224***
	(0.12)			(0.07)
h_pv	-0.004*			
	(0.00)			
Host country FE (j)	*Yes*	*Yes*	*Yes*	
Year FE	*Yes*	*Yes*	*Yes*	
Observations	272	272	272	272
$RESET$ test $p-values$	0.0108	0.0015	0.0008	0.2598
AIC	209.6325	18.3589	16.76538	17.41574
BIC	52859.1	-1414.011	-1129.637	5.18e+08
$Deviance$	54299.79263	21.07406607	311.0537281	518277567.7
$Dispersion$	211.2832	0.0823206	1.210326	2024522
$Bias$	0.0808287	0.0316178	0.0361513	1.048619
MSE	0.1333289	0.0808789	0.0888418	4.191317
$ErrorLoss$	0.2127956	0.2206041	0.2201401	1.592108

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

4.5 Further robustness checks

A drawback of the estimations presented in the previous Section is that by including host country fixed effects, the coefficients of the explanatory variables with low or none time-variability (such as distance, population or land area among others) could not be estimated because of perfect collinearity with the fixed effects (Baltagi et al. (2014)). Thereby, in this Section we replicate the analysis without host country fixed effects as a robustness check. Furthermore, this allows us to compare the GLMs estimated parameters with those provided by the BMA analysis conducted in Chapter 3. Tables 4.B.1 to 4.B.7 provide the estimated coefficients for the alternative GLM estimators for the whole dataset and for the different country-groups. The goodness-of-fit statistics are also provided at the bottom of the tables. We also provide a visual inspection of the residuals for the different estimators. Pearson residuals for each GLM estimator are provided in Figures 4.B.1 to 4.B.7. Whereas Figures 4.B.8 to 4.B.14 report the density of deviance residuals for each estimator. Overall, the goodness-of-fit statistics and the inspection of the residuals confirm the results reported in Section 4.4: NBPML is the preferred estimator followed by GPML. Looking at the estimated coefficients, we found that although the magnitude is slightly different from the model averaged estimated coefficients in Chapter 3, the sign of the FDI determinants identified remains very stable across samples and thus, confirm the BMA results. Bear in mind that in Chapter 3 we estimated a log-linear regression for the BMA analysis whereas here the use of non-linear estimators might explain the difference in the magnitude of the coefficients.

4.6 Concluding remarks

The gravity model has become a popular tool to identify the determinants of the bilateral distribution of FDI. Even though the theoretical foundations of the FDI gravity model are nowadays well-stablished, there is no consensus concerning its empirical application. Researchers have applied a variety of estimators that estimate the gravity model either in its additive (i.e. log-log form) or multiplicative form, each of them with its own advantages and drawbacks. Since Silva and Tenreyro (2006), empirical studies have applied multiplicative functional form estimators and, in particular, the PPML estimator. However, the recent literature has argued that the PPML estimator is not always the best performing estimator and thus, alternative estimators have been suggested. This paper aims to shed some light to the debate by examinig the determinants of German outward FDI by comparing several methods estimating gravity models in their multiplicative form in a GLM framework: PPML, GPML, NBPML and Gaussian-GLM. Thereby, our contributions are twofold. First, we conduct a comprehensive analysis to identify the robust determinants of German outward FDI in order to contribute to the scarce existent literature on this topic. Second, we contribute to the debate on the empirical application of the FDI gravity model by comparing the performance of alternative GLM estimators.

In this paper, we follow a model selection approach based on several goodness-of-fit statistics and graphical techniques in order to asses the performance of the different GLMs. We argue that the estimation of the gravity model by PPML, as has been frequently done and recommended in the literature, does not appear to fit the data so well for our application. Furthermore, we show that the Gaussian-GLM estimator performs poorly in comparison with the alternative estimators considered. Our analysis suggest that NBPML is the estimator best matched to our data followed by GPML.

Moreover, the paper takes the comparison of the alternative GLM estimators to data and provides insightful evidence on the determinants of German outward FDI. The main results of the paper are the following. First, an appropriate estimation of the robust determinants previously identified in the BMA analysis conducted in Chapter 3 confirms a parsimonious FDI specification. Second, a disaggregation into country-groups allows us to disentangle the diversity of German FDI strategies across regions. In particular, we found that German investors in developed countries are strongly attracted by countries' embeddedness into GVCs which is positively linked with VFDI. Furthermore, results confirm the important role of the EU integration as an attractive factor for FDI together with well-developed democratic institutions. As regards developing countries, labour factor endowments attracts MNEs seeking cost-efficiency. Besides, large markets size and institutions are also important determinants in these countries. Within developing countries, results show that the main strategy of German FDI in Asian countries is market-driven, whereas efficiency-seeking (VFDI) appears to prevail in Latin American countries. Finally, our results show that German FDI in core EU countries is mainly market-driven (HFDI), whereas VFDI play a key role in peripheral EU countries.

Appendix A

Table 4.A.1 Robust FDI determinants identified by BMA

Variables	Country-groups					
	Developed	Developing	Latin American	Asian	Core EU	Peripheral EU
ln_sum_gdp				•		
sim_gdp		•	•			
ln_sq_gdp_diff			•			
ln_h_pop			•			
ln_h_gdp		•				
ln_h_gdppc					•	
ln_distcap					•	
tdiff			•			
h_landlocked		•	•	•		
h_skilled	•					
ln_h_area			•			
ln_sq_educ_diff		•				
ln_skill_gdp_interact			•			
ln_h_yr$_{s}ch$			•	•		
ln_h_wage					•	
PTA		•				•
CU	•	•				•
BIT						•

(Continued)

Table 4.A.1 Robust FDI determinants identified by BMA *(Continued)*

Variables	Country-groups					
	Developed	Developing	Latin American	Asian	Core EU	Peripheral EU
ln_h_ER						•
ln_educ_open_interact		•				
skill_open_interact		•				
ln_exports					•	•
ln_imports		•				
ln_trade		•				
h_open			•			
ln_h_KOFGI		•				
ln_h_internet			•	•		
ln_h_lines		•	•		•	•
ln_h_cell		•				
h_pr			•			
h_cl				•		
h_va		•	•			
h_pv			•			
h_ge	•					
h_rq		•		•		
h_cc	•					

Notes: The table lists all variables with inclusion probabilities (PIP) above 0.50 computed by BMA.

Appendix B

Table 4.B.1 Robustness checks: Determinants of FDI in all recipient countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_sq_gdp_diff	0.147***			0.157***
	(0.04)			(0.02)
ln_sq_gdppc_diff		0.100**	0.096**	
		(0.05)	(0.04)	
ln_h_pop	1.101***	1.134***	1.129***	1.274***
	(0.07)	(0.06)	(0.06)	(0.10)
ln_h_gdppc	0.649***	0.789***	0.783***	1.218***
	(0.17)	(0.13)	(0.13)	(0.19)
ln_distcap		-0.372***	-0.370***	-0.346***
		(0.09)	(0.09)	(0.10)
h_landlocked	0.776***	0.408*	0.416*	0.502***
	(0.20)	(0.22)	(0.22)	(0.17)
h_skilled		-6.124***	-6.079***	
		(2.09)	(2.08)	
ln_h_area				
ln_sq_educ_diff		1.534***	1.533***	
		(0.42)	(0.42)	
sq_skill_diff		-15.706***	-15.622***	
		(4.85)	(4.85)	
h_productivity	3.769***			4.096***
	(0.72)			(0.72)
comlang_ethno		0.749***	0.724***	0.294***
		(0.29)	(0.27)	(0.09)
PTA		-1.187**	-1.126**	
		(0.53)	(0.48)	
CU	0.963***	1.536***	1.476***	1.335***
	(0.17)	(0.51)	(0.46)	(0.13)
FTA		1.394**	1.332***	1.046***
		(0.56)	(0.51)	(0.23)

(Continued)

Table 4.B.1 Robustness checks: Determinants of FDI in all recipient countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_h_ER	-0.129***			-0.193***
	(0.04)			(0.05)
ln_educ_open_interact	-0.065***	-1.594***	-1.591***	-0.101***
	(0.02)	(0.42)	(0.42)	(0.01)
h_open	1.328***	2.743***	2.735***	1.491***
	(0.11)	(0.50)	(0.50)	(0.11)
ln_h_KOFGI		2.849***	2.820***	-4.894***
		(0.94)	(0.94)	(1.33)
ln_h_cell	0.183*			
	(0.11)			
h_pr		0.405***	0.405***	
		(0.11)	(0.11)	
h_cl		-0.193*	-0.194*	
		(0.11)	(0.11)	
h_va	0.027***	0.052***	0.052***	0.014***
	(0.01)	(0.01)	(0.01)	(0.00)
h_rl	-0.014***	-0.018***	-0.018***	
	(0.01)	(0.00)	(0.00)	
Host country FE (j)	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	986	986	986	986
$RESET$ test $p-values$	0.0005	0.0029	0.0048	0.0000
AIC	2849.192	18.62292	18.29608	20.42803
BIC	2792923	-5987.7	-5457.276	4.04e+10
$Deviance$	2799520.589	561.2735461	1091.697192	4.03529e+10
$Dispersion$	2925.309	0.5908143	1.149155	4.23e+07
$Bias$	0.2687628	0.2802414	0.2817137	-0.1068201
MSE	2.24651	1.70628	1.722817	2.448829
$ErrorLoss$	0.6742898	0.609595	0.6094354	0.8333995

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Table 4.B.2 Robustness checks: Determinants of FDI in developed countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_sq_gdppc_diff	0.052*	0.069**		0.012**
	(0.03)	(0.03)		(0.01)
ln_h_urban	-1.565**			
	(0.77)			
ln_h_pop	1.365***	1.149***	1.078***	1.193***
	(0.11)	(0.20)	(0.21)	(0.11)
h_skilled		-6.147***	-6.472***	
		(1.79)	(1.68)	
ln_h_area	-0.255***	-0.191**	-0.188*	-0.462***
	(0.09)	(0.10)	(0.10)	(0.05)
sq_skill_diff	-36.079***	-32.833***	-38.149***	
	(8.53)	(7.96)	(7.92)	
h_productivity				
comlang_off	2.844***	1.771***	2.044***	3.216***
	(0.55)	(0.36)	(0.34)	(0.17)
comlang_ethno	-2.280***	-1.731***	-2.119***	-2.631***
	(0.43)	(0.46)	(0.46)	(0.14)
PTA	-0.749***	-1.082***	-0.793**	-1.218***
	(0.25)	(0.37)	(0.33)	(0.13)
CU	0.414**	0.444*		
	(0.17)	(0.27)		
FTA		1.268***	0.958**	
		(0.47)	(0.38)	
EIA				
ln_h_ER		-0.133**	-0.128**	-0.599***
		(0.06)	(0.06)	(0.06)
ln_exports		0.350*	0.398**	0.213**
		(0.19)	(0.19)	(0.10)
h_open	0.707***	1.263***	1.290***	
	(0.21)	(0.34)	(0.33)	

(Continued)

Table 4.B.2 Robustness checks: Determinants of FDI in developed countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_h_KOFGI	5.170**	6.629***	6.843***	
	(2.38)	(2.03)	(2.33)	
h_va	0.052***	0.067***	0.069***	0.054***
	(0.02)	(0.01)	(0.01)	(0.01)
h_pv		0.014***	0.015***	
		(0.01)	(0.01)	
h_ge	0.022*	0.032**		0.010***
	(0.01)	(0.01)		(0.00)
h_cc	-0.028*	-0.062***	-0.047***	
	(0.02)	(0.01)	(0.01)	
Host country FE (j)	*No*	*No*	*No*	*No*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	646	646	646	646
RESET test $p-values$	0.0004	0.0568	0.0947	0.0000
AIC	3989.985	19.5456	19.31716	21.10034
BIC	2566846	-3522.188	-3252.141	5.09e+10
Deviance	2570831.641	437.9410947	727.4003632	5.08816e+10
Dispersion	4173.428	0.71559	1.182765	8.21e+07
Bias	0.329332	0.3327661	0.344227	0.0256873
MSE	2.692575	2.274213	2.443341	3.482975
ErrorLoss	0.6878338	0.6718497	0.6809272	1.030777

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Table 4.B.3 Robustness checks: Determinants of FDI in developing countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp	-0.576***	-0.641**	-0.649**	-0.387*
	(0.17)	(0.29)	(0.29)	(0.20)
ln_sq_gdppc_diff	0.944***	1.017***	1.015***	1.325***
	(0.26)	(0.15)	(0.15)	(0.34)
ln_h_gdp	1.008***	1.100***	1.104***	1.233***
	(0.18)	(0.21)	(0.21)	(0.18)
tdiff	0.130***	0.166***	0.166***	
	(0.03)	(0.03)	(0.03)	
h_landlocked	-1.540***	-1.850***	-1.859***	-1.265***
	(0.32)	(0.29)	(0.28)	(0.26)
ln_sq_educ_diff	0.943***	0.761***	0.768***	1.149***
	(0.28)	(0.24)	(0.24)	(0.35)
sq_skill_diff				
h_oil	-2.727***	-3.161***	-3.160***	
	(0.74)	(0.48)	(0.48)	
h_productivity		-1.337**	-1.348**	
		(0.58)	(0.58)	
PTA	-0.263**	-0.279**	-0.279**	
	(0.11)	(0.11)	(0.11)	
CU	1.188***	1.526***	1.528***	0.611**
	(0.29)	(0.28)	(0.28)	(0.26)
ln_h_ER	-0.147***	-0.137***	-0.137***	-0.072**
	(0.03)	(0.02)	(0.02)	(0.03)
ln_educ_open_interact	-1.516***	-1.409***	-1.416***	-1.613***
	(0.26)	(0.28)	(0.28)	(0.34)
skill_open_interact		-4.784**	-4.818**	
		(2.13)	(2.11)	
ln_imports	-0.601***	-0.440***	-0.441***	-0.665***
	(0.09)	(0.14)	(0.14)	(0.16)

(Continued)

Table 4.B.3 Robustness checks: Determinants of FDI in developing countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_trade	0.710***	0.514***	0.516***	0.683***
	(0.18)	(0.15)	(0.15)	(0.19)
h_open	1.522***	1.738***	1.744***	1.917***
	(0.33)	(0.31)	(0.31)	(0.53)
ln_h_KOFGI	-1.826*	-3.098***	-3.098***	
	(0.95)	(0.80)	(0.80)	
ln_h_lines	-0.390***	-0.486***	-0.488***	-0.162*
	(0.08)	(0.12)	(0.12)	(0.09)
ln_h_cell	0.440***	0.417***	0.421***	0.479***
	(0.07)	(0.07)	(0.07)	(0.07)
h_va				
h_rq	0.014***	0.012***	0.012***	0.014***
	(0.00)	(0.00)	(0.00)	(0.00)
Host country FE (j)	*No*	*No*	*No*	*No*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	340	340	340	340
RESET test $p-values$	0.3803	0.0099	0.0118	0.4220
AIC	260.2651	17.02308	15.55123	17.23141
BIC	83439.69	-1834.006	-1524.192	5.39e+08
Deviance	85310.77774	37.08508851	346.89952	538787443.7
Dispersion	265.7657	0.1155299	1.080684	1683711
Bias	0.050158	0.0545369	0.0536475	0.0264666
MSE	0.1448175	0.1155629	0.1154063	0.2785388
ErrorLoss	0.2880479	0.2538302	0.2534671	0.3889425

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Table 4.B.4 Robustness checks: Determinants of FDI in Latin American countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp	1.087***	1.007***	1.017***	1.510***
	(0.11)	(0.07)	(0.07)	(0.07)
ln_sq_gdp_diff	-2.265***	-2.853***	-2.803***	-2.163***
	(0.13)	(0.30)	(0.31)	(0.21)
ln_h_pop	-4.733***	-6.247***	-6.114***	-4.342***
	(0.51)	(0.74)	(0.78)	(0.29)
tdiff	1.585***	2.048***	2.006***	1.438***
	(0.16)	(0.22)	(0.23)	(0.08)
ln_h_area	3.825***	5.069***	4.956***	3.100***
	(0.55)	(0.62)	(0.66)	(0.16)
ln_skill_gdp_interact	0.466***	0.606***	0.597***	0.371***
	(0.06)	(0.09)	(0.09)	(0.07)
ln_h_yr_sch	-3.072***	-3.646***	-3.551***	-3.369***
	(0.21)	(0.34)	(0.34)	(0.53)
h_open	-2.086***	-2.683***	-2.666***	-2.357***
	(0.22)	(0.35)	(0.36)	(0.20)
ln_h_lines	-0.536***	-0.564***	-0.560***	-0.355***
	(0.10)	(0.10)	(0.10)	(0.06)
ln_h_internet	0.378***	0.453***	0.436***	0.392***
	(0.04)	(0.06)	(0.06)	(0.04)

(Continued)

Table 4.B.4 Robustness checks: Determinants of FDI in Latin American countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
h_pr	0.180***	0.166***	0.163***	0.225***
	(0.03)	(0.03)	(0.03)	(0.02)
h_va	-0.016***	-0.032***	-0.031***	
	(0.01)	(0.01)	(0.01)	
h_pv	0.011***	0.013***	0.013***	0.007***
	(0.00)	(0.00)	(0.00)	(0.00)
Host country FE (j)	*No*	*No*	*No*	*No*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	119	119	119	119
$RESET$ test $p-values$	0.6053	0.0000	0.0000	0.0000
AIC	33.38816	16.3758	13.23627	14.11082
BIC	2354.179	-537.7506	-419.7523	8462575
Deviance	2894.2202	2.290398285	120.2887007	8463114.665
Dispersion	25.61257	0.020269	1.064502	74894.82
Bias	0.0044703	0.0096235	0.0109141	-0.054751
MSE	0.0287948	0.019474	0.0196284	0.0589891
ErrorLoss	0.1244174	0.1111525	0.1114212	0.1666536

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Table 4.B.5 Robustness checks: Determinants of FDI in Asian countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_sum_gdp	3.269***	3.146***	3.154***	3.442***
	(0.05)	(0.11)	(0.10)	(0.12)
ln_h_gdppc	-1.198***	-0.892***	-0.917***	-1.224***
	(0.12)	(0.19)	(0.19)	(0.16)
h_landlocked	-2.487***	-2.442***	-2.444***	-2.298***
	(0.10)	(0.18)	(0.17)	(0.17)
ln_yr_sch_d	4.447***	3.473***	3.553***	4.387***
	(0.34)	(0.65)	(0.63)	(0.54)
ln_h_ER	-0.125***	-0.104***	-0.105***	-0.154***
	(0.02)	(0.02)	(0.02)	(0.02)
ln_h_internet	0.163***	0.148***	0.149***	0.175**
	(0.05)	(0.05)	(0.05)	(0.07)
h_cl	0.261**	0.343***	0.336***	
	(0.12)	(0.11)	(0.11)	
h_va	0.030***	0.031***	0.031***	0.018***
	(0.01)	(0.01)	(0.01)	(0.00)
Host country FE (j)	*No*	*No*	*No*	*No*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	119	119	119	119
RESET test $p-values$	0.6646	0.0085	0.0123	0.1185
AIC	79.43134	17.5981	15.11886	15.51694
BIC	7761.756	-535.4646	-417.4107	3.45e+07
Deviance	8301.797165	4.576329076	122.6302135	34530207.62
Dispersion	73.46723	0.0404985	1.085223	305577.1
Bias	0.0239674	0.0192282	0.0202486	0.0268637
MSE	0.048546	0.0398685	0.0401288	0.066558
ErrorLoss	0.1498828	0.1521747	0.151418	0.1702422

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Table 4.B.6 Robustness checks: Determinants of FDI in core EU countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp	1.551***	1.446***	1.445***	1.431***
	(0.05)	(0.08)	(0.08)	(0.11)
ln_h_gdppc	1.097***	1.217***	1.215***	1.147***
	(0.20)	(0.25)	(0.25)	(0.18)
ln_distcap	2.910***	2.862***	2.862***	3.420***
	(0.08)	(0.05)	(0.05)	(0.19)
h_landlocked	0.924***	0.852***	0.852***	0.870***
	(0.04)	(0.05)	(0.05)	(0.06)
ln_h_area	-0.634	-0.463***	-0.463***	-0.436***
	(0.02)	(0.06)	(0.06)	(0.09)
ln_h_wage		1.463**	1.461**	2.183**
		(0.60)	(0.59)	(1.04)
ln_exports				
ln_h_internet				0.098*
				(0.05)
ln_h_lines	-0.774***	-0.883***	-0.883***	-1.316***
	(0.29)	(0.25)	(0.25)	(0.31)
Host country FE (j)	*No*	*No*	*No*	*No*
Year FE	*Yes*	*Yes*	*Yes*	*Yes*
Observations	136	136	136	136
RESET test $p-values$	0.0126	0.2278	0.2294	0.0000
AIC	781.6989	22.83718	19.71349	20.43026
BIC	104004.7	-631.0991	-497.1577	5.36e+09
Deviance	104638.4807	2.633394886	136.5748167	5359297344
Dispersion	811.151	0.0204139	1.05872	4.15e+07
Bias	0.0012256	0.0096816	0.0096505	-0.0690539
MSE	0.023577	0.0196526	0.0196535	0.0489789
ErrorLoss	0.1218555	0.1128783	0.1128674	0.1642884

Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Table 4.B.7 Robustness checks: Determinants of FDI in peripheral EU countries, 1996-2012.

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
sim_gdp	0.429***			0.408***
	(0.07)			(0.07)
h_urban	-1.876***	-3.479***	-3.466***	-1.290***
	(0.22)	(0.88)	(0.85)	(0.16)
contig	1.972***	1.831***	1.859***	1.855***
	(0.20)	(0.25)	(0.23)	(0.21)
ln_distcap	1.583***	1.488***	1.500***	1.507***
	(0.24)	(0.15)	(0.15)	(0.27)
h_productivity	-1.693***	-3.273***	-3.208***	-1.661**
	(0.60)	(0.75)	(0.70)	(0.65)
PTA	0.312***	0.781***	0.764***	0.331***
	(0.07)	(0.17)	(0.16)	(0.06)
CU		-0.327*	-0.312*	
		(0.17)	(0.16)	
BIT	-1.676***	-1.422***	-1.442***	-1.629***
	(0.13)	(0.19)	(0.18)	(0.13)
ln_h_ER	0.437***	0.447***	0.450***	0.400***
	(0.02)	(0.05)	(0.05)	(0.03)
ln_exports				
ln_imports		-0.378**	-0.385**	
		(0.19)	(0.19)	
ln_trade		0.988***	0.983***	
		(0.21)	(0.20)	

(Continued)

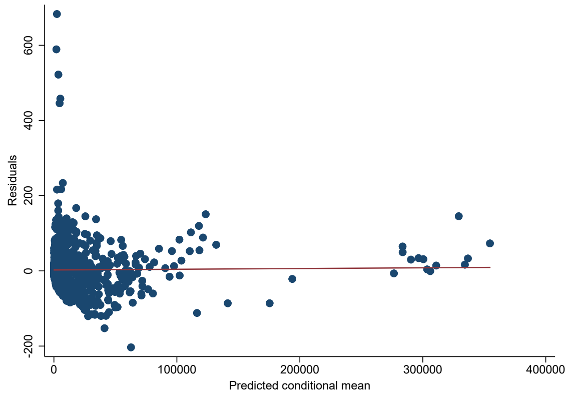
Table 4.B.7 Robustness checks: Determinants of FDI in peripheral EU countries, 1996-2012. *(Continued)*

	GLMs			
	PPML	GPML	NBPML	Gaussian GLM
ln_h_rail	0.842***	0.356***	0.359***	0.906***
	(0.11)	(0.13)	(0.12)	(0.10)
ln_h_lines	-0.420**	-0.576**	-0.565**	-0.395**
	(0.20)	(0.24)	(0.24)	(0.18)
ln_h_cell	0.246***			0.129*
	(0.09)			(0.08)
h_pv				
Host country FE (j)	*Yes*	*Yes*	*Yes*	
Year FE	*Yes*	*Yes*	*Yes*	
Observations	272	272	272	272
RESET test $p-values$	0.1034	0.0147	0.0064	0.7490
AIC	330.5292	18.38799	17.07957	17.83051
BIC	85743.01	-1410.356	-1136.116	7.90e+08
Deviance	87183.69853	30.33500804	304.5751082	790471856
Dispersion	339.2362	0.1180351	1.185117	3075766
Bias	0.1096489	0.0487685	0.0506312	0.1691215
MSE	0.1946108	0.1294503	0.1301796	0.2521434
ErrorLoss	0.2676427	0.2474676	0.2460399	0.3002481

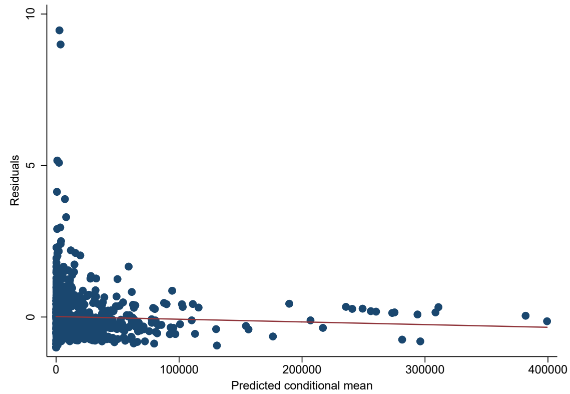
Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Fig. 4.B.1 Robustness checks: GLMs estimators for all host countries: Predictions and Pearson residuals.

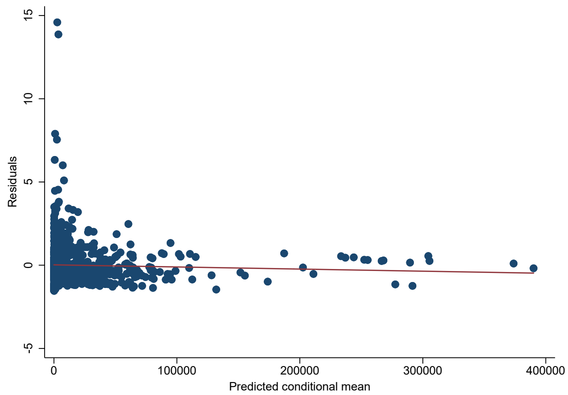
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

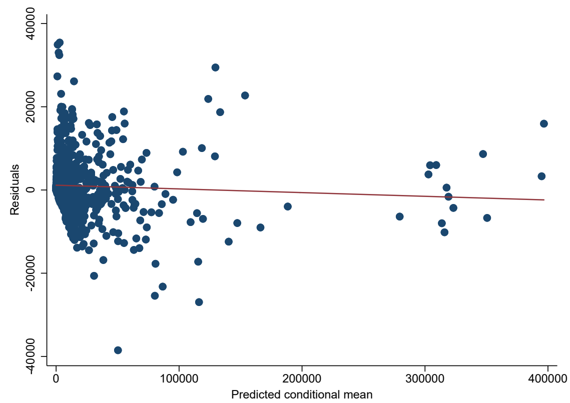
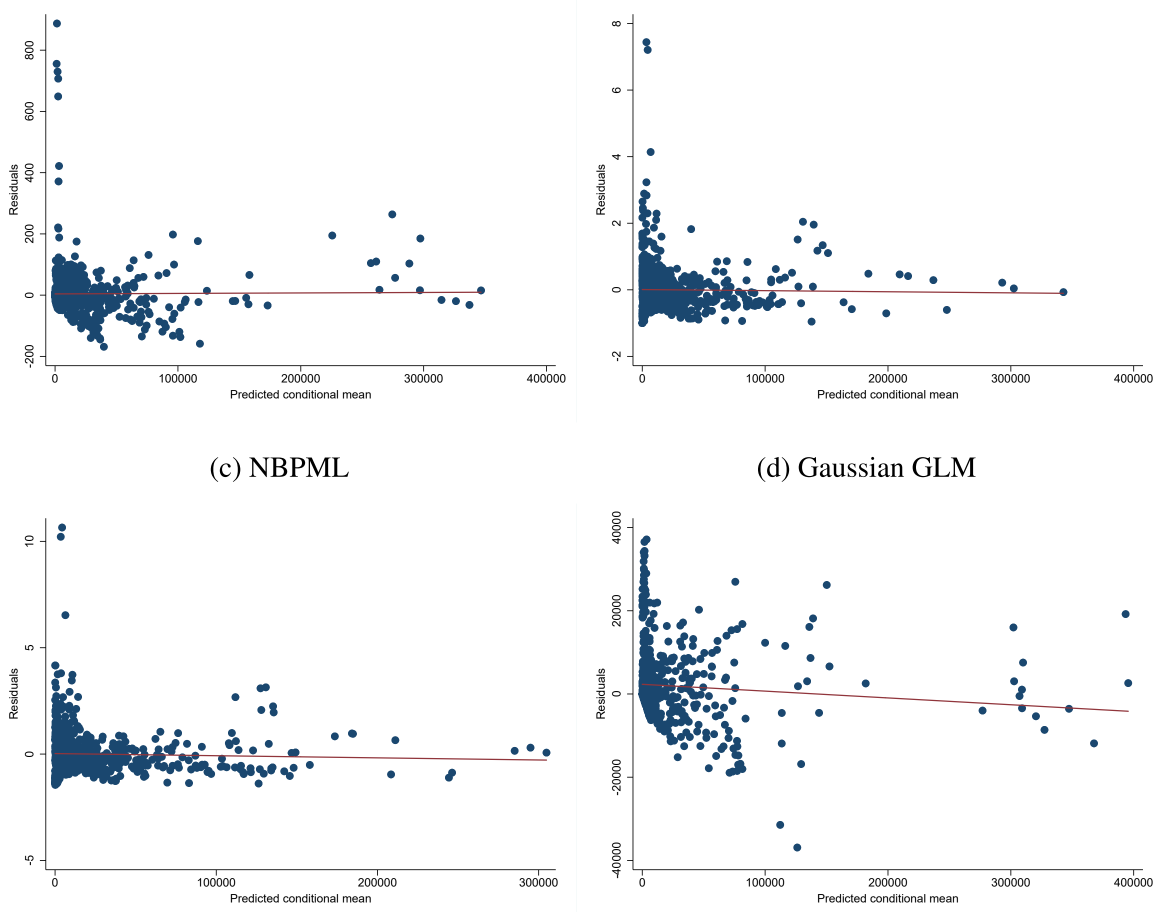


Fig. 4.B.2 Robustness checks: GLMs estimators for developed countries: Predictions and Pearson residuals.



(a) PPML

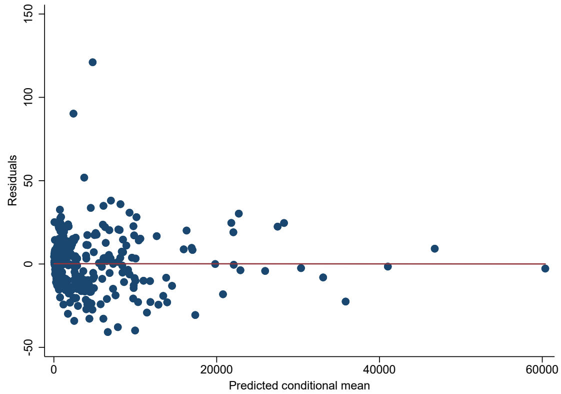
(b) GPML

(c) NBPML

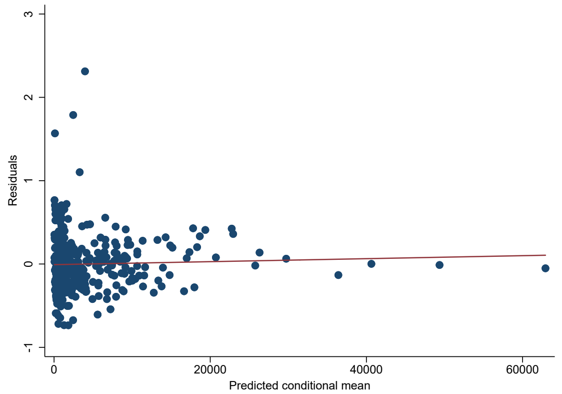
(d) Gaussian GLM

Fig. 4.B.3 Robustness checks: GLMs estimators for developing countries: Predictions and Pearson residuals.

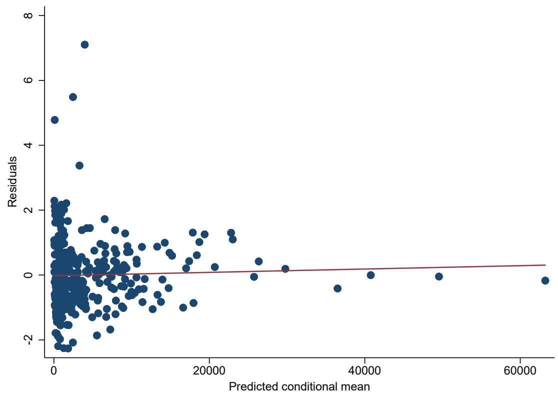
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

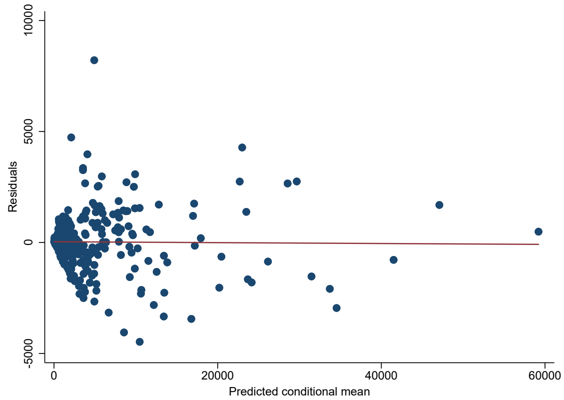


Fig. 4.B.4 Robustness checks: GLMs estimators for Latin American countries: Predictions and Pearson residuals.

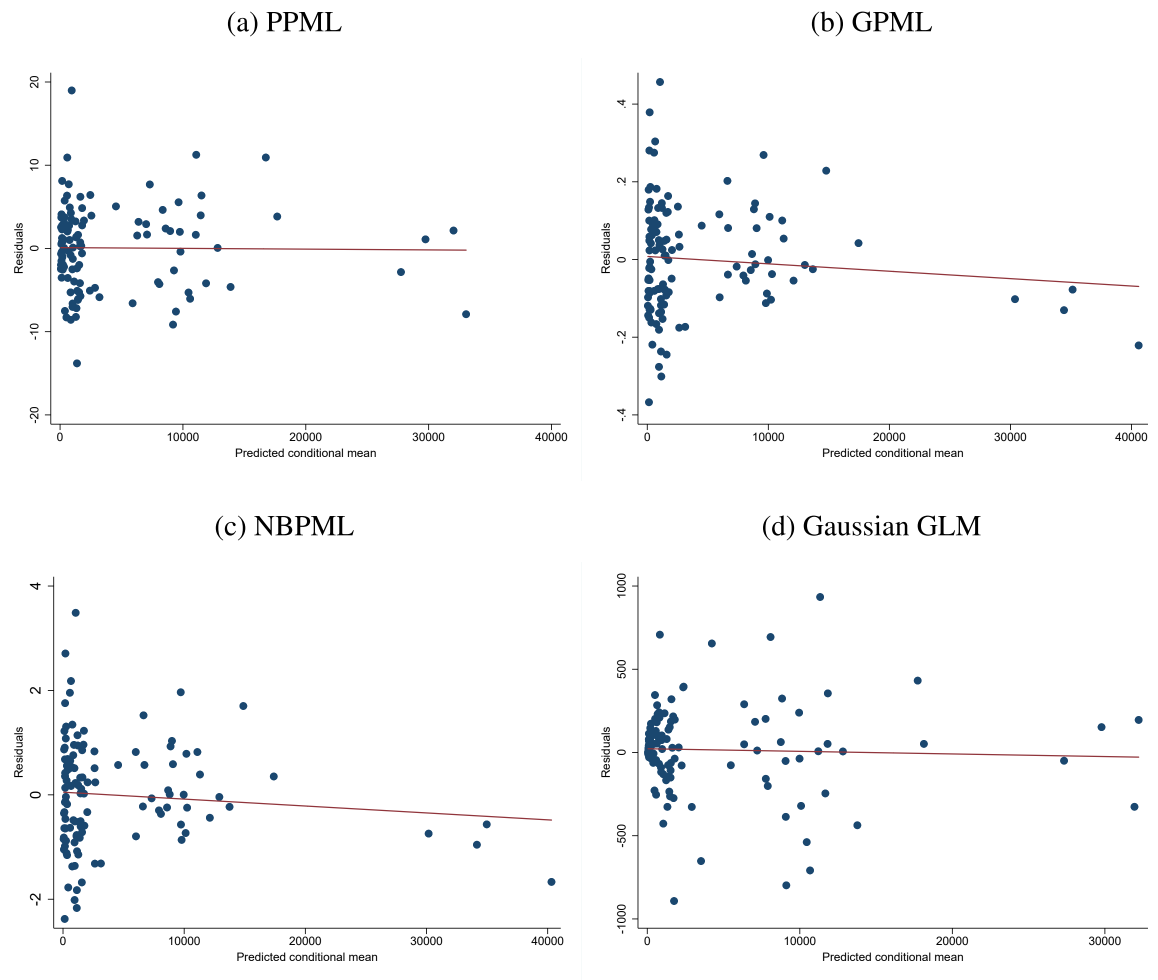
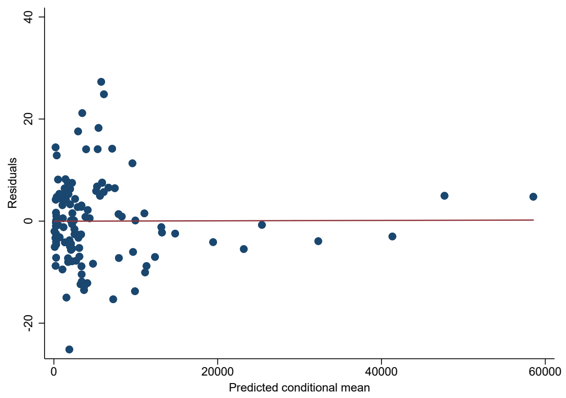
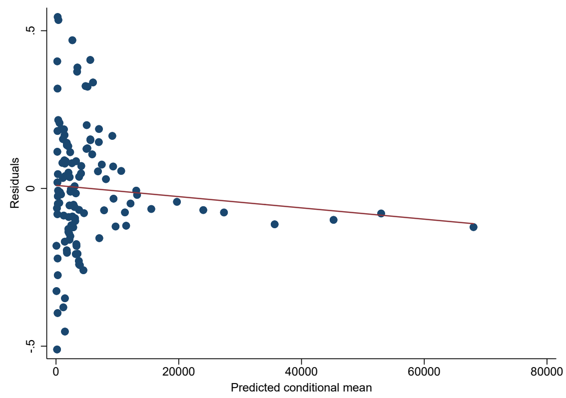


Fig. 4.B.5 Robustness checks: GLMs estimators for Asian countries: Predictions and Pearson residuals.

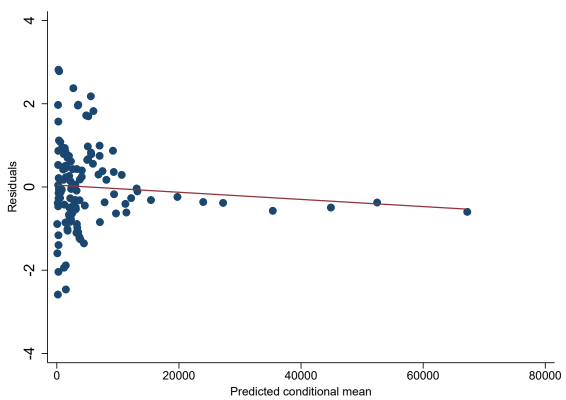
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

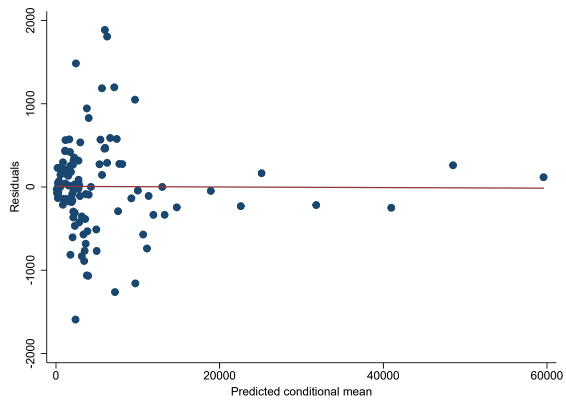
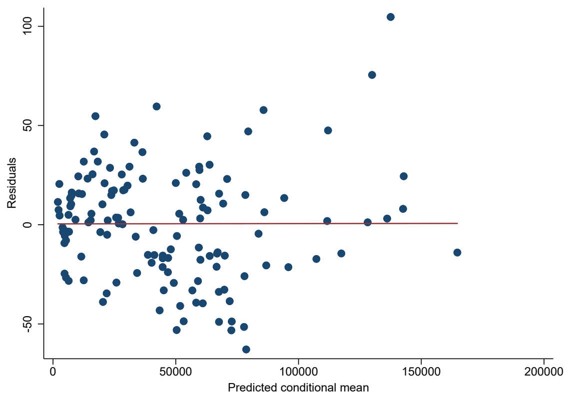
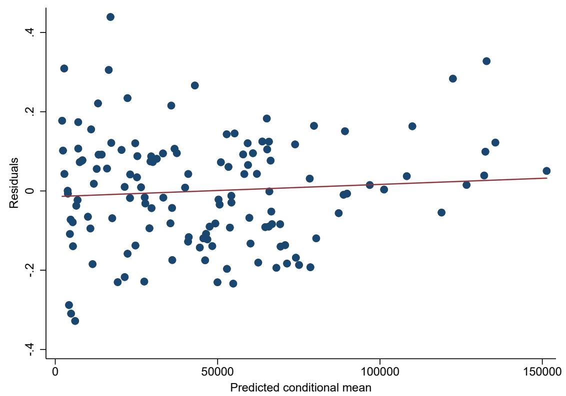


Fig. 4.B.6 Robustness checks: GLMs estimators for Core EU countries: Predictions and Pearson residuals.

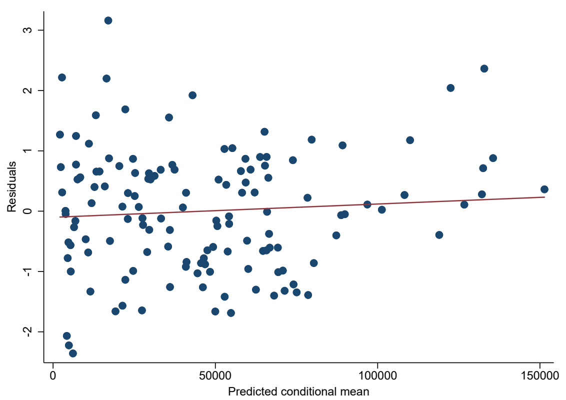
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

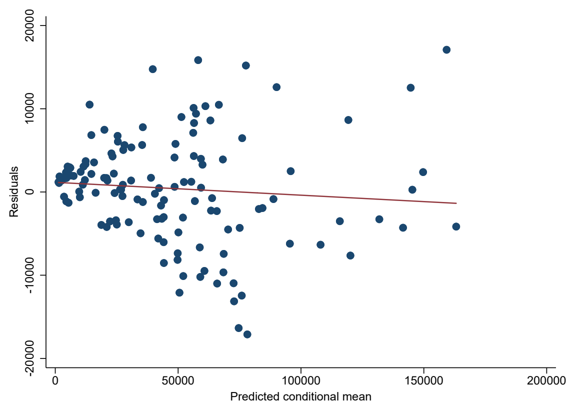


Fig. 4.B.7 Robustness checks: GLMs estimators for Peripheral EU countries: Predictions and Pearson residuals.

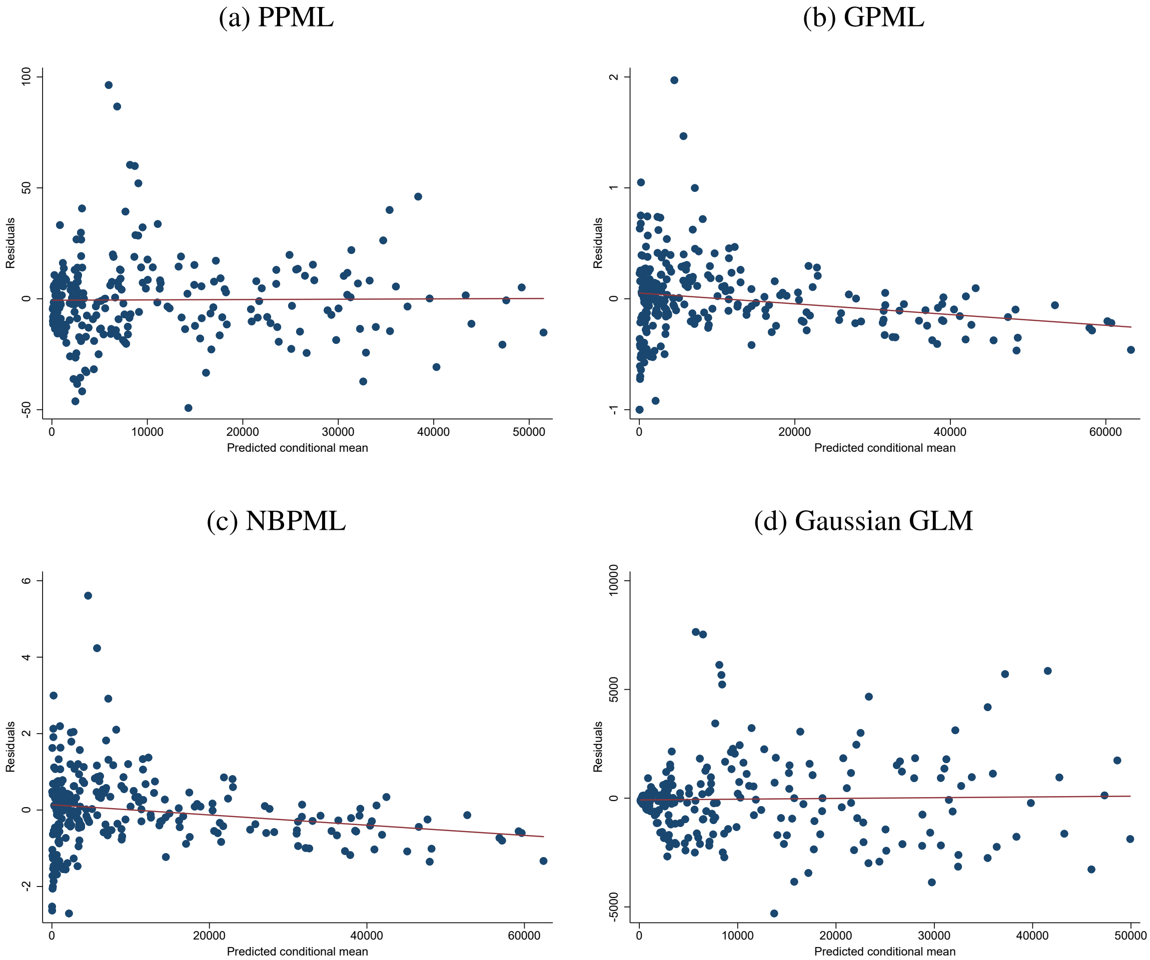
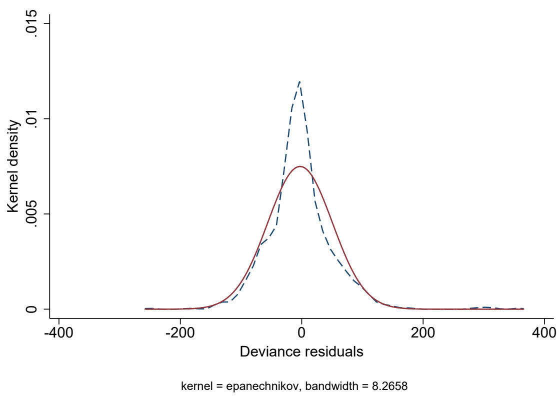
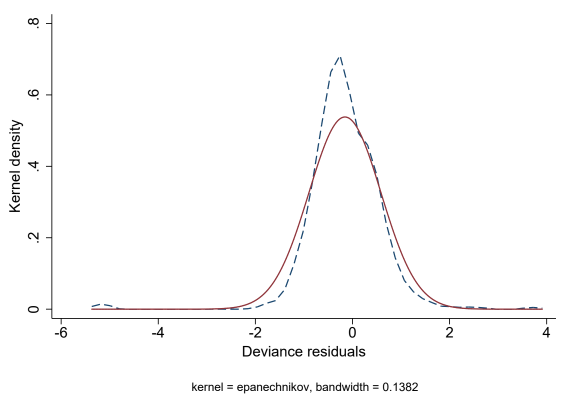


Fig. 4.B.8 Robustness checks: GLMs estimators for all host countries: Density of deviance residuals.

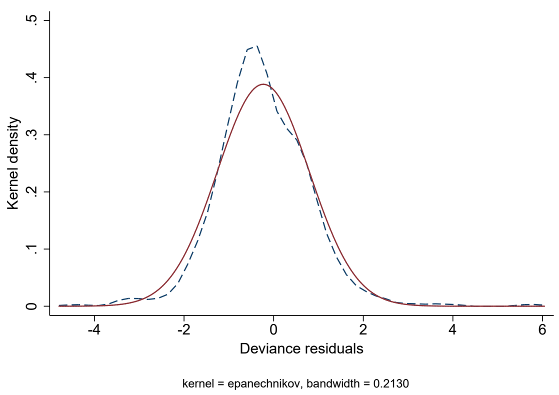
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

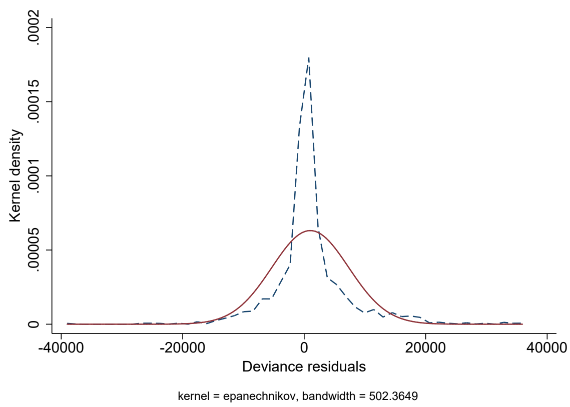


Fig. 4.B.9 Robustness checks: GLMs estimators for developed countries: Density of deviance residuals.

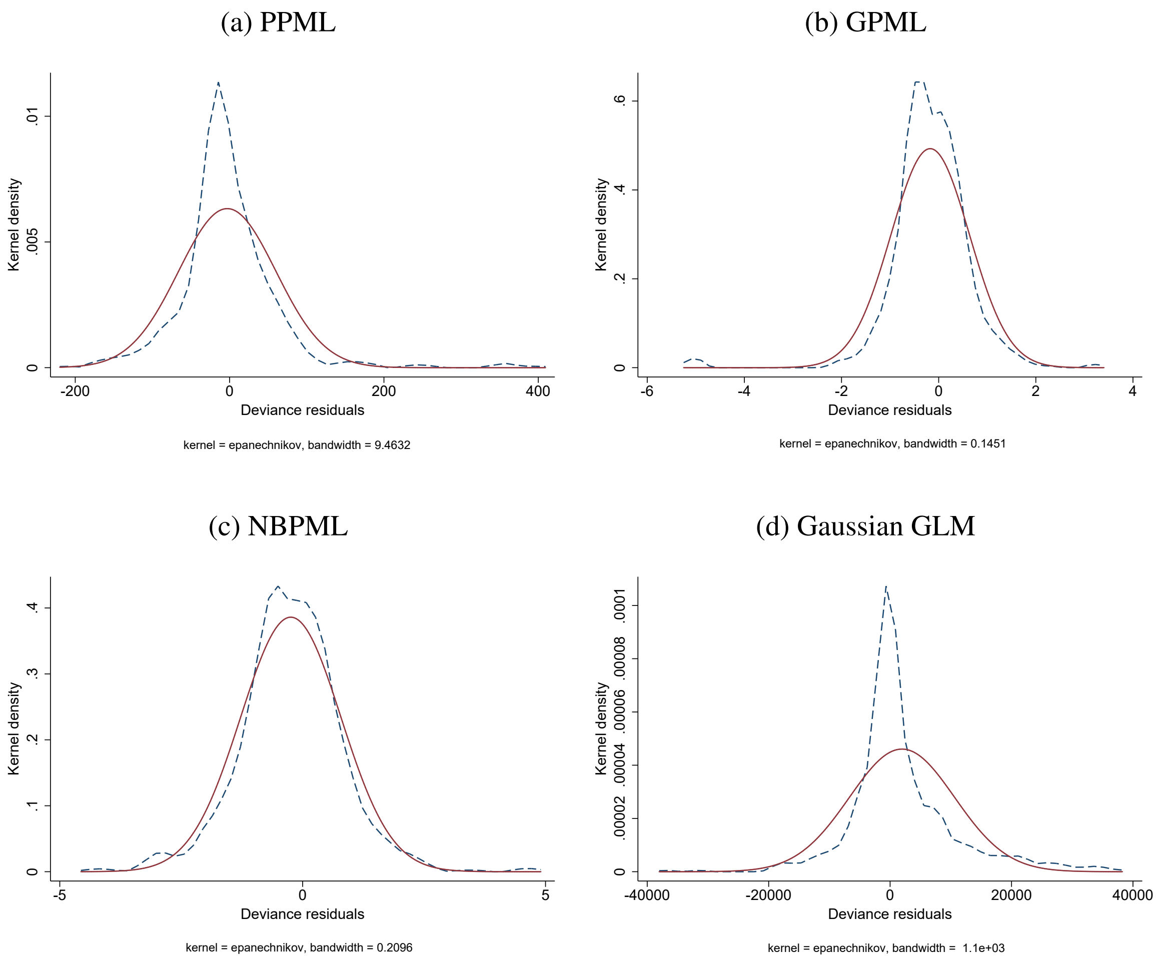


Fig. 4.B.10 Robustness checks: GLMs estimators for developing countries: Density of deviance residuals.

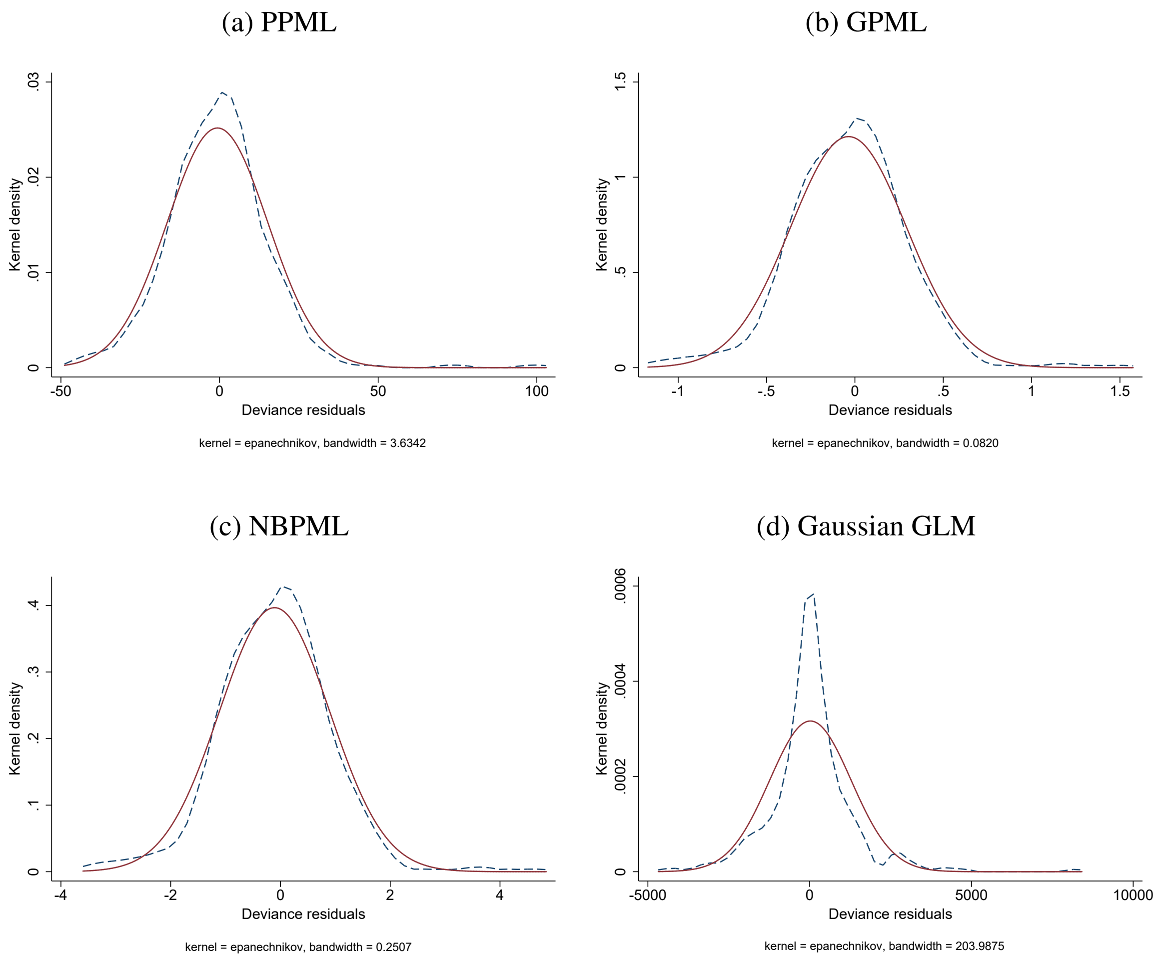
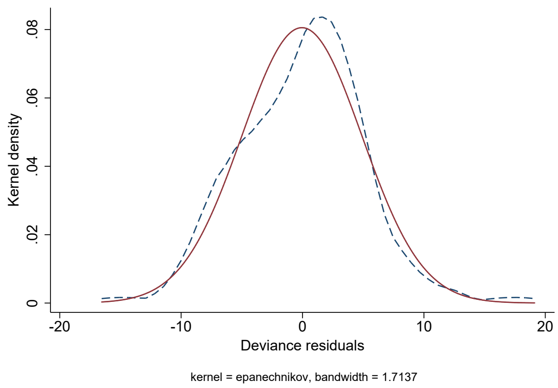
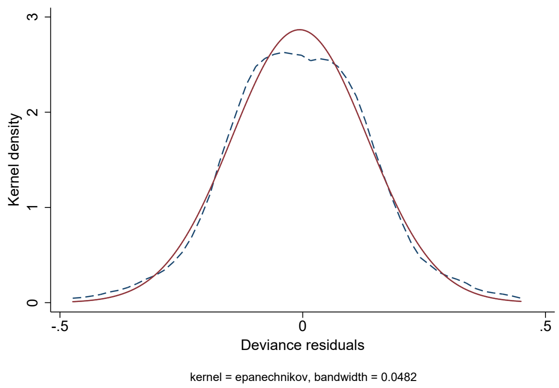


Fig. 4.B.11 Robustness checks: GLMs estimators for Latin American countries: Density of deviance residuals.

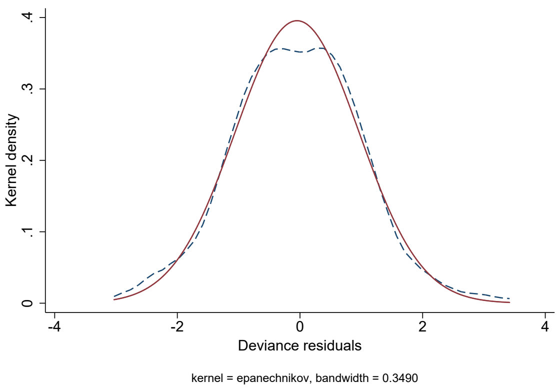
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

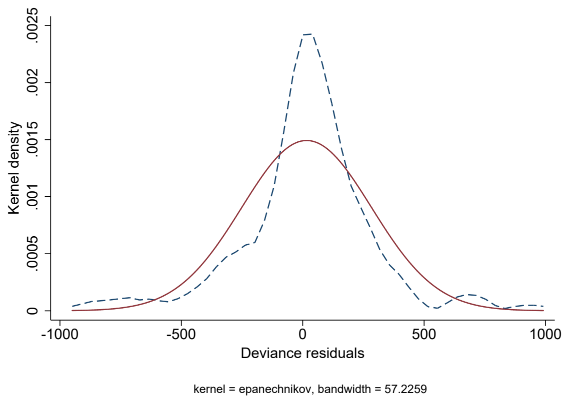
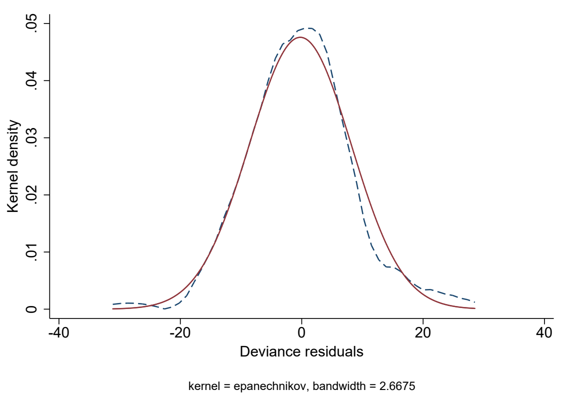
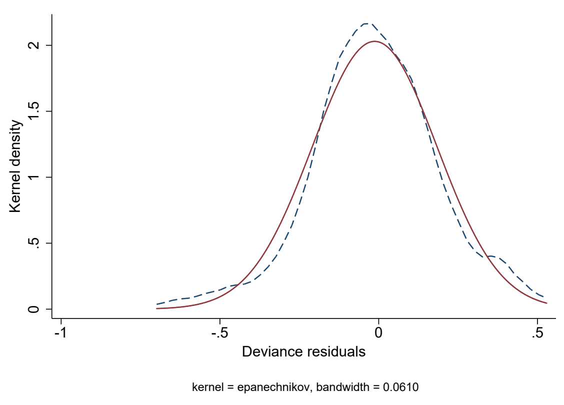


Fig. 4.B.12 Robustness checks: GLMs estimators for Asian countries: Density of deviance residuals.

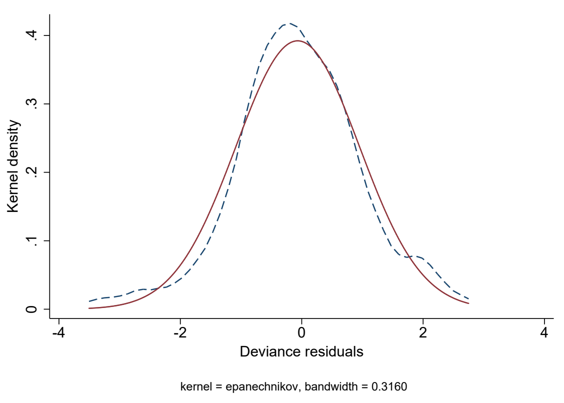
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

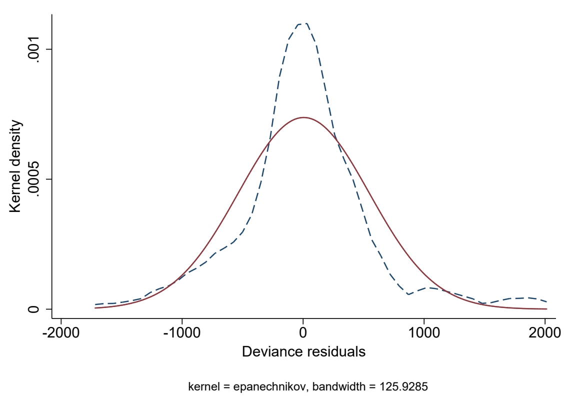
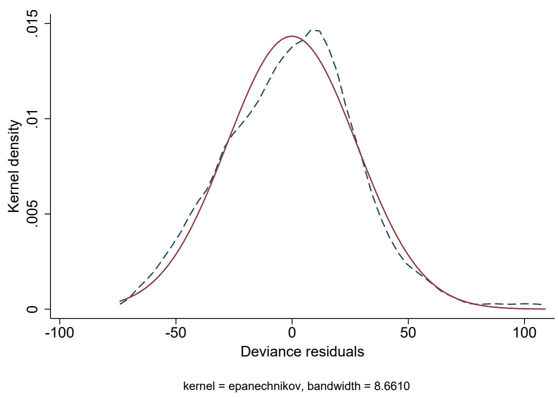
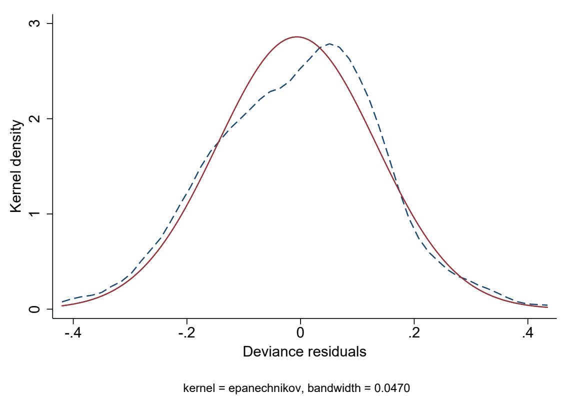


Fig. 4.B.13 Robustness checks: GLMs estimators for Core EU countries: Density of deviance residuals.

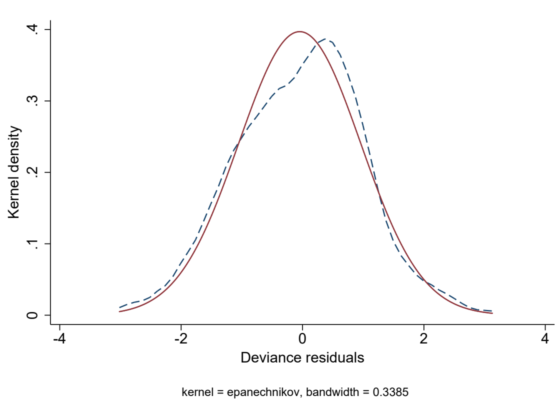
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

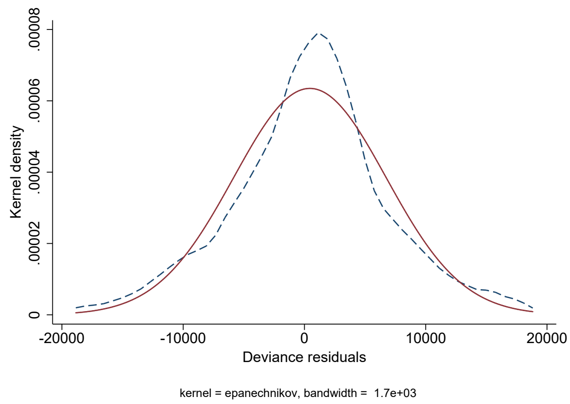
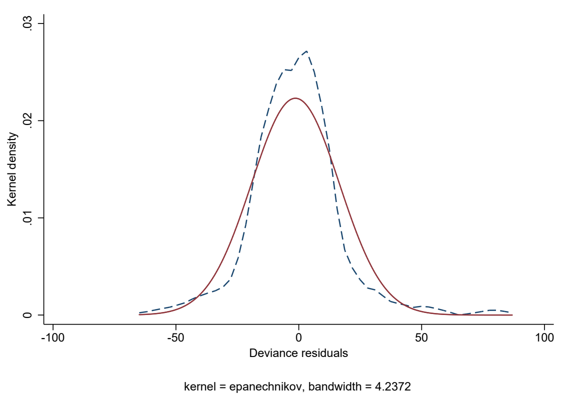
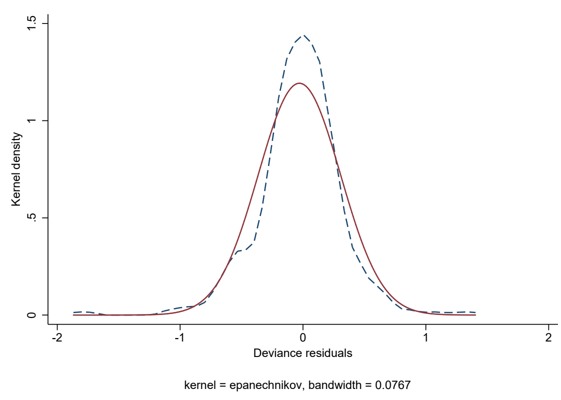


Fig. 4.B.14 Robustness checks: GLMs estimators for Peripheral EU countries: Density of deviance residuals.

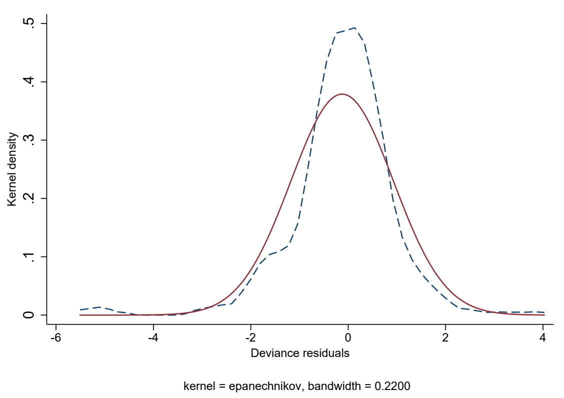
(a) PPML



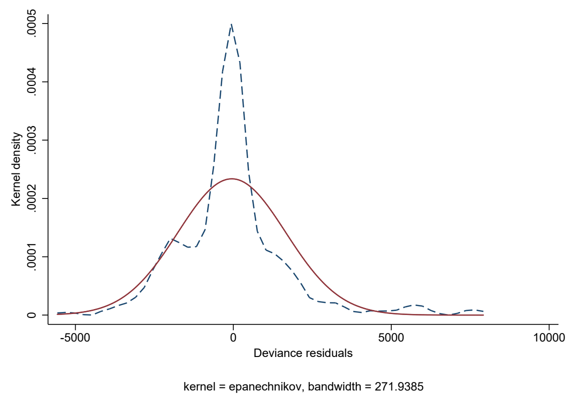
(b) GPML



(c) NBPML



(d) Gaussian GLM



Chapter 5

Conclusions

Over the last decades, FDI has attracted considerable attention of researchers and policy makers due to its remarkable growth, geographical pattern and associated potential benefits. Nevertheless, discussing the ability of national and regional governments to attract FDI requires a thorough understanding of the factors driving MNEs activities. This has been particularly important in the European Union where the internal market triggered cross-border investments within and outside its members.

Theoretical literature on FDI has developed several alternative theories, thus putting forward different variables as driving forces of cross-border investments. Particularly, the gravity model approach has been successfully and frequently applied in several empirical studies. However, concerns have been raised recently regarding the lack of consensus in the modelization of FDI due to extensive differences in terms of the variables considered, model specification and estimation methods of the FDI gravity model across studies.

In this context, the aim of the present doctoral dissertation has been to contribute to the literature by investigating the driving forces of MNEs activities to and from European countries, both at the regional and national level, tackling the variable selection and model uncertainty problems faced in the modelization of FDI.

For this purpose, in chapter 2, given our interest in explaining the geographical distribution of FDI across regions, we decided to focus on inward FDI stock in Spain. There are two main reasons for choosing this country as the focus of our interest. First, Spain was the largest capital importer since the launching of the euro, within peripheral EU countries. Second, FDI is highly unevenly distributed across regions. Consequently, in this chapter we strive to identify the long-run FDI determinants in Spanish regions. To that end, we conduct an exploratory factor analysis (EFA) that allows us to address the collinearity problem that arise when considering an ad hoc set of variables put forward by the literature as potential drivers of FDI and then, we estimate an extended gravity model.

The two chapters thereafter deal with the modelization of German outward FDI. We focus exclusively on Germany because it is considered one of the main investors not only among European countries but also among developed economies worldwide. Furthermore, in the context of the current expansion of Global Value Chains (GVCs) and the associated complex integration strategies of MNEs, Germany has established itself as the core of the European production hub. Accordingly, because we are interested in examining the long-run German FDI determinants, chapter 3 addresses the selection of the appropriate set of variables to include in a regression model to explain the pattern of FDI stocks. To overcome the variable selection problem we apply a Bayesian Model Averaging approach that has the advantage of attaching probabilities to any of the possible model specifications over the model space. We conduct this analysis for country-groups rather than for the total sample in order to take into account heterogenuos FDI motivations in different countries that would be disregarded otherwise.

Finally, following up on the results from chapter 3, the fourth chapter addresses uncertainty in the econometric specification of the FDI gravity model. We compare the performance of alternative Generalized Linear Model (GLM) estimators in order to select the proper estimator for our data set that precisely estimates the coefficients of the robust German outward FDI determinants.

Accordingly, the chapters included in this dissertation empirically contribute to extant discussions in the literature and add to the knowledge in several aspects. First, because FDI flows are volatile through time, we consider FDI stock data is more suitable when the interest of the resarcher is the long-run determinants of FDI. To the best of our knowledge, we are one of the first studies to use FDI stock data to explore the long-run FDI determinants in Spanish regions. Second, we acknowledge the model uncertainty problem in the FDI literature and propose to adopt a Bayesian Model Averaging (BMA) approach to overcome such a challenge. In relation to this, we also add to the literature by comparing several methods estimating gravity models in their multiplicative form in a GLM framework. Another contribution relates to the large number of variables considered in the empirical analysis as potential FDI determinants. Third, both data sets used in this dissertation cover the most recent period available for bilateral FDI stock data and include extensive coverage in terms of the countries considered. Another key contribution is to disaggregate destination countries into different country-groups in order to better disentangle different FDI motivations. Finally, our investigation adds to the previously rather overlooked literature devoted to the understanding of the driving forces of German MNEs activities.

In broad terms, our findings are in line with recent contributions that point towards a parsimonous FDI specification. In relation to the most appropriate estimation method for the gravity model, our analysis in chapter 4 provides further evidence in favor of the Negative Binomal Pseudo Maximum Likelihood (NBPML) estimator. Furthermore, a few general policy implications can be drawn from our findings related to the factors that should be emphasized to attract FDI either at the regional or national level.

At the regional level, our findings in chapter 2 show a statistically significant positive impact of the *Competitivenes and agglomeration* factor on inward FDI stock that ranges from 0.6% to 0.7% across the different model specifications. This would mean that initiatives to improve the relative competitiveness of the regions are especially important for attracting FDI. Such initiatives include promoting internationalization of the firms and improve factor endowments. We find also a positive impact of the *Productive capacity* factor, at the 10% level of statistical significance. In this respect, initiatives to encourage an adequate transport infrastructure would also increase the attraction of FDI. Interestingly, we do not find statistical support for the *Economic potential* factor in our chosen specifications. This is a notable outcome as it implies that at the regional level FDI is not market seeking but efficiency seeking. As regards additional explanatory variables, our findings also emphasize the geographic disadvantage of being landlocked to attracting FDI with the exception of being the capital city. Not surprisingly, there is no impact of *Distance* from the regional standpoint. We also find evidence of a positive (and statistically significant) impact of the regional location quotient for the industry sector on inward FDI stock. This outcome seems to indicate that the creation of solid industry clusters would also strengthen FDI attractiveness. Although these findings are provided for the case of Spain, it could also be extended to other countries with a federal system where local governments can influence FDI attraction.

At the national level, the Bayesian Model Averaging analysis conducted in chapter 3 for different recipient country-groups, and subsequent estimation of the parameters of interest in chapter 4 reveal different set of key factors for FDI attraction into different country-groups. Firstly, our analysis in chapter 3 enables the identification of those variables with the highest probabilities to be included in the true model of bilateral FDI patterns. The results highlight GDP and population measures as critical variables for explaining FDI in all country-groups. We find also robust evidence of distance and other geography measures for all subsamples with the exception of developed countries. As regards variables associated to factor endowments and productivity, our results suggest that they appear to be more relevant for FDI in developing countries and particularly, Latin American countries. Cultural and historical factors, in turn, appear to be robust determinants only for developed countries. Interestingly, competitiveness, captured by the exchange rate, does not have a high inclusion probability for Latin American countries. It should also be mentioned that trade and investment treaties exert high inclusion probabilities for both developed and developing countries; yet when we further disaggregate our country-groups they appear to be robust only for peripheral EU countries. Something similar happens with variables related to trade openness, however no statistical evidence is found for this group of variables in Asian countries. Our analysis reveals also statistical support for telecomunications infrastructure in all country-groups as well as transport infrastructure in peripheral EU countries. Finally, the most striking result is the notably relevance of institutions as FDI determinants in all country-groups, with the exception of core EU countries.

Secondly, the comparative analysis of GLM estimators in chapter 4 provides a more accurate estimation of the model that yield information on the magnitude and sign of the robust FDI determinants. For developed countries, we find a positive impact of an effective host country government and control of corruption on FDI of 2.2% and 1.9%, respectively. Furthermore, results show a positive (and highly significant) impact of the EU Customs Union on German outward FDI, presenting a coefficient estimate of 0.776. We find also (weakly significant) evidence of drivers linked to vertical strategies. On the other hand, for FDI directed towards developing countries, our results show a positive impact of differences in education levels of 0.609% on FDI. We also find that FDI into these economies is motivated by access to large markets. Additionally, our findings highlight the key role of trade treaties and trade openness variables. A striking result is the (highly significant) positive impact of the EU-Turkey custom union, implying three times more German FDI into Turkey. Good institutions are also an attractive factor for FDI in developing economies, yet for this country-group the voice and accountability and regulatory quality indeces are the relevant institutional factors for FDI. Further exploring FDI into developing countries shows that market access is the most relevant attractive factor in Asian countries. In this respect, we find a positive (and highly significant) impact of the sum of the two countries' real GDPs on outward FDI stock of 1.757%. Meanwhile, factors associated with efficiency-seeking FDI are key for investment in Latin America. Finally, considering intra-European FDI, results show that access to large European markets is the primary driver of FDI in "core" EU countries. Particularly, we find a positive (and highly significant) impact of market size on German FDI of 3.214%. On the contrary, our results show a statistically significant impact of trade agreements on FDI directed towards peripheral EU economies. These findings seem to indicate export-platform FDI motivations, highlighting the role of countries' embeddedness into the global value chains network.

The research conducted in this dissertation presents some limitations that also give rise to potential lines of future research. In relation to the modelization of FDI from a regional perspective, the availability of FDI stock data disaggregated across sectors would allow to provide more detailed advice concerning the regional FDI determinants. Concerning our BMA approach, it would also be of interest extending the analysis to account for the selection bias that results from the log-linearization. This line of research, motivated by the work of Eicher et al. (2012), would allow us to explore the FDI determinants in the intensive (i.e. how much to invest in a particular location) and extensive (i.e. the decision to invest) FDI margins. Finally, it could also be investigated the role of third-country effects as determinants of bilateral FDI: the decision to invest in a particular host country may also depend on the FDI directed towards neighbouring countries. In this respect, incorporating the spatial interdependence in FDI data would also be very helpful to this field of research.

Conclusiones

En las últimas décadas, la IED ha atraído una considerable atención de investigadores y responsables de la formulación de políticas debido a su notable crecimiento, patrón geográfico y beneficios potenciales asociados. Sin embargo, debatir sobre la capacidad de los gobiernos nacionales y regionales para atraer la IED requiere una comprensión profunda de los factores que impulsan las actividades de las EMNs. Esto ha sido particularmente importante en la Unión Europea, donde el mercado interno provocó inversiones transfronterizas entre sus miembros y con terceros países.

La literatura teórica sobre IED ha desarrollado diversas teorías alternativas, por consiguiente se han postulado diferentes variables como fuerzas impulsoras de las inversiones transfronterizas. En particular, el modelo de gravedad se ha utilizado con éxito en numerosos estudios empíricos. Sin embargo, recientemente la falta de consenso en la modelización de la IED ha suscitado inquietudes en los investigadores debido a las amplias diferencias en los estudios empíricos en términos de las variables consideradas, la especificación y los métodos de estimación del modelo de gravedad de la IED.

En este contexto, el objetivo de la presente disertación ha sido contribuir a la literatura mediante la investigación de las fuerzas impulsoras de las actividades de las EMNs hacia y desde los países europeos, tanto a nivel regional como nacional, abordando los problemas de selección de variables e incertidumbre del modelo que se enfrentan al modelizar la IED.

Para este propósito, en el capítulo 2, dado nuestro interés en explicar la distribución geográfica de la IED entre las regiones, decidimos centrarnos en el stock de IED entrante en España. Hay dos razones principales para elegir este país como foco de nuestro interés. En primer lugar, España fue el mayor importador de capital desde el lanzamiento del euro, entre los países periféricos de la UE. En segundo lugar, la IED se distribuye de manera muy desigual entre las regiones. En consecuencia, en este capítulo tratamos de identificar los determinantes de la IED a largo plazo en las regiones españolas. Con ese fin, llevamos a cabo un análisis factorial exploratorio (*exploratory factor analysis*; EFA) que nos permite abordar el problema de colinealidad que surge al considerar un conjunto de variables ad hoc presentado por la literatura como posibles determinantes de la IED y luego, estimamos un modelo de gravedad extendida.

Los dos capítulos posteriores tratan sobre la modelización de la IED saliente de Alemania. Nos centramos exclusivamente en Alemania porque se considera uno de los principales inversores no solo entre los países europeos sino también entre las economías desarrolladas en todo el mundo. Además, en el contexto de la expansión actual de las cadenas globales de valor y las estrategias complejas de integración asociadas a las EMNs, Alemania se ha establecido como el núcleo del centro de producción europeo. En consecuencia, puesto que estamos interesados en examinar los determinantes de la IED alemana a largo plazo, el capítulo 3 aborda la selección del conjunto apropiado de variables a incluir en un modelo de regresión para explicar la distribución del stock de IED. Para solventar el problema de selección de variables, aplicamos un enfoque de Promedio Bayesiano de Modelos (*Bayesian Model Averaging*; BMA) que tiene la ventaja de asociar probabilidades a todas las combinaciones de modelos posibles. Llevamos a cabo este análisis por grupos de países en lugar de para el total de la muestra con el fin de tener en cuenta las motivaciones heterogéneas de la IED en diferentes países que de otro modo no serían consideradas.

Finalmente, partiendo de los resultados del capítulo 3, el cuarto capítulo aborda la incertidumbre en la especificación econométrica del modelo de gravedad de la IED. Comparamos el desempeño de estimadores alternativos en el marco de Modelos Lineales Generalizados (*Generalized Linear Models*; GLMs) para seleccionar el estimador adecuado para nuestra base de datos que estime con precisión los coeficientes de los determinantes robustos de la IED saliente de Alemania.

En consecuencia, los capítulos incluidos en esta disertación contribuyen empíricamente a las discusiones existentes en la literatura en varios aspectos. En primer lugar, debido a que los flujos de IED son volátiles en el tiempo, consideramos que los datos de stock de IED son más adecuados cuando el interés del investigador son los determinantes a largo plazo de la IED. Hasta donde sabemos, proporcionamos uno de los primeros estudios en utilizar datos de stock de IED para investigar los determinantes de la IED a largo plazo en las regiones españolas. En segundo lugar, reconocemos el problema de incertidumbre acerca del modelo en la literatura de la IED y proponemos adoptar un enfoque BMA para solventar dicho desafío. En relación con esto último, también contribuimos a la literatura mediante la comparación de diversos métodos de estimación del modelo de gravedad en su forma multiplicativa en un marco GLM. Otra contribución está relacionada con la gran cantidad de variables consideradas en el análisis empírico como posibles determinantes de la IED. En tercer lugar, ambas bases de datos utilizadas en esta disertación cubren el período más reciente disponible para los datos bilaterales de stock de IED e incluyen una amplia cobertura en términos de los países considerados. Otra contribución clave consiste en la desagregación de los países de destino en diferentes grupos de países con el objetivo de esclarecer las diferentes motivaciones de la IED. Finalmente, nuestra investigación contribuye a la escasa literatura dedicada a la comprensión de las fuerzas impulsoras de las actividades de las EMNs alemanas.

En términos generales, nuestros resultados están en línea con las contribuciones recientes que apuntan hacia una especificación parsimoniosa de la IED. En relación con el método de estimación más apropiado para el modelo de gravedad, nuestro análisis en el capítulo 4 proporciona más evidencia a favor del estimador de Pseudo Máxima Verosimilitud de la Binomial Negativa (*Negative Binomal Pseudo Maximum Likelihood*; NBPML). Además, los resultados obtenidos permiten extraer algunas implicaciones políticas relacionadas con los factores que se deben enfatizar para atraer la IED tanto a nivel regional como nacional.

A nivel regional, los resultados obtenidos en el capítulo 2 muestran un impacto positivo y estadísticamente significativo del factor *Competitividad y efectos de aglomeración* en el stock de IED entrante que varía de 0.6% a 0.7% en las diferentes especificaciones del modelo. Esto sugiere que iniciativas enfocadas a mejorar la competitividad relativa de las regiones son especialmente importantes para atraer la IED. Dichas iniciativas incluyen promover la internacionalización de las empresas y mejorar la dotación de factores. También encontramos un impacto positivo del factor *Capacidad productiva*, con un 10% de significación. A este respecto, las iniciativas para fomentar una infraestructura de transporte adecuada también potenciarían la atracción de la IED. Curiosamente, no encontramos un efecto estadísticamente significativo para el factor *Potencial económico o de mercado* en nuestras especificaciones preferidas. Este es un resultado relevante, ya que implica que a nivel regional la IED no está motivada por la búsqueda de mercados sino de eficiencia. Con respecto a variables explicativas adicionales, nuestros resultados enfatizan también la desventaja geográfica de las regiones sin salida al mar para la atracción de la IED, a excepción de la capital. Como se esperaba, la *Distancia* no tiene un impacto desde el punto de vista regional. También encontramos evidencia de un impacto positivo (y estadísticamente significativo) del cociente de localización regional para el sector industrial en el stock de IED entrante. Este resultado parece indicar que la creación de clusters industriales sólidos también favorecería el atractivo de la IED. Aunque estos resultados se proporcionan para el caso de España, también podrían extenderse a otros países con un sistema federal donde los gobiernos locales pueden influir en la atracción de la IED.

A nivel nacional, el análisis BMA realizado en el capítulo 3 para diferentes grupos de países receptores, y la posterior estimación de los parámetros de interés en el capítulo 4 revelan diversos factores clave para la atracción de la IED en diferentes grupos de países. En primer lugar, nuestro análisis en el capítulo 3 permite la identificación de aquellas variables con las probabilidades más altas de ser incluidas en el verdadero modelo del patrón de IED bilateral. Los resultados destacan el PIB y las medidas de población como variables críticas para explicar la IED en todos los grupos de países. También encontramos evidencia sólida de la distancia y otras medidas geografícas para todas las submuestras, con excepción de los países desarrollados. Con respecto a las variables asociadas a la dotación de factores y la productividad, nuestros resultados sugieren que parecen ser más relevantes para la IED en los países en desarrollo y, particularmente, en los países latinoamericanos. Los factores culturales e históricos, a su vez, parecen ser determinantes sólidos solo para los países desarrollados. Curiosamente, la competitividad, capturada por el tipo de cambio, no tiene una alta probabilidad de inclusión para los países latinoamericanos. También debe mencionarse que los tratados de comercio e inversión tienen altas probabilidades de inclusión tanto para los países desarrollados como para los países en desarrollo; sin embargo, cuando desagregamos aún más nuestros grupos de países, parecen ser robustos solo para los países periféricos de la UE. Algo similar sucede con las variables relacionadas con la apertura comercial, sin embargo, no se encuentra evidencia estadística para este grupo de variables en los países asiáticos. Nuestro análisis revela también evidencia estadística para la infraestructura de telecomunicaciones en todos los grupos de países, así como la infraestructura de transporte en los países periféricos de la UE. Finalmente, el resultado más sorprendente es la relevancia de las instituciones como determinantes de la IED en todos los grupos de países, con la excepción de los países centrales de la UE.

En segundo lugar, el análisis comparativo de los estimadores GLM en el capítulo 4 proporciona una estimación más precisa del modelo que aporta información sobre la magnitud y el signo de los determinantes robustos de la IED. Para los países desarrollados, encontramos un impacto positivo de los indicadores institucionales relacionados con la efectividad gubernamental en el país receptor y el control de la corrupción en la IED de 2.2% y 1.9%, respectivamente. Además, los resultados muestran un impacto positivo (y muy significativo) de la Unión Aduanera de la UE en la IED saliente de Alemania, presentando un coeficiente estimado de 0.776. También encontramos evidencia (débilmente significativa) de determinantes vinculados a estrategias verticales. Por otro lado, para la IED dirigida a los países en desarrollo, nuestros resultados muestran un impacto positivo de las diferencias en los niveles de educación del 0.609% sobre la IED. También encontramos que la IED en estas economías está motivada por el acceso a grandes mercados. Además, nuestros resultados resaltan el papel clave de los tratados comerciales y las variables de apertura comercial. Un resultado sorprendente es el impacto positivo (y altamente significativo) de la unión aduanera UE-Turquía, que implica tres veces más IED alemana en Turquía. La calidad de las instituciones también constituye un factor atractivo para la IED en las economías en desarrollo, sin embargo, para este grupo de países, los indicadores relacionados con la expresión y rendición de cuentas y la calidad regulatoria son los factores institucionales relevantes para la IED. Una investigación más profunda de la IED en los países en desarrollo muestra que el acceso a los mercados es el factor atractivo más relevante en los países asiáticos. A este respecto, encontramos un impacto positivo (y muy significativo) de la suma de los PIB reales de los dos países en el stock de IED saliente de 1.757%. Mientras tanto, los factores asociados con la búsqueda de eficiencia son clave para la inversión en América Latina. Finalmente, considerando la IED intraeuropea, los resultados muestran que el acceso a grandes mercados europeos es el principal impulsor de la IED en los países centrales de la UE. Particularmente, encontramos un impacto positivo (y altamente significativo) del tamaño de mercado en la IED alemana de 3.214%. Por el contrario, nuestros resultados muestran un impacto estadísticamente significativo de los acuerdos comerciales sobre la IED dirigida a las economías periféricas de la UE. Estos resultados parecen indicar motivaciones de IED relacionadas con establecer plataformas de exportación, destacando el papel de la integración de los países en las cadenas globales de valor.

La investigación realizada en esta disertación presenta algunas limitaciones que dan lugar a posibles líneas de investigación futuras. En relación con la modelización de la IED desde una perspectiva regional, la disponibilidad de datos de stock de IED desagregados por sectores permitiría proporcionar recomendaciones más específicas sobre los determinantes regionales de la IED. Con respecto a nuestro enfoque BMA, también sería interesante extender el análisis para tener en cuenta el sesgo de selección que resulta de la linealización logarítmica. Esta línea de investigación, motivada por el trabajo de Eicher et al. (2012), nos permitiría explorar los determinantes de los márgenes intensivos (es decir, cuánto invertir en una determinada localización) y extensivos de la IED (es decir, la decisión de invertir). Finalmente, también podría investigarse el papel de los efectos de terceros países como determinantes de la IED bilateral: la decisión de invertir en un país determinado puede depender también de la IED dirigida a los países vecinos. A este respecto, incorporar la interdependencia espacial de la IED también sería muy útil para este campo de investigación.

References

Adam, A. and Filippaios, F. (2007). Foreign direct investment and civil liberties: A new perspective. *European Journal of Political Economy*, 23(4):1038–1052.

Agiomirgianakis, G., Asteriou, D., and Papathoma, K. (2003). The Determinants of Foreign Direct Investment: A Panel Data Study for the OECD Countries. *Discussion Paper Series 03/06, Department of Economics, School of Social Sciences, City University, London*.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with Gravitas : A Solution to the Border Puzzle. *The American Economic Review*, 93(1):170–192.

Antonakakis, N. and Tondl, G. (2015). Robust determinants of OECD FDI in developing countries: Insights from bayesian model averaging. *Cogent Economics and Finance*, 3(1):1–25.

Asiedu, E. (2002). On the Determinants of Foreign Direct Investment to Developing Countries: Is Africa Different? *World Development*, 30(1):107–19.

Asiedu, E. (2006). Foreign Direct Investment in Africa: The Role of Natural Resources, Market Size, Government Policy, Institutions and Political Instability. *World Economy*, 29 (1):63–77.

Asiedu, E., Dzigbede, K., and Nti-Addae, A. (2015). Foreign Direct Investment, Natural Resources and Employment in Sub-Saharan Africa. In Zedillo, E., Cattaneo, O., and Wheeler, H., editors, *Africa at a Fork in the Road: Taking Off or Disappointment Once Again?*, pages 395–414. Yale Center for the Study of Globalization, New Haven.

Assunçao, S., Forte, R., and Teixeira, A. (2011). Location Determinants of FDI: a Literature Review. *FEP Working Paper, No. 433, University of Porto*.

Bae, C. and Jang, Y. J. (2014). The Impact of Free Trade Agreements on Foreign Direct Investment: The Case of Korea. *East Asian Economic Review*, 17(4):417–444.

Bajo-Rubio, O. and Sosvilla-rivero, S. (1994). An Econometric Analysis of Foreign Direct Investment in Spain, 1964-89. *Southern Economic Journal*, 61(1):104–120.

Baldwin, R., Di Nino, V., Fontagné, L., De Santis, R., and Taglioni, D. (2008). Study on the Impact of the Euro on Trade and Foreign Direct Investment. *European Economy Economics Paper, 321*.

Baldwin, R. and Taglioni, D. (2007). Trade Effects of the Euro: a Comparison of Estimators. *Journal of Economic Integration*, 22(4):780–818.

Baltagi, B., Egger, P., and Pfaffermayr, M. (2014). Panel Data Gravity Models of International Trade, in B. H. Baltagi (ed.),The Oxford Handbook of Panel Data, Chapter 20(Oxford: OxfordUniversity Press), 608–42.

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2007). Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics*, 140(1):260–281.

Barba Navaretti, G. and Venables, A. (2004). Multinational Firms in the World Economy. *Princeton University Press, Princeton, NJ*.

Barbieri, M. M. and Berger, J. (2004). Optimal predictive model selection. *Annals of Statistics*, 32(3):870–897.

Baronchelli, A. and Uberti, T. E. (2018). Exports and FDI: comparing networks in the new millenium. *Working Paper CRENoS 201813, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia*.

Basile, R., Castellani, D., and Zanfei, A. (2008). Location choices of multinational firms in Europe: The role of EU cohesion policy. *Journal of International Economics*, 74(2):328–340.

Basu, D. (2019). Bias of OLS Estimators due to Exclusion of Relevant Variables and Inclusion of Irrelevant Variables. *Oxford Bulletin of Economics and Statistics*, pages 1–26.

Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012). Criteria for bayesian model choice with application to variable selection. *Annals of Statistics*, 40(3):1550–1577.

Bénassy-Quéré, A., Coupet, M., and Mayer, T. (2007). Institutional determinants of foreign direct investment. *World Economy*, 30(5):764–782.

Bénassy-Quéré, A., Fontagné, L., and Lahrèche-Révil, A. (2005). How does FDI react to corporate taxation? *International Tax and Public Finance*, 12(5):583–603.

Berden, K., Bergstrand, J. H., and Etten, E. V. (2012). Governance , Globalization , and Selection into Foreign Direct.

Berger, A., Busse, M., Nunnenkamp, P., and Roy, M. (2013). Do trade and investment agreements lead to more FDI? Accounting for key provisions inside the black box. *International Economics and Economic Policy*, 10(2):247–275.

Bergstrand, J. H. and Egger, P. (2004). A Knowledge-and-Physical-Capital Model of International Trade, Foreign Direct Investment, and Outsourcing: Part I, Developed Countries. *Notre Dame University. Manuscript*.

Bergstrand, J. H. and 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.

Bertola, G., Driffill, J., James, H., Sinn, H. W., Sturm, J. E., and Valentinyi, Á. (2013). EEAG (2013), The EEAG Report on the European Economy, "European Imbalances". *CESifo Group Munich*, pages 55–72.

Bevan, A. A. and Estrin, S. (2004). The determinants of foreign direct investment into European transition economies. *Journal of Comparative Economics*, 32(4):775–787.

Blanchard, O. and Acalin, J. (2016). What does measured FDI actually measure? *Policy Brief No. PB16-17. Washington, DC: Peterson Institute of International Economics.*, pages 1–7.

Blonigen, B., Davies, R., and Head, K. (2003). Estimating the Knowledge-Capital Model of the Multinational Enterprise: Comment. *American Economic Review*, 93(3):980–994.

Blonigen, B., Davies, R., Waddell, G., and Naughton, H. (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review*, 51:1303–1325.

Blonigen, B. and Piger, J. (2014). Determinants of foreign direct investment. *Canadian Journal of Economics*, 47(3):775–812.

Blonigen, B. A. (1997). Firm-specific assets and the link between exchange rates and foreign direct investment. *American Economic Review*, 87(3):447–465.

Blonigen, B. A. (2005). A review of the empirical literature on FDI determinants. *Atlantic Economic Journal*, 33(4):383–403.

Bosquet, C. and Boulhol, H. (2014). Applying the GLM Variance Assumption to Overcome the Scale-Dependence of the Negative Binomial QGPML Estimator. *Econometric Reviews*, 33(7):772–784.

Botric, V. and Skuflic, L. (2006). Main Determinants of Foreign Direct Investment in the Southeast European Countries. *Transition Studies Review*, 13(2):359–377.

Brainard, S. L. (1997). An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade. *American Economic Review*, 87(4):520–544.

Brenton, P. and Di Mauro, F. (1999). Economic Integration and FDI: An Empirical Analysis of Foreign Investment in the EU and in Central and Eastern Europe. *Empirica*, 26(2):95–121.

Bruno, R., Campos, N., Estrin, S., and Tian, M. (2016). Technical Appendix to "The Impact of Brexit on Foreign Investment in the UK ". Gravitating Towards Europe : An Econometric Analysis of the FDI Effects of EU Membership. *Retrieved from http://cep.lse.ac.uk/pubs/download/brexit03_technical_paper.pdf.*

Buch, C., Kleinert, J., Lipponer, A., and Toubal, F. (2005). Determinants and effects of foreign direct investment: evidence from German firm-level data. *Economic Policy*, 20(41):52–110.

Buch, C., Kleinert, J., and Toubal, F. (2003). Determinants of German FDI: New Evidence from Micro-Data. *Discussion Paper 09/03, Economic Research Centre of the Deutsche Bundesbank*.

Buch, C. M. and Kleinert, J. (2008). Exchange rates and FDI: Goods versus capital market frictions. *World Economy*, 31(9):1185–1207.

Burger, M., Van Oort, F., and Linders, G. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4(2):167–190.

Busse, M. and Hefeker, C. (2007). Political risk, institutions and foreign direct investment. *European Journal of Political Economy*, 23(2):397–415.

Camarero, M., Forte, A., Garcia-Donato, G., Mendoza, Y., and Ordoñez, J. (2015). Variable selection in the analysis of energy consumption-growth nexus. *Energy Economics*, 52:207–216.

Camarero, M., Gómez-Herrera, E., and Tamarit, C. (2018). New Evidence on Trade and FDI: how Large is the Euro Effect? *Open Economies Review*, 29(2):451–467.

Camarero, M. and Tamarit, C. (2004). Estimating the export and import demand for manufactured goods: The role of FDI. *Review of World Economics*, 140(3):347–375.

Cameron, C. and Trivedi, P. (2009). *Microeconometrics Using Stata*. Stata Press, College Station, Texas.

Campos, N. and Kinoshita, Y. (2003). Why Does FDI Go Where it Goes? New Evidence from the Transition Economies. *Working paper 03/228. International Monetary Fund, Washington, DC.*

Canton, E. and Solera, I. (2016). Greenfield Foreign Direct Investment and Structural Reforms in Europe: What Factors Determine Investments? *European Economy Discussion Paper*, 8022.

Carr, D. L., Markusen, J. R., and Maskus, K. E. (2001). Estimating the Knowledge-Capital Model of the Multinational Enterprise. *American Economic Review*, 91(3):693–708.

Cavallari, L. and D'Addona, S. (2013). Nominal and real volatility as determinants of FDI. *Applied Economics*, 45(18):2603–2610.

Chakrabarti, A. (2001). The Determinants of Foreign Direct Investment : Sensitivity Analyses of Cross-Country Regressions. *Kyklos*, 54(1):89–114.

Chan, M. W. L., Hou, K., Li, X., and Mountain, D. C. (2014). Foreign direct investment and its determinants: A regional panel causality analysis. *Quarterly Review of Economics and Finance*, 54(4):579–589.

Cheng, L. K. and Kwan, Y. K. (2000). What are the Determinants of Location of Foreign Direct Investment? The Chinese Experience. *Journal of International Economics*, 51(2):379–400.

Chidlow, A., Salciuviene, L., and Young, S. (2009). Regional determinants of inward FDI distribution in Poland. *International Business Review*, 18(2):119–133.

Copenhagen Economics (2006). Study on FDI and regional development. Final report, European Commission Directorate - General for Regional Policy, Denmark, available on-line: http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/fdi2006.pdf.

Daude, C. and Stein, E. (2007). The quality of institutions and foreign direct investment. *Economics and Politics*, 19(3):317–344.

Dellis, K., Sondermann, D., and Vansteenkiste, I. (2017). Determinants of FDI inflows in advanced economies: Does the quality of economic structures matter? *ECB Working Paper, No. 2066, European Central Bank,Frankfurt, Germany. https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2066.en.pdf.*

Demekas, D. G., Horváth, B., Ribakova, E., and Wu, Y. (2007). Foreign direct investment in European transition economies-The role of policies. *Journal of Comparative Economics*, 35(2):369–386.

Desbordes, R., Koop, G., and Vicard, V. (2018). One size does not fit all... panel data: Bayesian model averaging and data poolability. *Economic Modelling*, 75:364–376.

Di Giovanni, J. (2005). What drives capital flows? The case of cross-border M&A activity and financial deepening. *Journal of International Economics*, 65(1):127–149.

Dimitropoulou, D., McCann, P., and Burke, S. P. (2013). The determinants of the location of foreign direct investment in UK regions. *Applied Economics*, 45(27):3853–3862.

Dunning, J. (1977). Trade, Location of Economic Activity and the MNE: A Search for an Eclectic Approach. In Ohlin B., Hesselborn PO., W. P., editor, *The International Allocation of Economic Activity*. Palgrave Macmillan UK, London.

Dunning, J. (1979). Explaining Changing Patterns of International Production: In Defence of the Eclectic Theory. *Oxford Bulletin of Economics and Statistics*, 41(4):269–295.

Dunning, J. (2000). The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review*, 9(2):163–190.

Eaton, J. and Tamura, A. (1994). Bilateralism and Regionalism, Japanese and US Trade and Direct Foreign Investment Patterns. *Journal of the Japanese and International Economics*, 8:478–510.

Economou, F. and Hassapis, C. (2015). Foreign direct investment inflows determinants in four South European economies. *Investment Management and Financial Innovations*, 12(3):182–189.

Economou, F., Hassapis, C., Philippas, N., and Tsionas, M. (2017). Foreign Direct Investment Determinants in OECD and Developing Countries. *Review of Development Economics*, 21(3):527–542.

Egea Román, M. P. and López Pueyo, C. (1991). Un estudio sobre la distribución geográfica de la inversión extranjera directa en España. *Información Comercial Española*, 696-697:105–118.

Egger, P. and Merlo, V. (2007). The impact of bilateral investment treaties on FDI dynamics. *World Economy*, 30(10):1536–1549.

Egger, P. and Pfaffermayr, M. (2004a). Distance, Trade and FDI: A Hausman-Taylor SUR Approach. *Journal of Applied Econometrics*, 19(2):227–246.

Egger, P. and Pfaffermayr, M. (2004b). The impact of bilateral investment treaties on foreign direct investment. *Journal of Comparative Economics*, 32(4):788–804.

Egger, P. and Winner, H. (2005). Evidence on corruption as an incentive for foreign direct investment. *European Journal of Political Economy*, 21(4):932–952.

Egger, P. H. and Staub, K. E. (2016). GLM estimation of trade gravity models with fixed effects. *Empirical Economics*, 50(1):137–175.

Eicher, T. S., Helfman, L., and Lenkoski, A. (2012). Robust FDI determinants: Bayesian Model Averaging in the presence of selection bias. *Journal of Macroeconomics*, 34(3):637–651.

Ekholm, K., Forslid, R., and Markusen, J. (2007). Export-platform Foreign Direct Investment. *Journal of the European Economic Association*, 5(4):776–795.

Ethier, W. J. and Markusen, J. R. (1996). Multinational firms, technology diffusion and trade. *Journal of International Economics*, 41(1-2):1–28.

Faeth, I. (2009). Determinants of foreign direct investment - A tale of nine theoretical models. *Journal of Economic Surveys*, 23(1):165–196.

Fernández, C., Ley, E., and Steel, M. F. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16(5):563–576.

Franco, C. (2013). Exports and FDI motivations: Empirical evidence from U.S. foreign subsidiaries. *International Business Review*, 22(1):47–62.

García-Donato, G. and Forte, A. (2015). BayesVarSel: Bayes Factors, Model Choice and Variable Selection In Linear Models. *R package version 1.6.1. URL http://CRAN.R-project.org/package=BayesVarSel.*

Garcia-Donato, G. and Forte, A. (2018). BayesVarSel: Bayesian Testing, Variable Selection and model averaging in Linear Models using R. *The R Journal*, 10(1):155–174.

García-Donato, G. and Martínez-Beneito, M. A. (2013). On sampling strategies in Bayesian variable selection problems with large model spaces. *Journal of the American Statistical Association*, 108(501):340–352.

George, E. and McCulloch, R. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7(2):339–373.

Globerman, S. and Shapiro, D. (2002). Global foreign direct investment flows: The role of governance infrastructure. *World Development*, 30(11):1899–1919.

Gómez-Herrera, E. (2013). Comparing alternative methods to estimate gravity models of bilateral trade. *Empirical Economics*, 44(3):1087–1111.

Gouel, C., Guimbard, H., and Laborde, D. (2012). A Foreign Direct Investment database for global CGE models. *CEPII Working Paper No 2012-08.*

Guadalupe, M., Kuzmina, O., and Thomas, C. (2012). Innovation and Foreign Ownership. *American Economic Review*, 102(7):3594–3627.

Gunnella, V., Al-Haschimi, A., Benkovskis, K., Chiacchio, F., de Soyres, F., Franco-Bedoya, S., Di Lupidio, B., Fidora, M., Frohm, E., Gradeva, K., Lopez-Garcia, P., Koester, G., Nickel, C., Osbat, C., Pavlova, E., Schmitz, M., Schroth, J., Skudelny, F., Tagliabracci, A., Vaccarino, E., Wörz, J., and Dorrucci, E. (2019). The impact of global value chains on the euro area economy. *Occasional Paper Series, No 221, ECB, Frankfurt am Main, April.*

Gutiérrez-Portilla, P., Maza, A., and Villaverde, J. (2019). Has the crisis affected Spanish investment strategy abroad? A spatial panel data approach. *World Economy*, 42(4):1032–1056.

Gutiérrez-Portilla, P., Maza, A., Villaverde, J., and Hierro, M. (2016). Foreign direct investment in the Spanish regions: What are the influencing factors? *Journal of Regional Research*, 35(67).

Hadjit, A. and Moxon-Browne, E. (2005). Foreign Direct Investment in Turkey: The Implications of EU Accession. *Turkish Studies*, 6(3):321–340.

Hair, J., Black, W., Babin, B., and Anderson, R. (2010). *Multivariate Data Analysis*. Pearson Prentice Hall, 7th edition.

Harms, P. and Ursprung, H. W. (2002). Do civil and political repression really boost foreign direct investments? *Economic Inquiry*, 40(4):651–663.

Head, C., Ries, J., and Swenson, D. (1999). Attracting foreign manufacturing: Investment promotion and agglomeration. *Regional Science and Urban Economics*, 29(2):197–218.

Head, K. and Mayer, T. (2014). Gravity Equations: Workhorse,Toolkit, and Cookbook. *In Handbook of international economics (Vol. 4, pp. 131-195). Elsevier.*

Head, K. and Ries, J. (1996). Inter-city competition for foreign investment: Static and dynamic effects of China's Incentive areas. *Journal of Urban Economics*, 40(1):38–60.

Head, K. and Ries, J. (2008). FDI as an outcome of the market for corporate control: Theory and evidence. *Journal of International Economics*, 74(1):2–20.

Helpman, E. (1984). A Simple Theory of International Trade with Multinational Corporations. *Journal of Political Economy*, 92(3):451–471.

Helpman, E. (1985). Multinational Corporations and Trade Structure. *Review of Economic Studies*, LII:443–457.

Helpman, E. (2006). Trade, Fdi, and the Organization of Firms. *Journal of Econolmic Literature*, XLIV(September):589–630.

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

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4):382–417.

Horstmann, I. J. and Markusen, J. R. (1987). Strategic Investments and the Development of Multinationals. *International Economic Review*, 28(1):109–121.

Kandogan, Y. (2012). Regional foreign direct investment potential of the states within the US. *Journal of Economics and Business*, 64(4):306–322.

Kaufmann, D., Kraay, A., and Zoido-Lobatón, P. (1999). Aggregating Governance Indicators. *World Bank Policy Research Paper No. 2195*.

Kersan-Skabic, I. and Tijanic, L. (2014). Regional Determinants of Foreign Direct Investments in Croatia. *Transylvanian Review of Administrative Sciences*, (SI):70–89.

Kleinert, J. and Toubal, F. (2010). Gravity for FDI. *Review of International Economics*, 18(1):1–13.

Lankhuizen, M., de Groot, H., and Linders, G.-J. (2011). The trade-off between foreign direct investments and exports: The role of multiple dimensions of distance. *World Economy*, 34(8):1395–1416.

Leamer, E. (1978). *Specification Searches*. Wiley, New York.

Man, G. (2015). Competition and the growth of nations: International evidence from Bayesian model averaging. *Economic Modelling*, 51:491–501.

Manning, W. G. and Mullahy, J. (2001). Estimating log models: to transform or not to transform? *Journal of Health Economics*, 20(4):461–494.

Markusen, J. and Maskus, K. (2001). Reconciling Theory and Evidence. *In Topics in Empirical International Economics: A Festschrift in Honor of Robert E. Lipsey. University of Chicago Press*, (January):71–98.

Markusen, J. and Maskus, K. (2002). Discriminating Among Alternative Theories of the Multinational Enterprise. *Review of International Economics*, 10(4):694–707.

Markusen, J. R. and Venables, A. J. (1998). Multinational firms and the new trade theory. *Journal of International Economics*, 46(2):183–203.

Martin, W. and Pham, C. (2008). Estimating the Gravity Model When Zero Trade Flows Are Important.

Martínez-Galán, E. and Fontoura, M. P. (2019). Global value chains and inward foreign direct investment in the 2000s. *World Economy*, 42(1):175–196.

Martínez-Martín, J. (2011). General equilibrium long-run determinants for Spanish FDI: A spatial panel data approach. *SERIEs*, 2(3):305–333.

Martínez-San Román, V., Bengoa, M., and Sánchez-Robles, B. (2016). Foreign direct investment, trade integration and the home bias: evidence from the European Union. *Empirical Economics*, 50(1):197–229.

Martínez-Zarzoso, I. (2013). The log of gravity revisited. *Applied Economics*, 45(3):311–327.

Martínez-Zarzoso, I., Felicitas, N., and Horsewood, N. (2009). Are regional trading agreements beneficial?: Static and dynamic panel gravity models. *The North American Journal of Economics and Finance*, 20(1):46–65.

Matyas, L. (2017). *The Econometrics of Multi-dimensional Panels*. Springer.

McCullagh, P. and Nelder, J. (1989). Generalized linear models. *2nd ed. (London: Chapman and Hall)*.

Medvedev, D. (2012). Beyond Trade: The Impact of Preferential Trade Agreements on FDI Inflows. *World Development*, 40(1):49–61.

Mitchell, J., Pain, N., and Riley, R. (2011). The drivers of international migration to the UK: A panel-based Bayesian model averaging approach. *Economic Journal*, 121(557):1398–1444.

Mitze, T., Alecke, B., and Untiedt, G. (2008). Trade, FDI and Cross-Variable Linkages: A German (Macro-) Regional Perspective. *Munich Personal RePEc Archive*, (12245).

Mitze, T., Alecke, B., and Untiedt, G. (2010). Trade-FDI linkages in a simultaneous equations system of gravity models for German regional data. *International Economics*, 122:121–162.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Neumayer, E. and Spess, L. (2005). Do bilateral investment treaties increase foreign direct investment to developing countries? *World Development*, 33(10):1567–1585.

Ng, A., Ibrahim, M. H., and Mirakhor, A. (2016). Does trust contribute to stock market development? *Economic Modelling*, 52:239–250.

OECD (2017). Germany: Trade and Investment Statistical Note. *OECD Publishing, Paris. Retrieved from www.oecd.org/investment/trade-investment-gvc.htm*.

Pelegrín, A. (2002). Inversión extranjera directa. Factores determinantes de la localización regional. *Papeles de Economía Española*, 93:122–133.

Pelegrín, A. (2003). Regional Distribution of Foreign Manufacturing Investment in Spain. Do agglomeration economies matter? *Institut d'Economia de Barcelona. Working paper 2003/6*.

Pelegrín, A. and Bolancé, C. (2008). Regional Foreign Direct Investment in Manufacturing. Do Agglomeration Economies Matter? *Regional Studies*, 42(4):505–522.

Petkova, K., Stasio, A., and Zagler, M. (2018). On the Relevance of Double Tax Treaties. *WU International Taxation Research Paper Series No. 2018-05. WU Vienna University of Economics and Business, Universität Wien, Vienna.*

Pfaffermayr, M. (2019). Gravity models, PPML estimation and the bias of the robust standard errors. *Applied Economics Letters*, pages 1–5.

Pham, T. H. H. (2017). Impacts of globalization on the informal sector: Empirical evidence from developing countries. *Economic Modelling*, 62:207–218.

Raftery, A. (1995). Bayesian model selection in social research. *Sociological methodology*, 25:111–163.

Ramondo, N. (2014). A quantitative approach to multinational production. *Journal of International Economics*, 93(1):108–122.

Ramondo, N. and Rodríguez-Clare, A. (2013). Trade, Multinational Production, and the Gains from Openness. *Journal of Political Economy*, 121(2):273–322.

Ramondo, N., Rodríguez-Clare, A., and Tintelnot, F. (2015). Multinational Production: Data and Stylized Facts. *American Economic Review. Papers and Proceedings*, 105(5):530–536.

Ramsey, J. (1969). Testing for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):350–371.

Razin, A. and Sadka, E. (2007). Productivity and Taxes as drivers of FDI. *NBER Working Paper No. 13094*.

Razin, A., Sadka, E., and Tong, H. (2008). Bilateral FDI flows: Threshold barriers and productivity shocks. *CESifo Economic Studies*, 54(3):451–470.

Roberts, H. (1965). Probabilistic Prediction. *Journal of the American Statistical Association*, 60(309):50–62.

Rodríguez, X. A. and Pallas, J. (2008). Determinants of foreign direct investment in Spain. *Applied Economics*, 40(19):2443–2450.

Russ, K. N. (2009). The New Theory of Foreign Direct Investment: Merging Trade and Capital Flows. *International Finance*, 12(1):107–119.

Russ, K. N. (2012). Exchange rate volatility and first-time entry by multinational firms. *Review of World Economics*, 148(2):269–295.

Sala-i Martin, X., Doppelhofer, G., and Miller, R. (2004). *Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estiamtes Approach,*, volume 94.

Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, 136(7 SPEC. ISS.):2144–2162.

Shapiro, S. and Wilk, M. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3):591–611.

Shin, H. (2019). What is behind the recent slowdown? *Presentation at the "Public Finance Dialogue" workshop arranged by German Federal Ministry of Finance and Centre for European Economic Research (ZEW).*

Siliverstovs, B. and Schumacher, D. (2009). Estimating gravity equations: To log or not to log? *Empirical Economics*, 36(3):645–669.

Silva, J. and Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Stein, E. and Daude, C. (2007). Longitude matters: Time zones and the location of foreign direct investment. *Journal of International Economics*, 71(1):96–112.

Talamo, G. (2007). Institution, FDI and the Gravity Model. *unpublished manuscript, Università di Palermo*.

UNCTAD (2018). UNCTAD Handbook of Statistics 2018, UN, New York, https://doi.org/10.18356/589b4ff4-en.

Venables, A. J. (1999). Fragmentation and multinational production. *European Economic Review*, 43(4-6):935–945.

Villaverde, J. and Maza, A. (2012). Foreign direct investment in Spain: Regional distribution and determinants. *International Business Review*, 21(4):722–733.

Wacker, K. M. (2013). On the measurement of FDI and its relationship to activities of multinational corporations. *ECB Working Paper No.1614*.

Wei, S.-J. (2000). How taxing is corruption on international investors? *The Review of Economics and Statistics*, 82(1):1–11.

Wei, Y. and Cao, Y. (2017). Forecasting house prices using dynamic model averaging approach: Evidence from China. *Economic Modelling*, 61:147–155.

Westerlund, J. and Wilhelmsson, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics*, 43(6):641–649.

Wheeler, D. and Mody, A. (1992). International investment location decisions. *Journal of International Economics*, 33(1-2):57–76.

World Bank Group, IDE-JETRO, OECD, UIBE, and World Trade Organization (2017). Global Value Chain Development Report 2017 : Measuring and Analyzing the Impact of GVCs on Economic Development. Washington, DC: World Bank.

Yang, Y. and Mallick, S. (2014). Explaining cross-country differences in exporting performance: The role of country-level macroeconomic environment. *International Business Review*, 23(1):246–259.