estructuraeconòmica
ECONOMÍA APLICADA II

ALTERNATIVE ESTIMATORS FOR THE FDI GRAVITY MODEL: AN APPLICATION TO GERMAN OUTWARD FDI

by

Mariam Camarero
Laura Montolio
and
Cecilio Tamarit

September 2019

Working Papers in Applied Economics

WPAE-2019-1907

ECONOMÍA APLICADA II

VNIVERSITAT DE VALÈNCIA

Departamento de Estructura Económica (Economía Aplicada II), Facultad de Economía, Universitat de València
Avda dels Tarongers, s/n, 46022 – Valencia, Spain
Phone +34 96 382 83 49 / Fax +34 96 382 83 54
http://www.estructuraeconomica.es
e-mail: investigacion.estructura@uv.es

ISSN 2172-3036

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

Mariam Camarero [†], Laura Montolio [‡] and Cecilio Tamarit [a§]

September 17, 2019

Abstract

Despite the sound theoretical foundations of FDI gravity models and its popularity in empirical studies, there is a lack of consensus regarding the econometric specification and the estimation of the gravity equation. This paper provides a comprehensive empirical evidence of the determinants of German outward FDI comparing several estimation methods in their multiplicative form. We use four versions of the Generalized Linear Model (GLM), namely, Poisson Pseudo Maximum Likelihood(PPML), Gamma Pseudo Maximum Likelihood (GPML), Negative Binomal Pseudo Maximum Likelihood (NBPML) and Gaussian-GLM. The results of the empirical application indicate that NBPML is the best performing estimator followed by GPML.

Keywords: FDI determinants; Outward Foreign Direct Investment; Germany; Gravity models; Generalized linear models
JEL classification: F21, F23, C13, C33

*The authors are grateful to Astrit Sulstarova for kindly providing us with the extended version of FDI stock data from the UNCTAD's Bilateral FDI Statistics database as well as to Stephan Klasen, Inma Martínez-Zarzoso and Felicitas Nowak-Lehmann for their advice and hospitality during a stay at the Department of Economics of the University of Goettingen. We also acknowledge the financial support from the Spanish Ministry of Science, Innovation and Universities (BES-2015-072395 and ECO2014-58991-C3-2-R and ECO2017-83255-C3-3-P projects) and the European Social Fund, as well as the Valencian regional government (Generalitat Valenciana-PROMETEO 2018/102 project). All remaining errors are ours.

[†]University Jaume I and INTECO, Department of Economics, Campus de Riu Sec, E-12080 Castellón (Spain). ORCID: 0000-0003-4525-5181

[‡]University of València and INTECO, Department of Applied Economics II, Av. dels Tarongers, s/n Eastern Department Building E-46022 Valencia, Spain

[§a]*Corresponding author:* University of València and INTECO, Department of Applied Economics II, Av. dels Tarongers, s/n Eastern Department Building E-46022 Valencia, Spain. email: `Cecilio.Tamarit@uv.es` ORCID: 0000-0002-0538-9882

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 (2006, 2011) 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 (2011), 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.

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:

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

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 1 can be log-linearized as

$$ln(AS_{ij}) = \alpha_1 + \zeta_1 ln(s_i) - \beta_1 ln(D_{ij}) + \xi_i ln(m_j) \tag{2}$$

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.2.1 we present estimation methods that estimate the gravity model in its additive form (i.e. log-log form), while in Sect.2.2 we report estimation methods estimating the gravity model in its multiplicative form.

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), Martinez-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)).

2.2. Multiplicative functional form estimators

Based on Jensen's inequality, that states that $E[ln(\epsilon_{ij})] \neq [lnE(\epsilon_{ij})]$, Silva and Tenreyro (2006) demonstrate that the OLS estimation of models, like the gravity model in Equation 2, in its additive functional form provides inconsistent estimates. This is because, under heteroskedastic data, the expected values of the log-linearized error term ($E[ln(\epsilon_{ij})]$) 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 (2011) 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 (2011), Head and Mayer (2014)).

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 (2011), 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:

$$y_i = exp(x_i\beta_i)\epsilon_i \tag{3}$$

where $\mathbb{E}(\epsilon_i|x) = 1$, 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 (Nelder and Wedderburn (1972); McCullagh and Nelder (1989)). 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 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 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.

3.1. Poisson pseudo maximum likelihood

Considering the gravity model in Equation 3, the PPML estimator computes β, the vector of parameters of interest, from the following first-order conditions:

$$\sum_{i=1}^{n}[y_i - exp(x_i\tilde{\beta})]x_i = 0 \tag{4}$$

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.

3.2. Gamma pseudo maximum likelihood

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

$$\sum_{i=1}^{n}[y_i - exp(x_i\check{\beta})]exp(-x_i\check{\beta})x_i = 0 \tag{5}$$

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 (2011), among others). GPML has also been widely used to address the zero-valued observations problem.

[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.

3.3. Negative binomial pseudo maximum likelihood

The NBPML estimator is defined by:

$$Pr[I_{ij}] = \frac{\Gamma(I_{ij} + \alpha^{-1})}{I_{ij}!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \left(\frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}} \right)^{I_{ij}} \tag{6}$$

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.

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.

3.4. Gaussian GLM

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

$$\sum_{i=1}^{n} [y_i - exp(x_i\check{\beta})]exp(x_i\check{\beta})x_i = 0 \tag{7}$$

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).

3.5. Data

The analysis makes use of the data set described in Camarero et al. (2019), which provides information on German outward FDI stock over the period 1996-2012 in 59 destination countries (38 developed and 21 developing). A detailed description of variable definitions and data sources is presented in Table A.1, while Table A.2 reports the 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.

Explanatory variables have been selected according to the results provided in Camarero et al. (2019). Camarero et al. (2019) apply a BMA analysis to identify the robust determinants of German outward FDI. To address the heterogeneity in the destinations, they conduct the analysis across different recipient-country groups: developed, developing, Latin American, Asian, EU core and EU peripheral. Their results provide the inclusion probabilities (PIP) of each potential FDI determinants; those variables with a PIP above 0.50 are considered robust determinants. A summary of their findings is reported in Table A.3. A drawback of this methodology is that it only computes model averaged estimated

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

coefficients. Thereby, this study aims to provide a more accurate estimation of the coefficients. In doing so, we depart from the variables obtained by the BMA analysis conducted in Camarero et al. (2019) and compare alternative GLM estimators applying a backward elimination (BE) procedure.[3]

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 (2011), 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 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 (2011).

Tables 2 to 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 1 to 7. 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

[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 estimators in order to gain further insights on the adequacy of the variance function in Figures 8 to 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]

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

[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 $(\beta * 100)$.

experiment and concludes that NBPML appears to be the best estimator for their application. Likewise, Martínez-Zarzoso (2011) 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 the study of Camarero et al. (2019) in which they proposed to tackle the heterogenity in FDI destinations by disaggregating the sample in different country-groups, we decide to present our results in two subsections. In Subsection 4.1, we report the results for the whole dataset. Then, in Subsection 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*).

4.1. Worldwide German FDI determinants

Table 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 Camarero et al. (2019). 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, as has been argued by Camarero et al. (2019), 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.

4.2. Region-specific German FDI determinants

4.2.1. German FDI in developed countries

Table 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)).

4.2.2. German FDI in developing countries

For developing countries, in turn, the estimated coefficients are reported in Table 4. Out of the 22 variables singled out by the BMA analysis in Camarero et al. (2019), 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 (*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 Camarero et al. (2019), 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

[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))

[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.

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.

4.2.3. German FDI in Latin American countries

Table 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.

4.2.4. German FDI in Asian countries

For Asian countries, in turn, the estimated coefficients are reported in Table 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.

4.2.5. German FDI in core EU countries

Table 7 reports the estimated coefficients for core EU countries. In this case, 6 out of the 9 variables identified by the BMA analysis in Camarero et al. (2019) 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.

4.2.6. German FDI in peripheral EU countries

For peripheral EU countries, in turn, the estimated coefficients are reported in Table 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 Pfaffermayr (2004b); Egger and Merlo (2007)).

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.

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 Camarero et al. (2019). Tables B.1 to 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 15 to 21. Whereas Figures 22 to 28 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: 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 Camarero et al. (2019), the sign of the FDI determinants identified remains very stable across samples and thus, confirm the BMA results. Bear in mind that Camarero et al. (2019) estimate 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.

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 Camarero et al. (2019) 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.

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.

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.

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), 608â42.

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.

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.

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.

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.

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. 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.

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.

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.

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.

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., Montolio, L., and Tamarit, C. (2019). What drives German Foreign Direct Investment? New evidence using Bayesian statistical techniques. *Economic Modelling, (forthcoming)*.

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

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.

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*.

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.

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

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

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

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.

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

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

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

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

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

Martin, W. and Pham, C. (2008). Estimating the gravity equation when zero trade flows are frequent.

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-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. (2011). The log of gravity revisited. *Applied Economics*, 45(3):311–327.

Martinez-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. (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.

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

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.

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.

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

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

Table 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 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.

Table 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 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.

Table 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: 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.

Table 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)
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.

Table 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.

Table 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.

Table 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 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.

Figure 1: GLMs estimators for all host countries: Predictions and Pearson residuals.

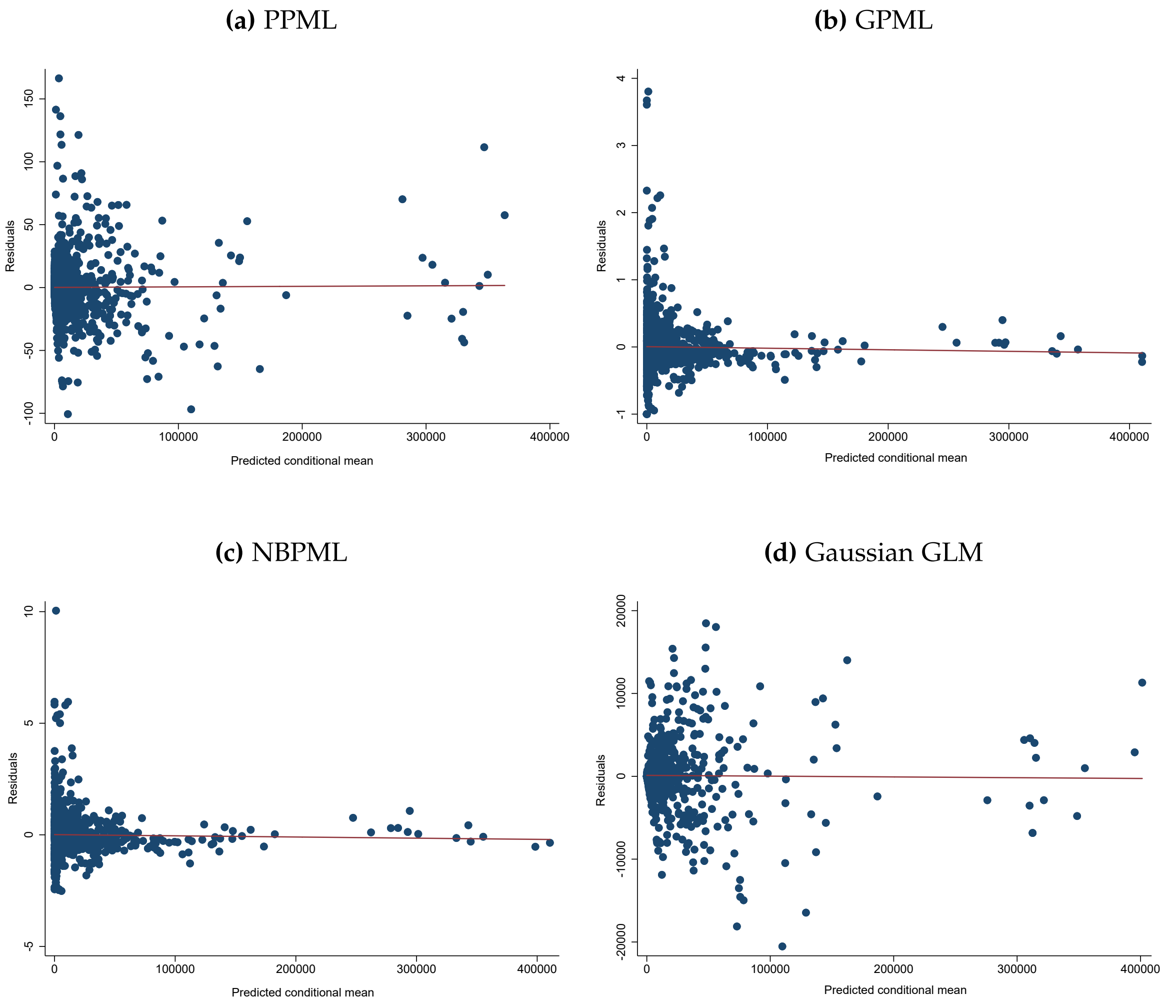
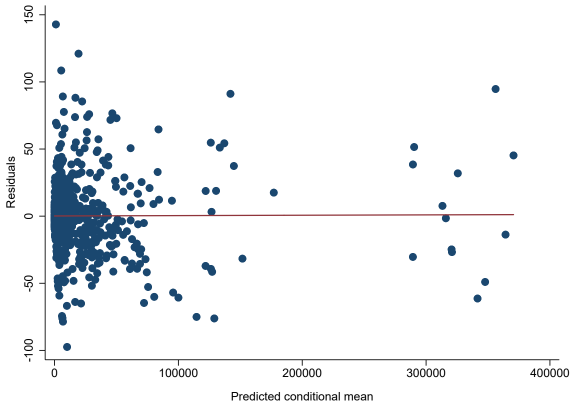
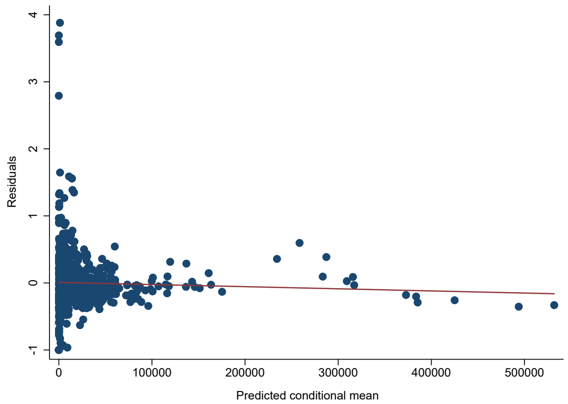


Figure 2: GLMs estimators for developed countries: Predictions and Pearson residuals.

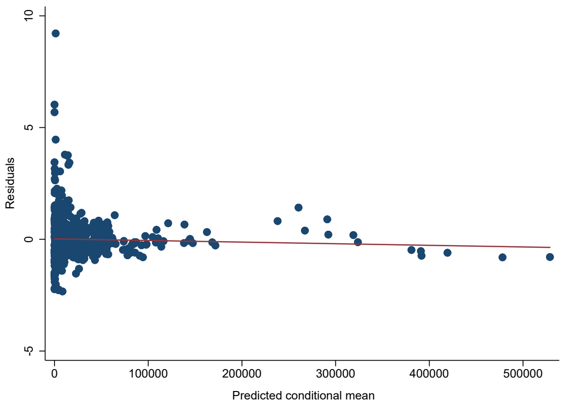
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

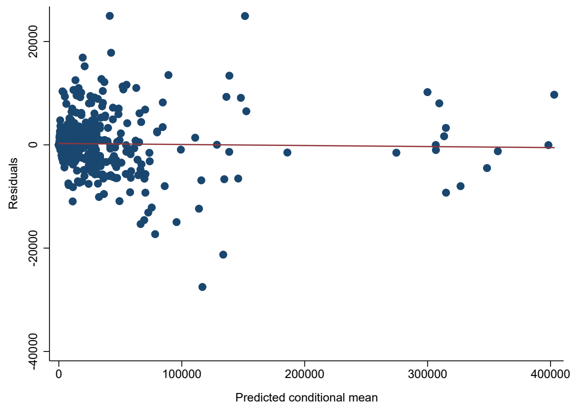
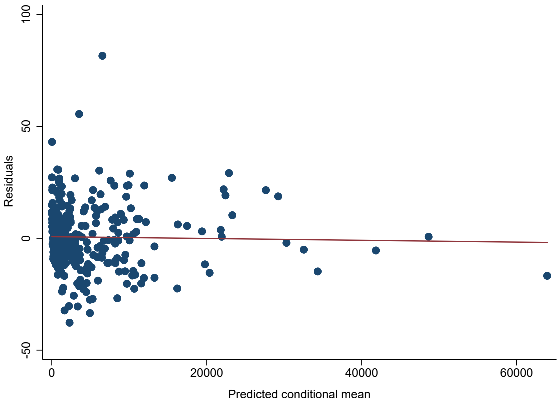
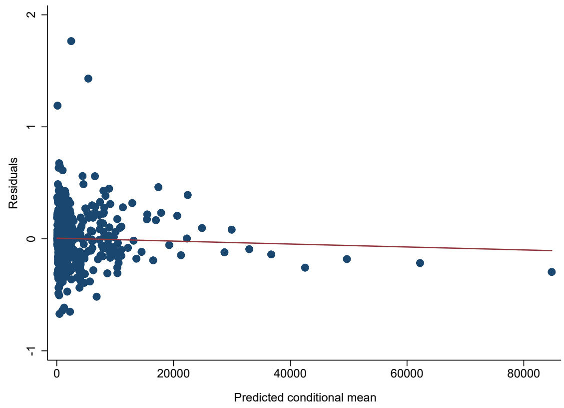


Figure 3: GLMs estimators for developing countries: Predictions and Pearson residuals.

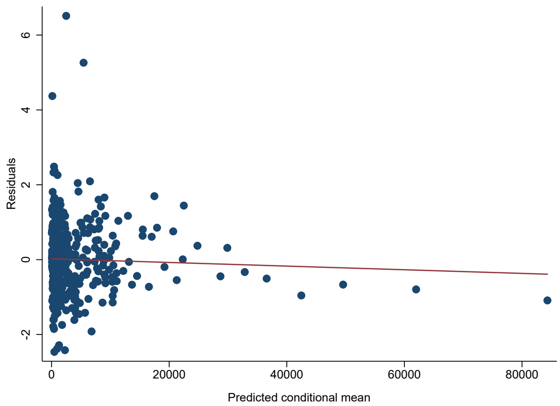
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

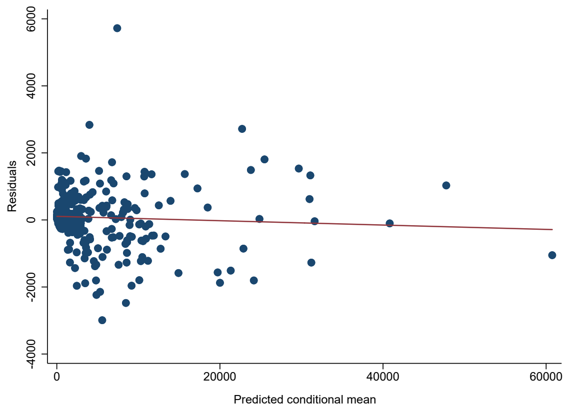
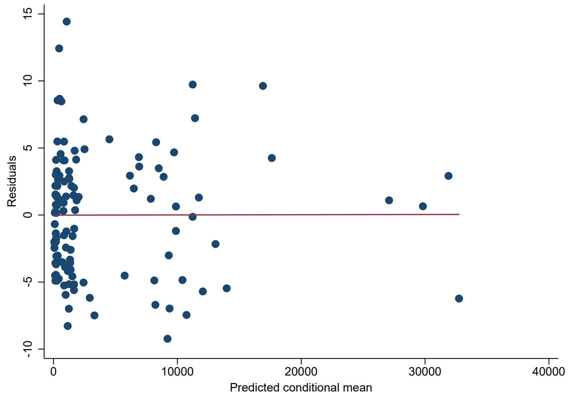
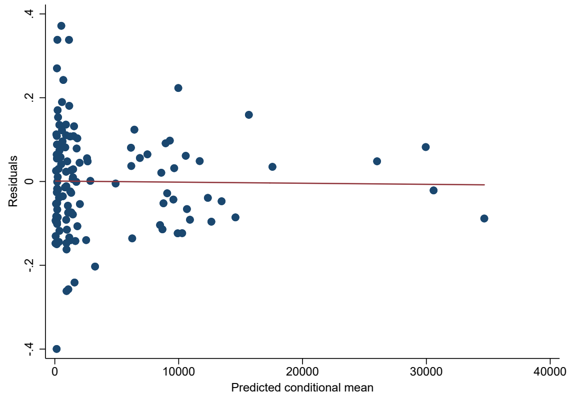


Figure 4: GLMs estimators for Latin American countries: Predictions and Pearson residuals.

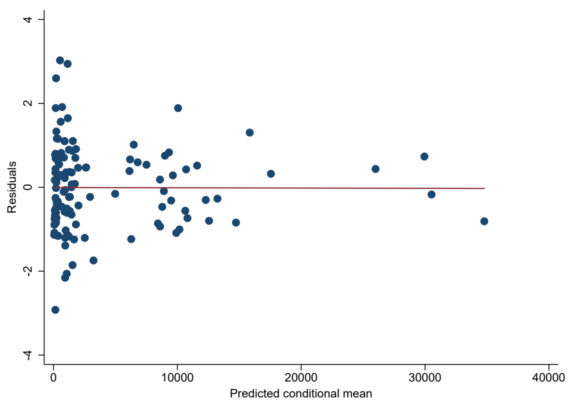
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

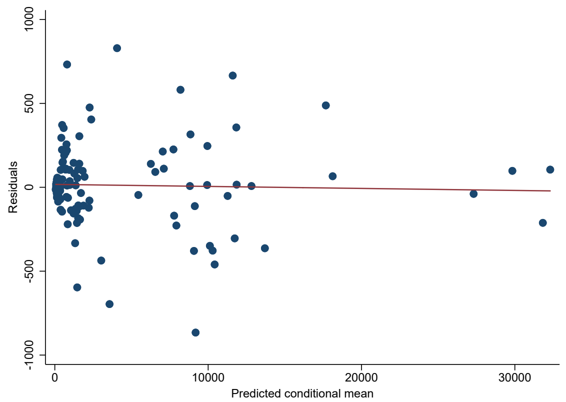


Figure 5: GLMs estimators for Asian countries: Predictions and Pearson residuals.

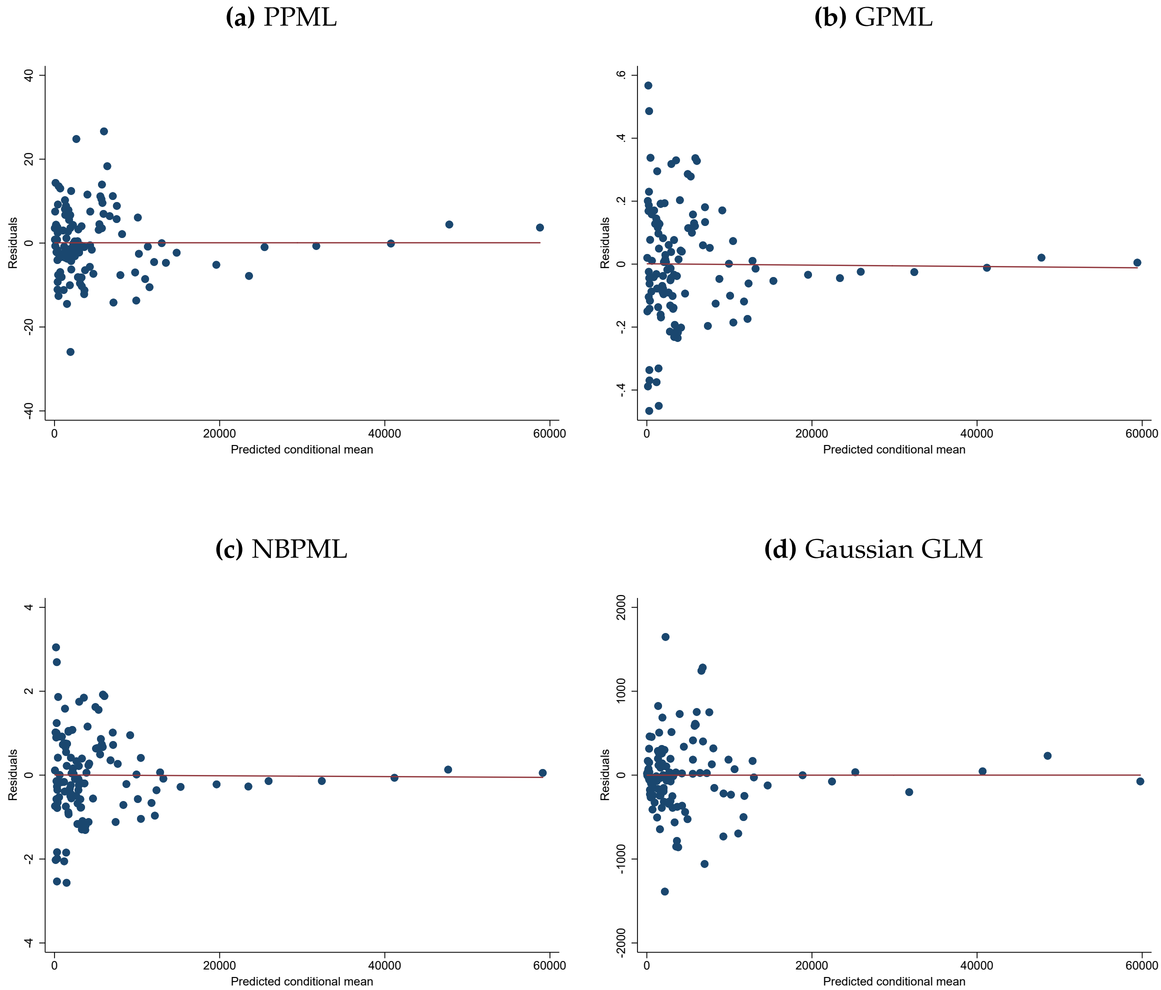
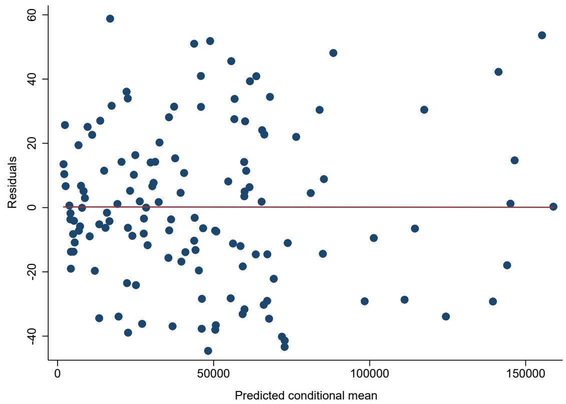
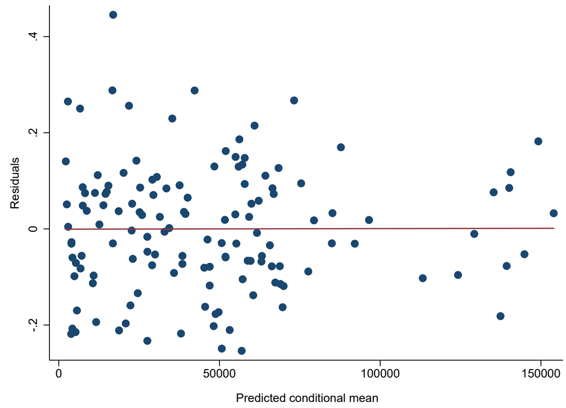


Figure 6: GLMs estimators for Core EU countries: Predictions and Pearson residuals.

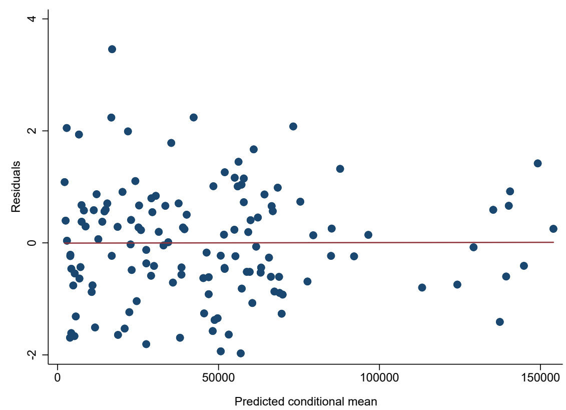
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

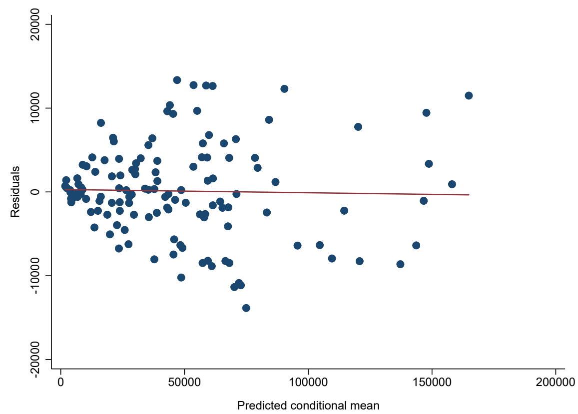
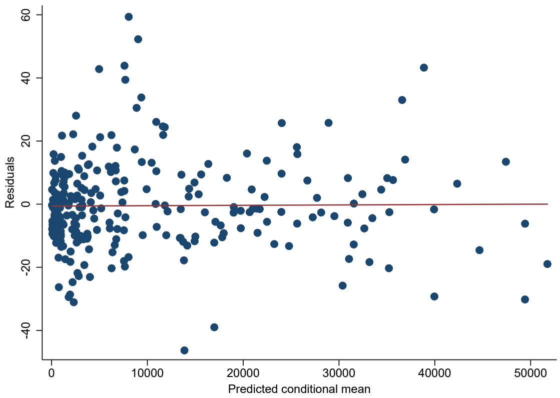
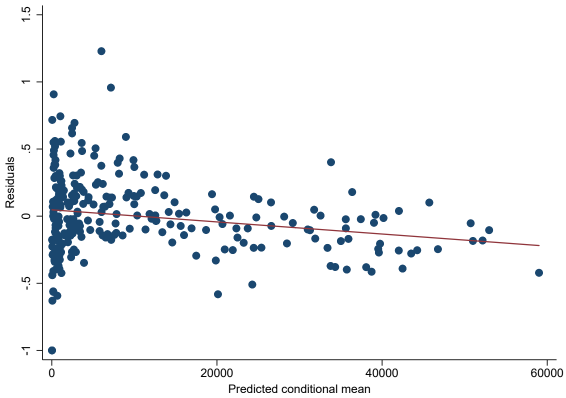


Figure 7: GLMs estimators for Peripheral EU countries: Predictions and Pearson residuals.

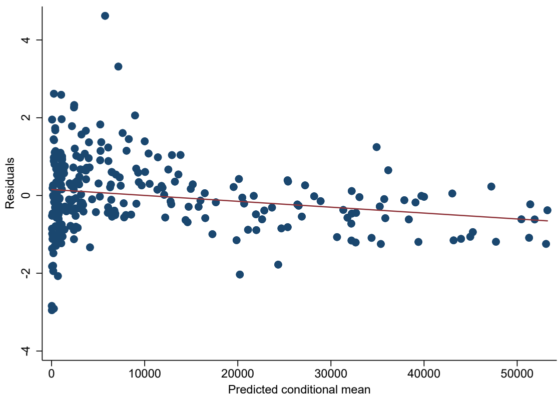
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

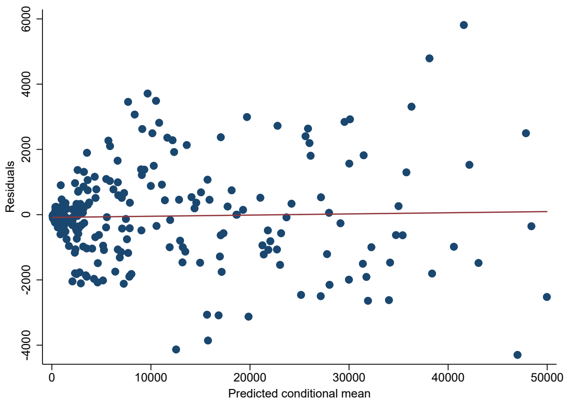
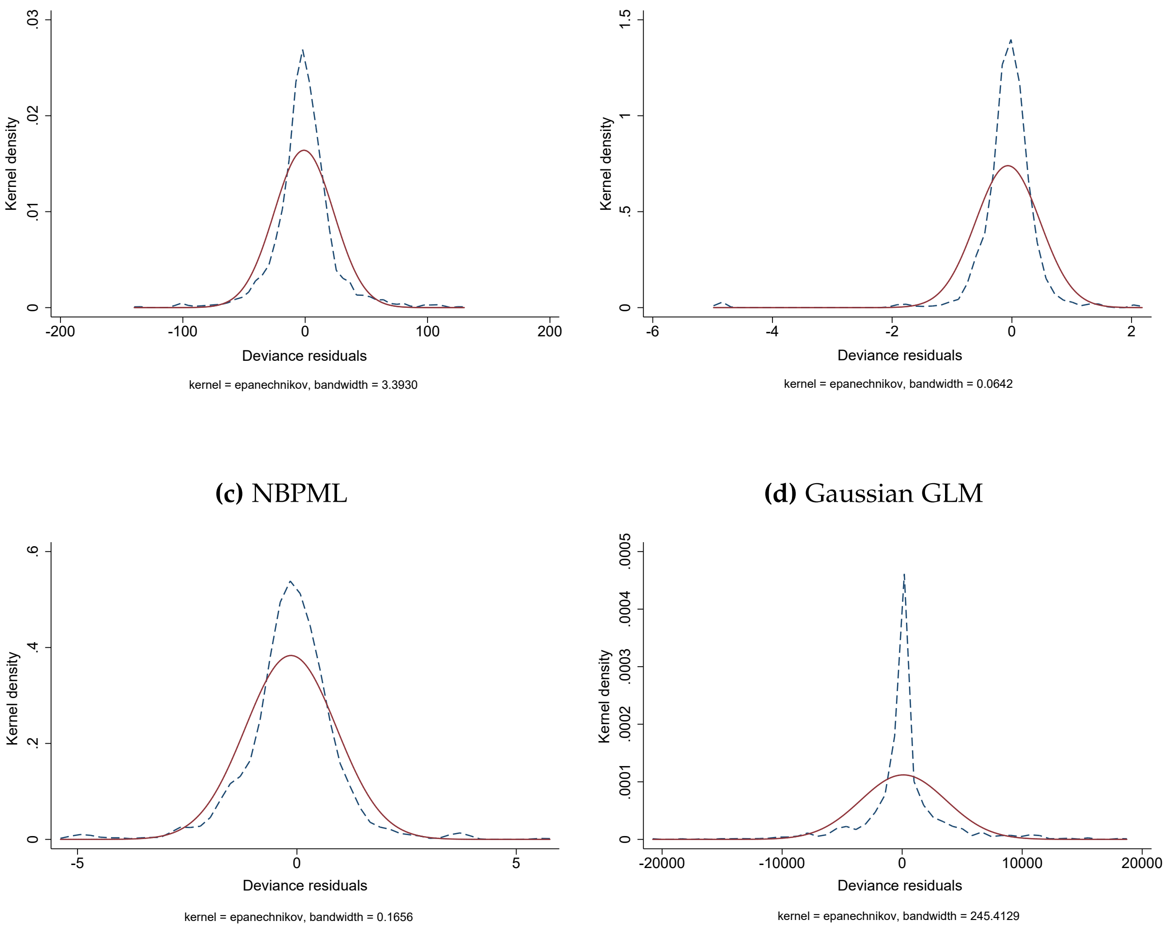


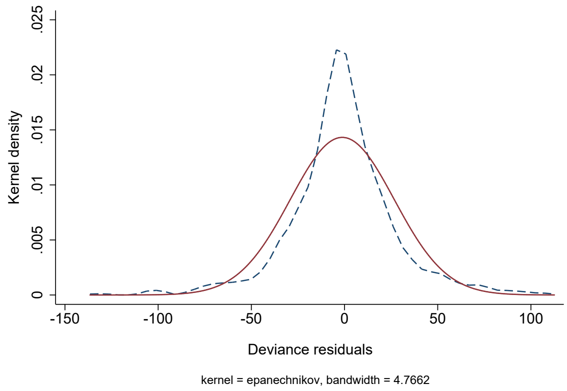
Figure 8: GLMs estimators for all host countries: Density of deviance residuals.



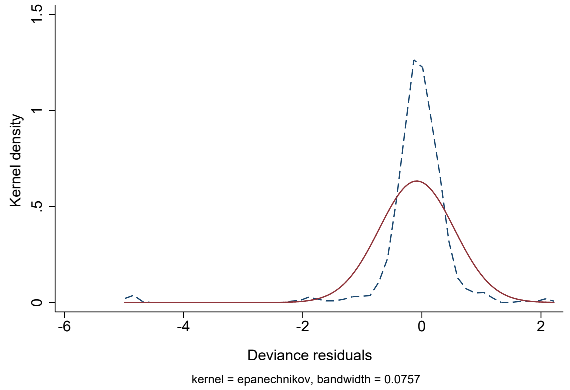
(a) PPML **(b)** GPML **(c)** NBPML **(d)** Gaussian GLM

Figure 9: GLMs estimators for developed countries: Density of deviance residuals.

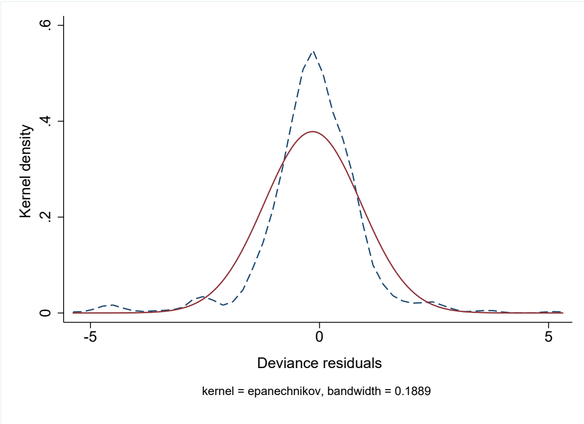
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

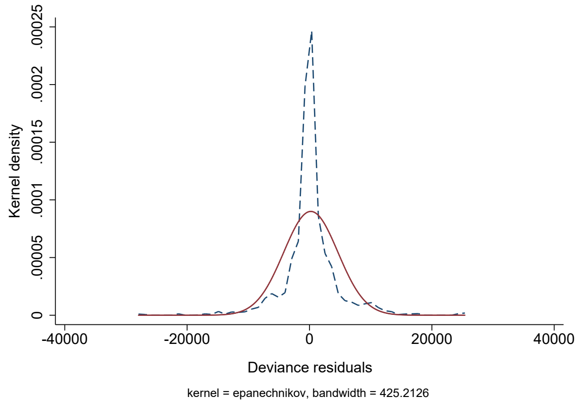
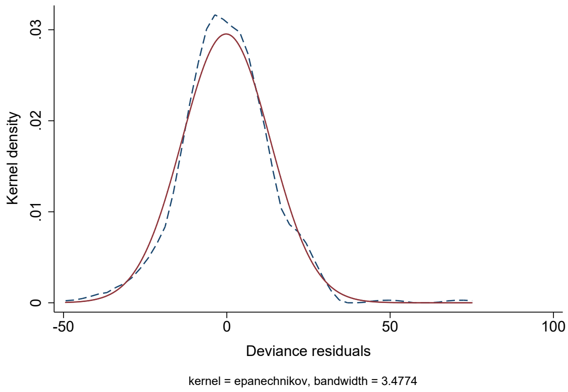
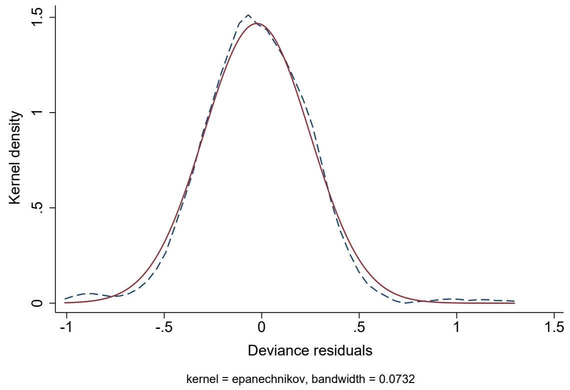


Figure 10: GLMs estimators for developing countries: Density of deviance residuals.

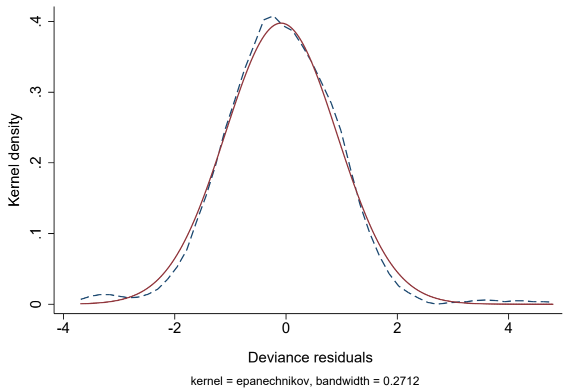
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

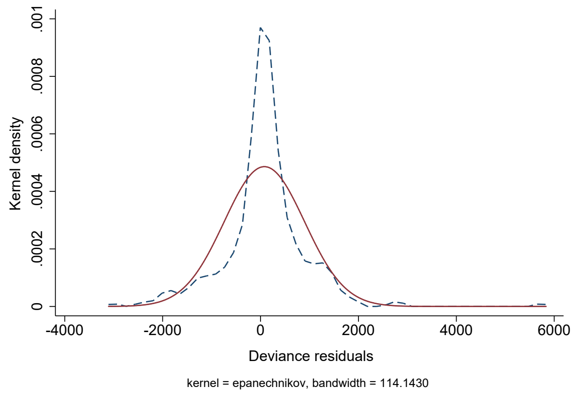
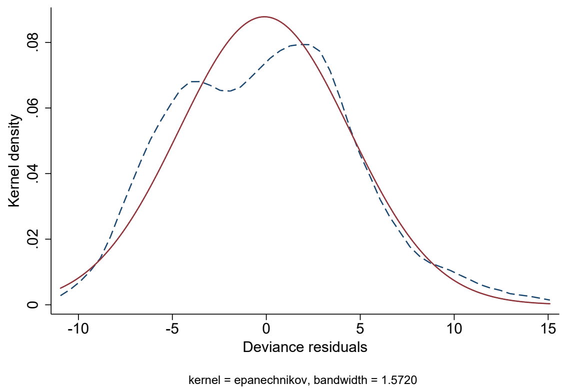
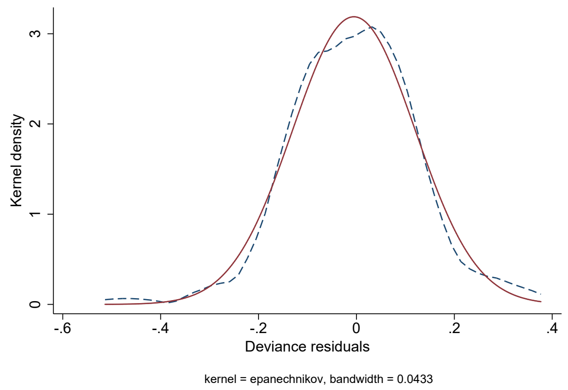


Figure 11: GLMs estimators for Latin American countries: Density of deviance residuals.

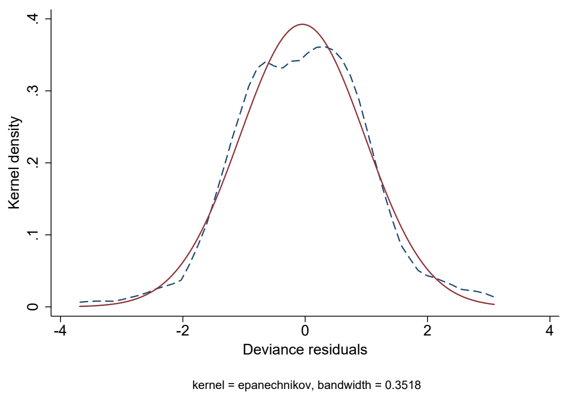
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

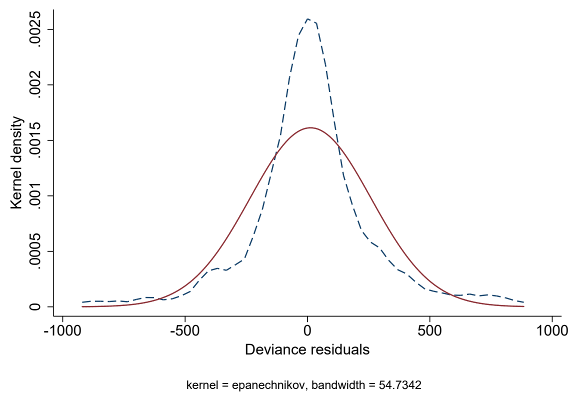
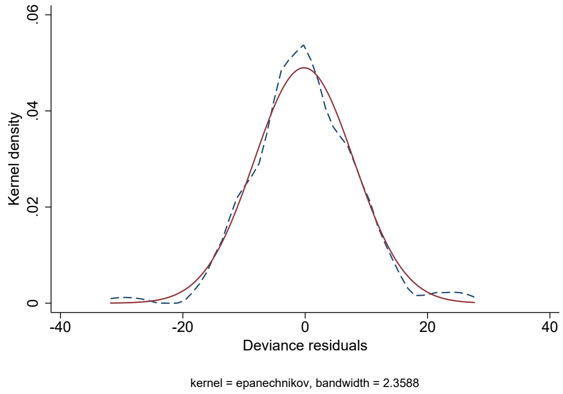
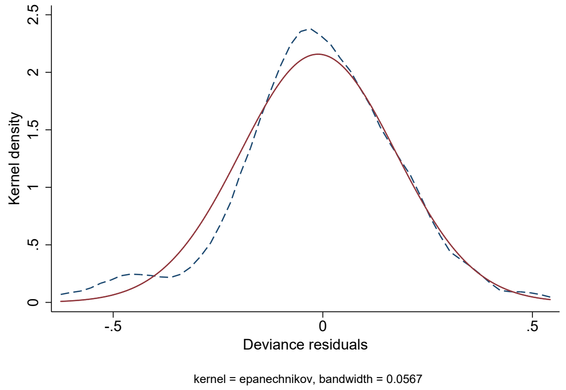


Figure 12: GLMs estimators for Asian countries: Density of deviance residuals.

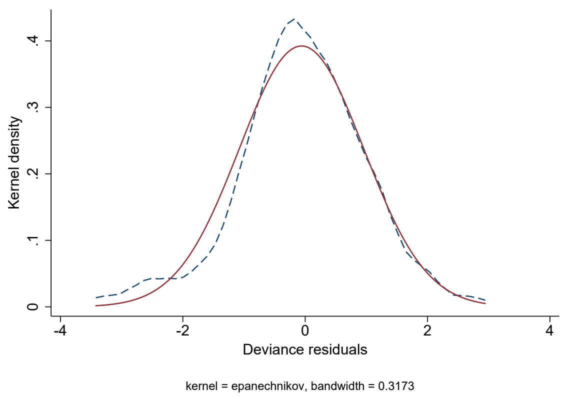
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

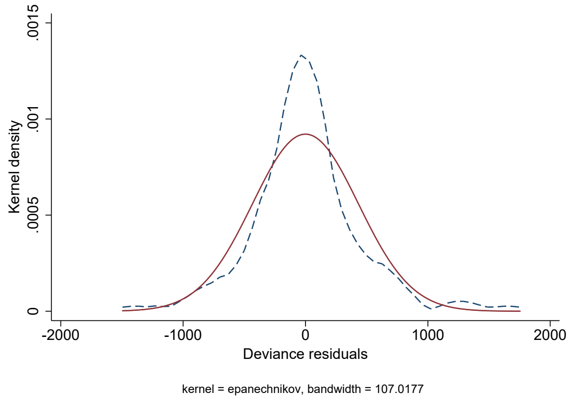


Figure 13: GLMs estimators for Core EU countries: Density of deviance residuals.

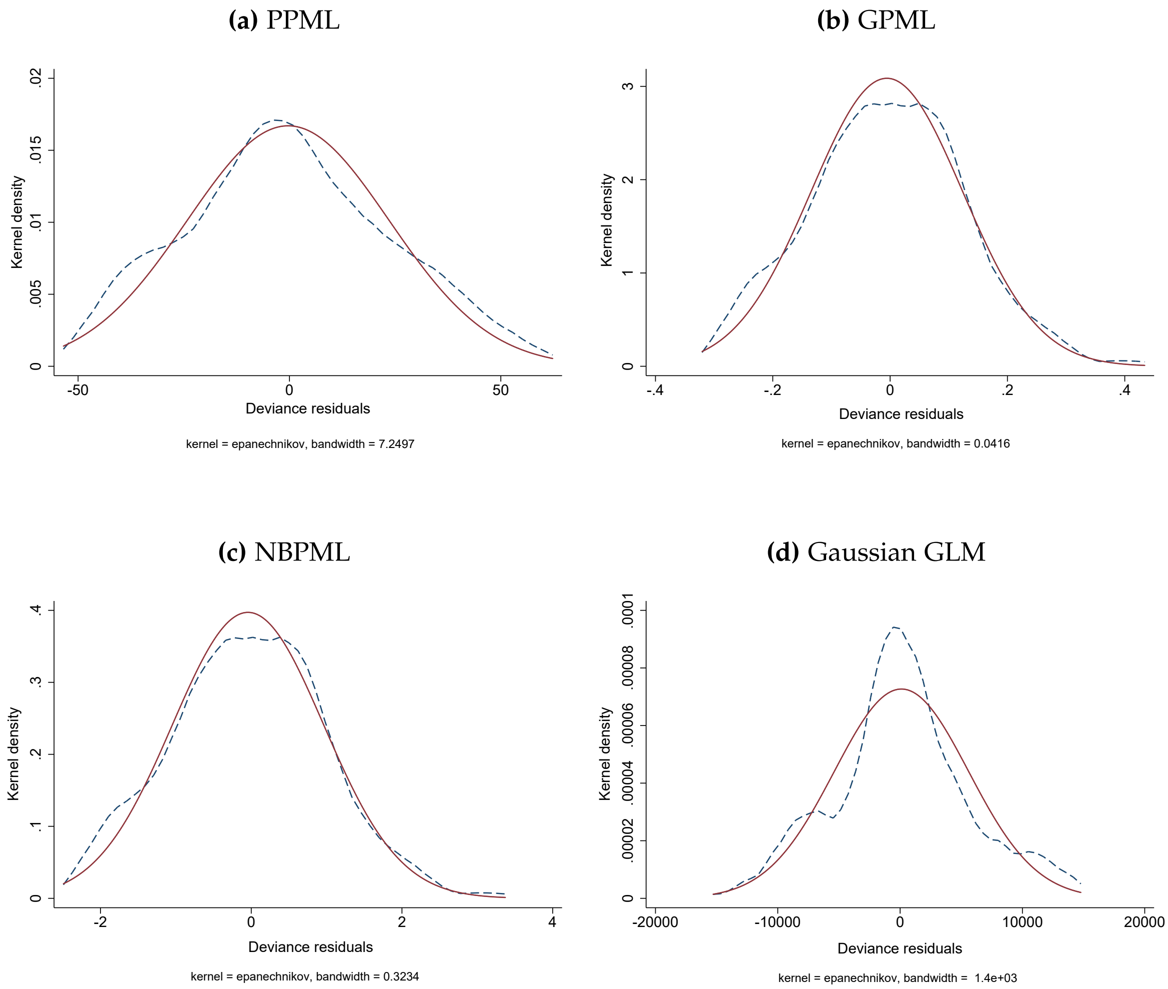
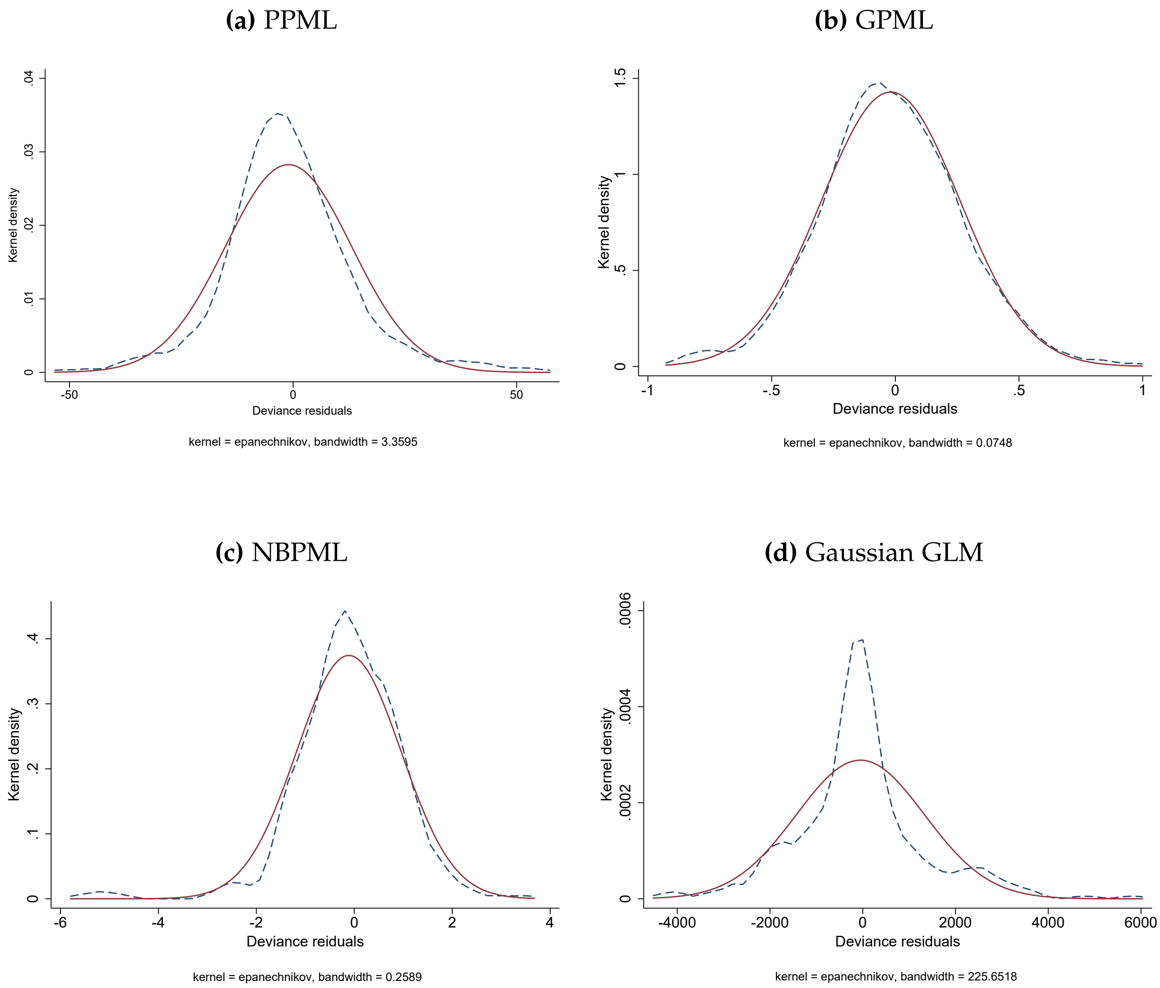


Figure 14: GLMs estimators for Peripheral EU countries: Density of deviance residuals.



(a) PPML

(b) GPML

(c) NBPML

(d) Gaussian GLM

APPENDIX A

Table 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 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 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 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.

Table A.3: 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 A.3: 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 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 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 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 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 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 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 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 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 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 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 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)

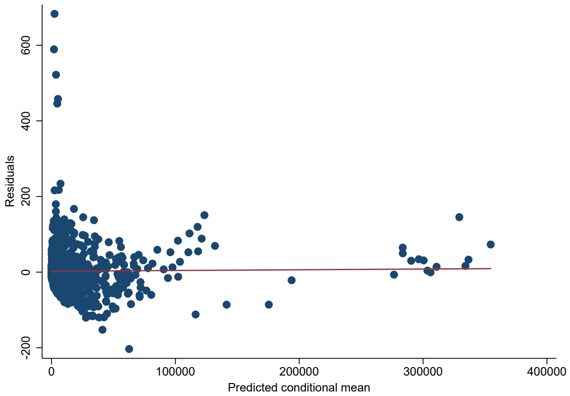
Table B.7: Determinants of FDI in peripheral EU countries, 1996-2012. *(Continued)*

	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

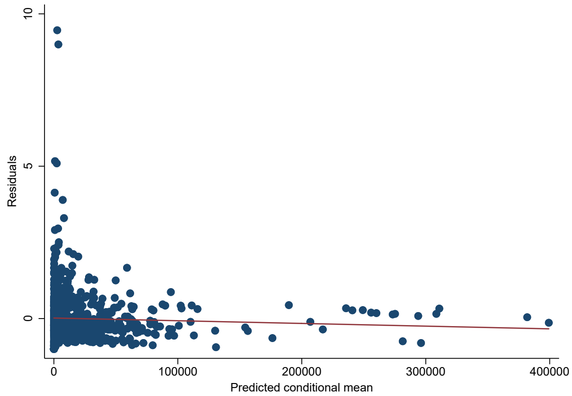
Notes: Country pair clustered standard errors are in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% significance levels, respectively. Posterior inclusion probabilities larger than 0.5.

Figure 15: Robustness checks: GLMs estimators for all host countries: Predictions and Pearson residuals.

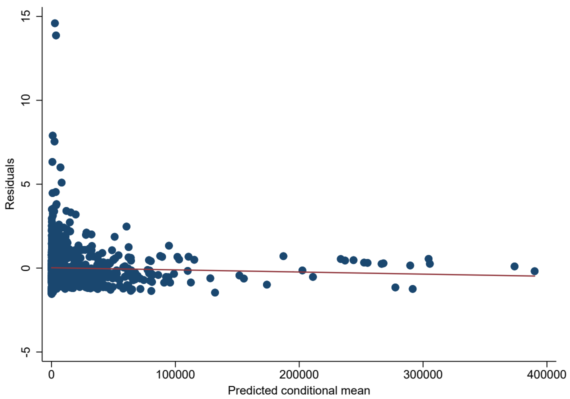
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

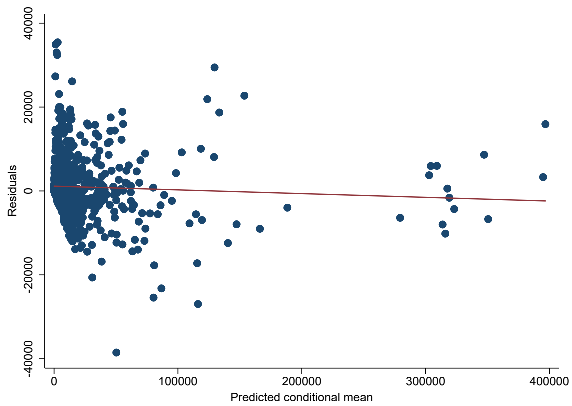
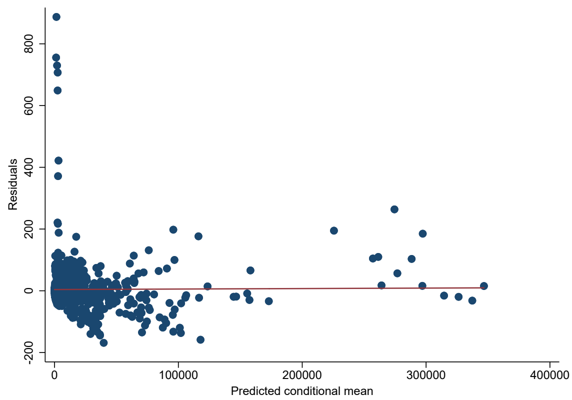
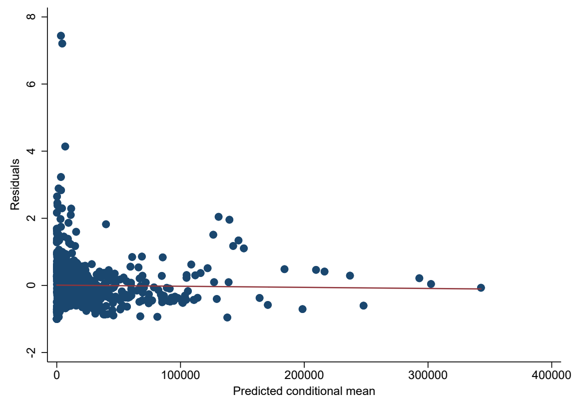


Figure 16: Robustness checks: GLMs estimators for developed countries: Predictions and Pearson residuals.

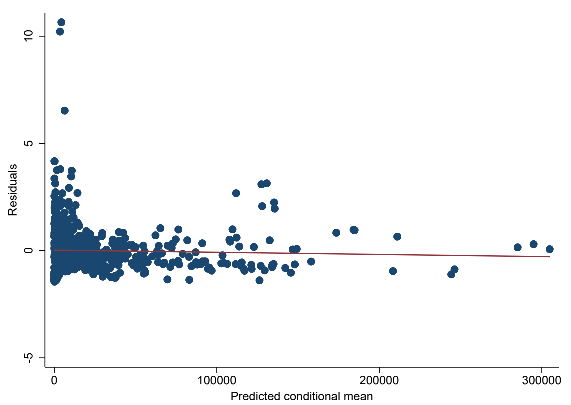
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

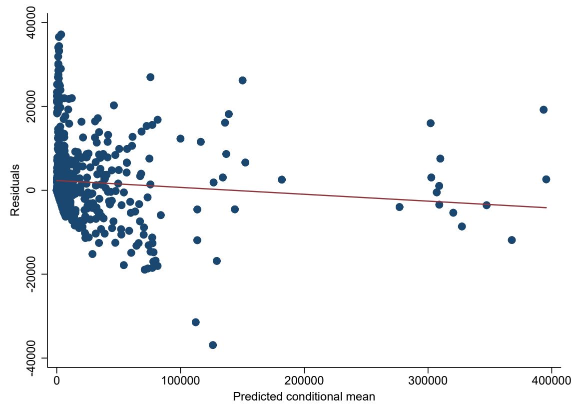
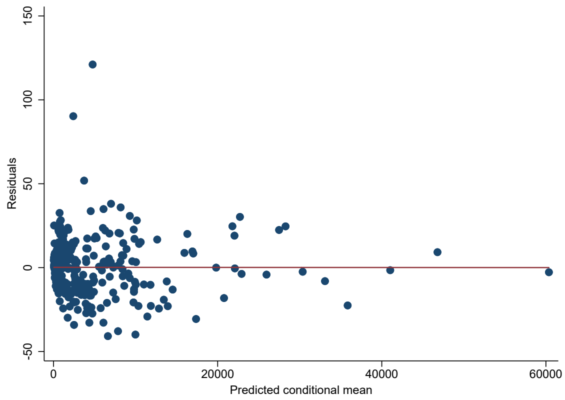
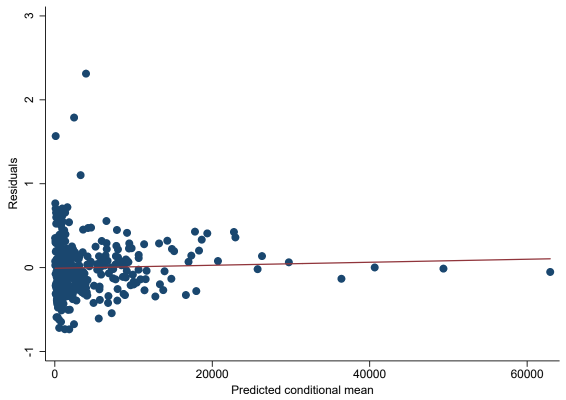


Figure 17: Robustness checks: GLMs estimators for developing countries: Predictions and Pearson residuals.

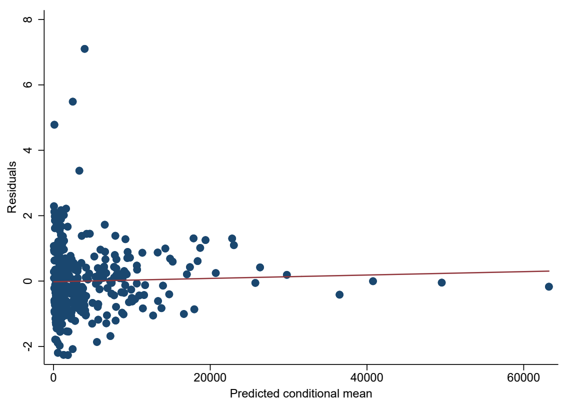
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

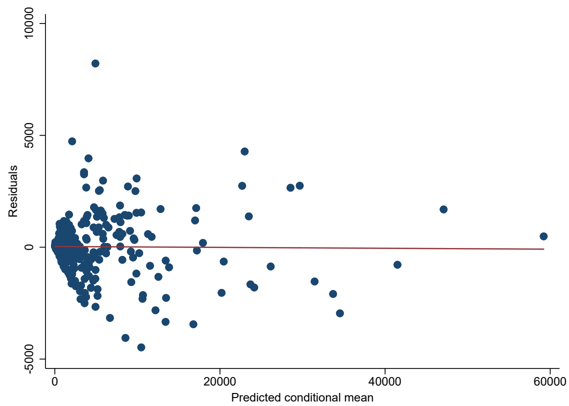
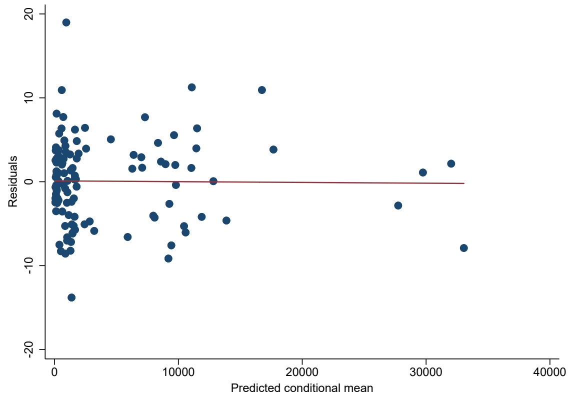
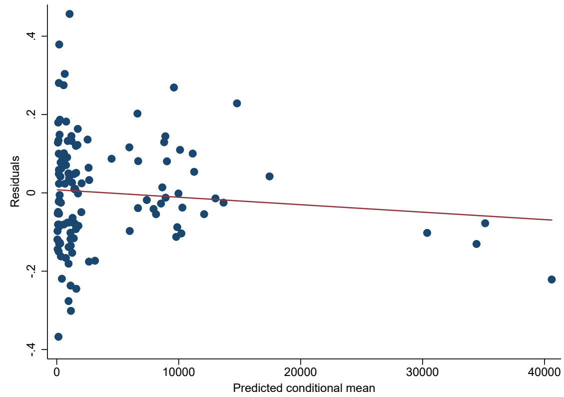


Figure 18: Robustness checks: GLMs estimators for Latin American countries: Predictions and Pearson residuals.

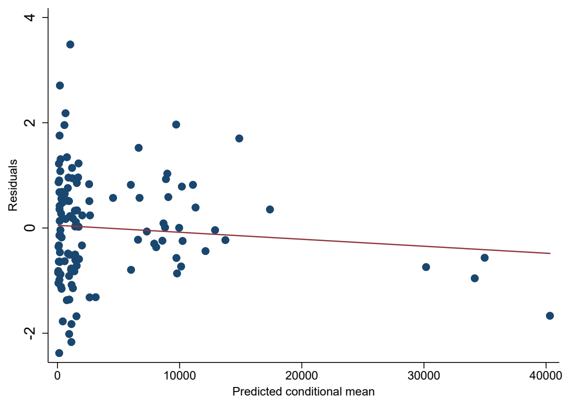
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

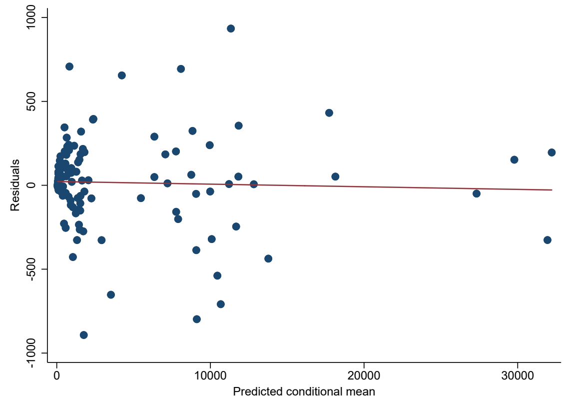
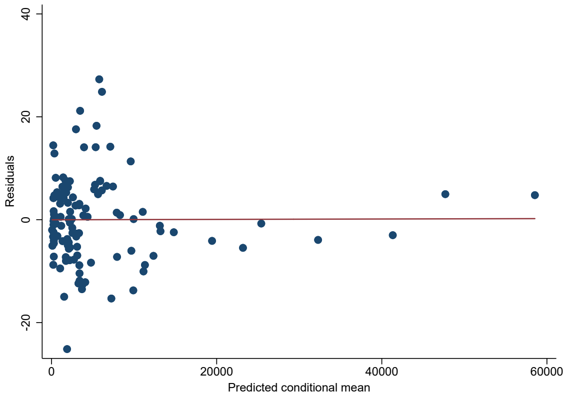
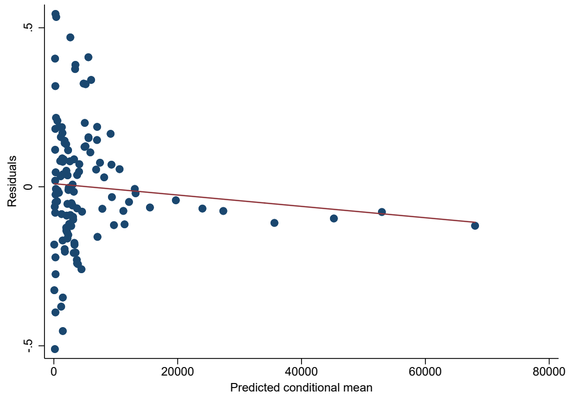


Figure 19: Robustness checks: GLMs estimators for Asian countries: Predictions and Pearson residuals.

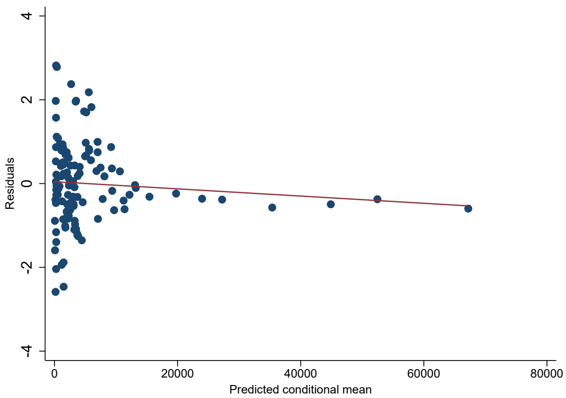
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

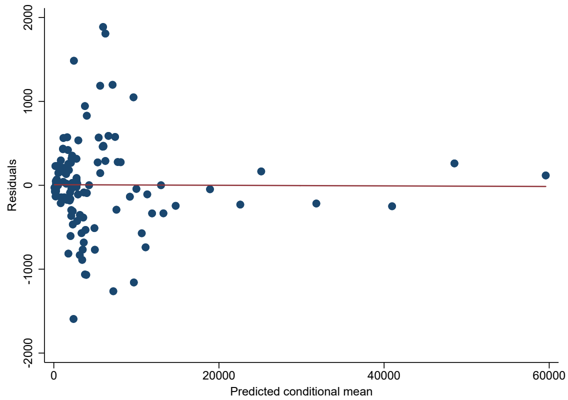
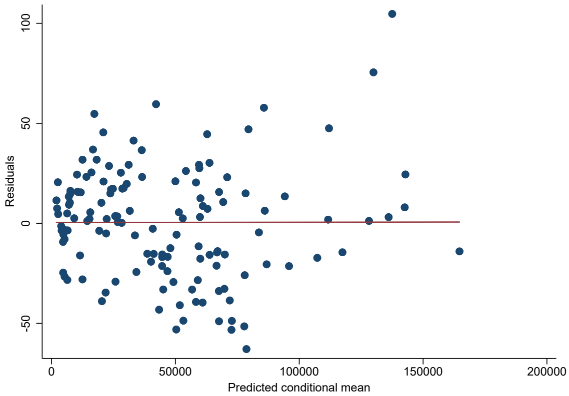
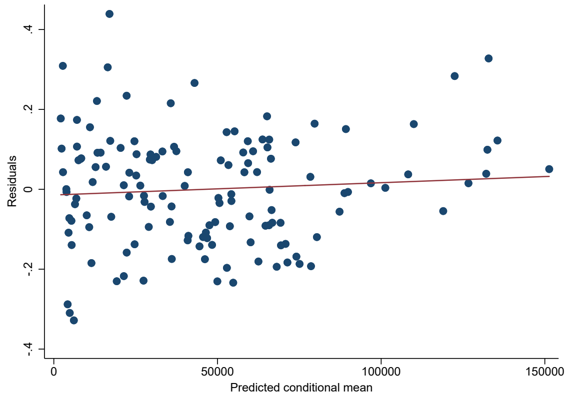


Figure 20: Robustness checks: GLMs estimators for Core EU countries: Predictions and Pearson residuals.

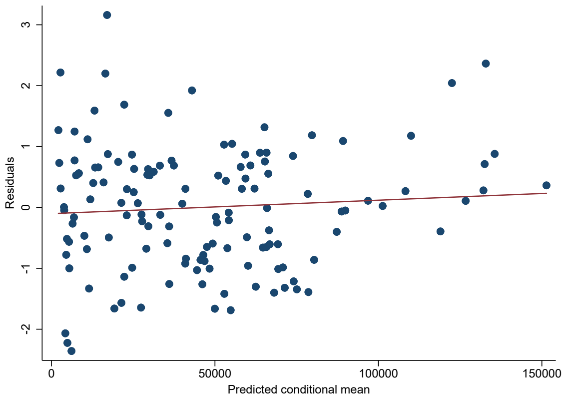
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

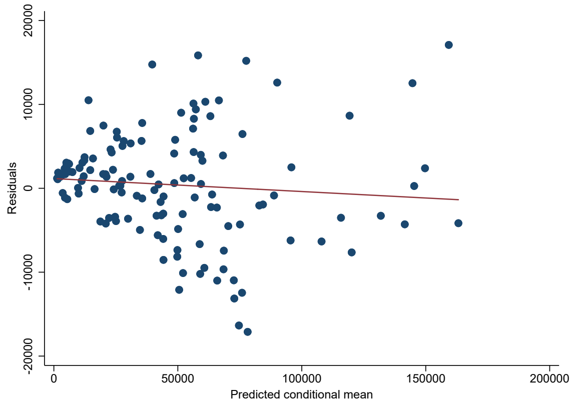
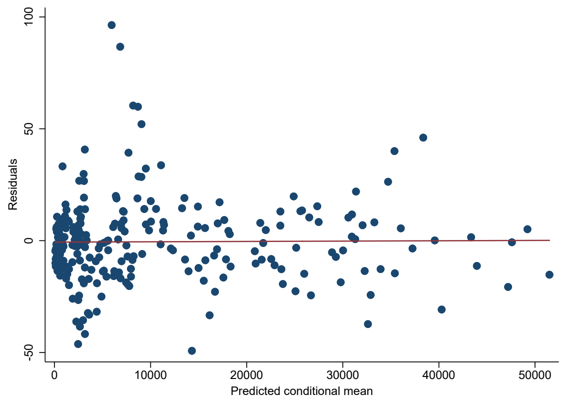
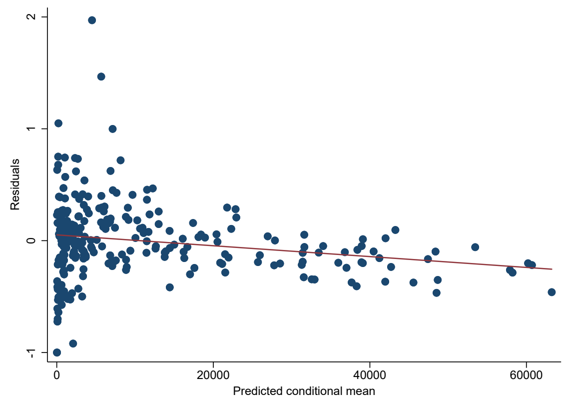


Figure 21: Robustness checks: GLMs estimators for Peripheral EU countries: Predictions and Pearson residuals.

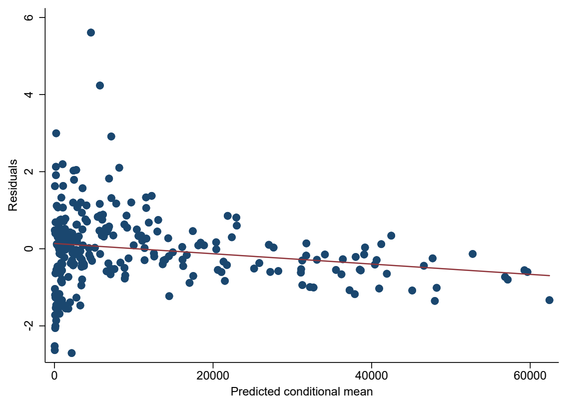
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

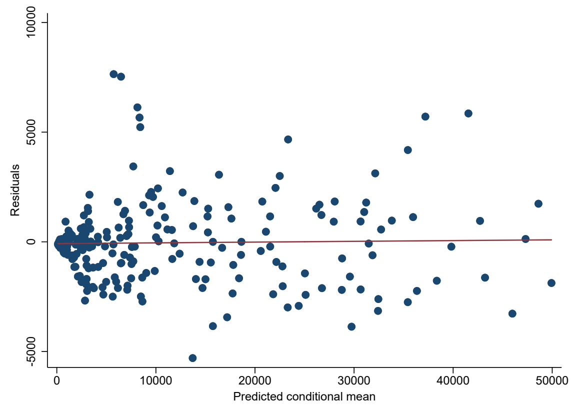


Figure 22: Robustness checks: GLMs estimators for all host countries: Density of deviance residuals.

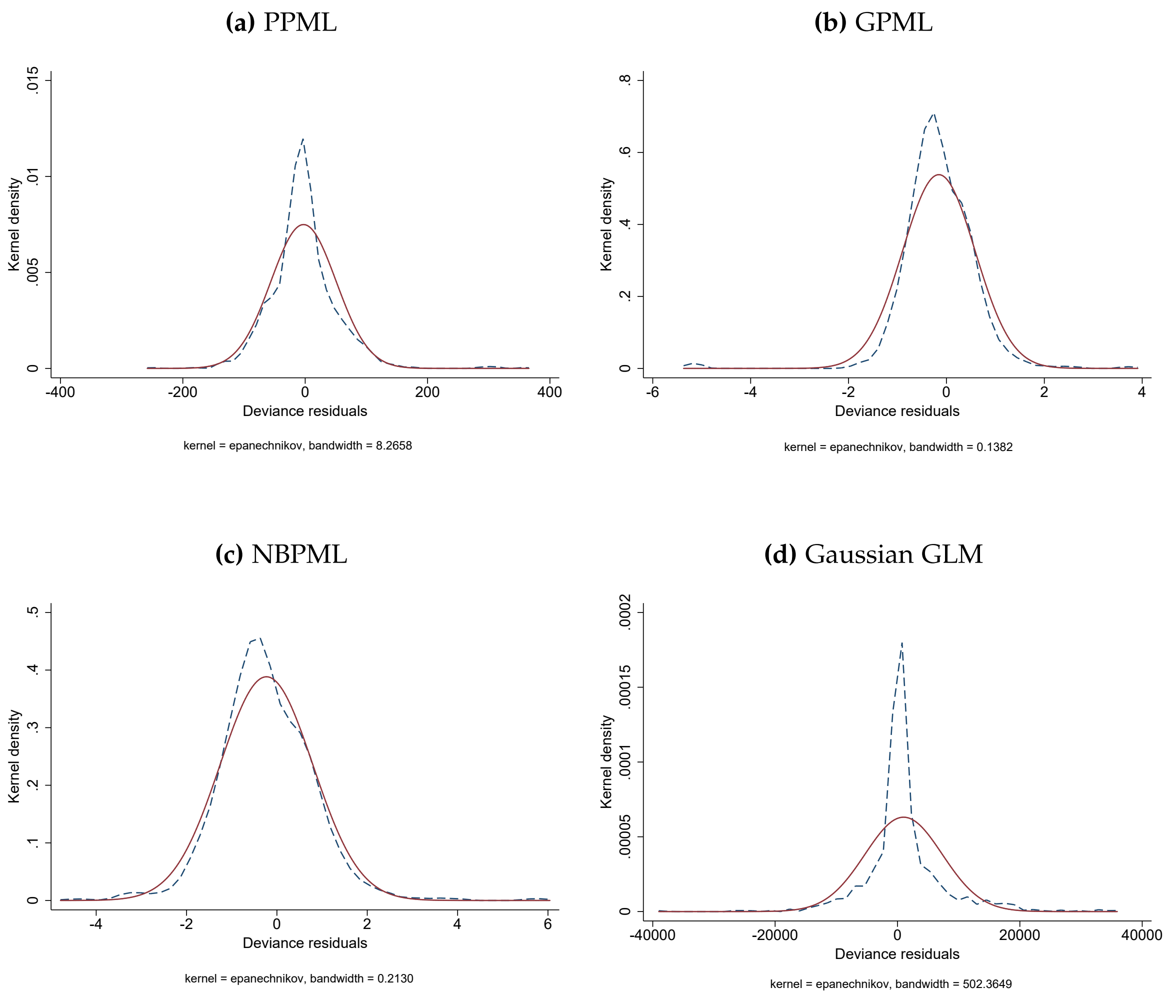
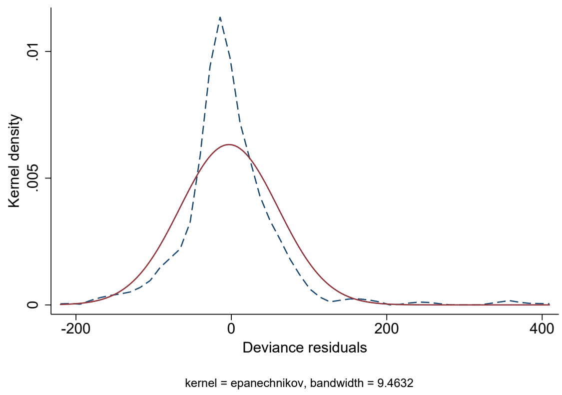
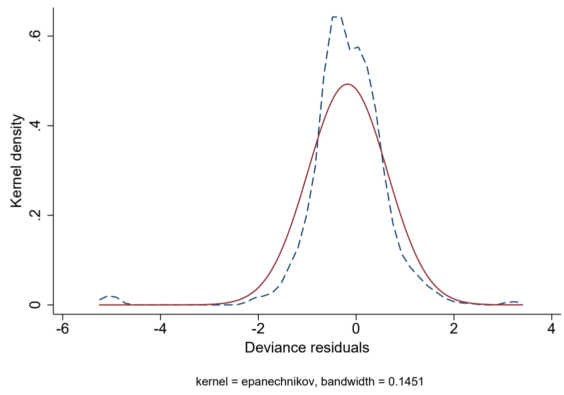


Figure 23: Robustness checks: GLMs estimators for developed countries: Density of deviance residuals.

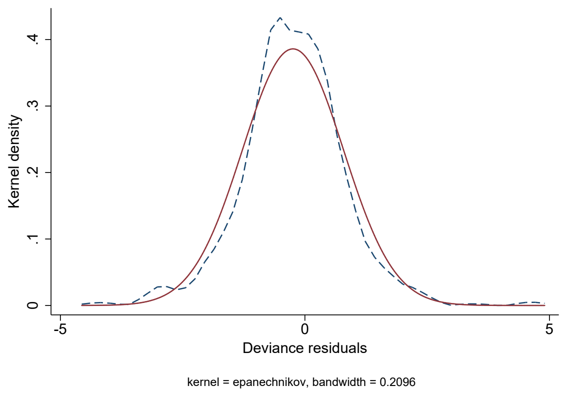
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

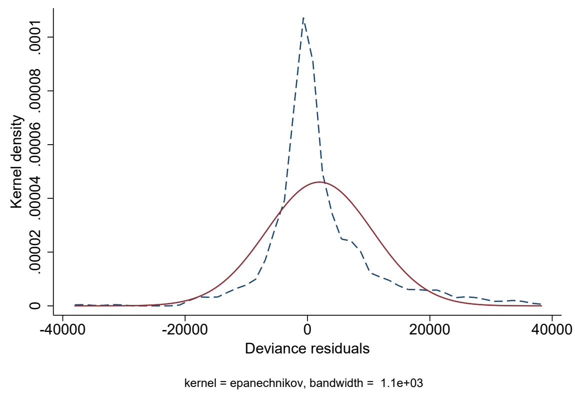
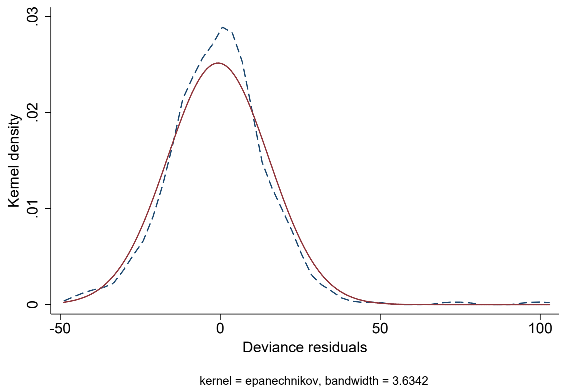
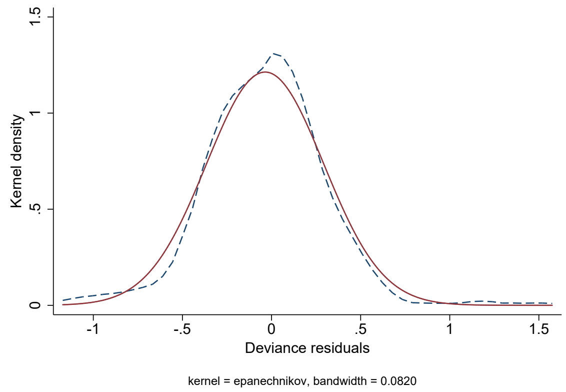


Figure 24: Robustness checks: GLMs estimators for developing countries: Density of deviance residuals.

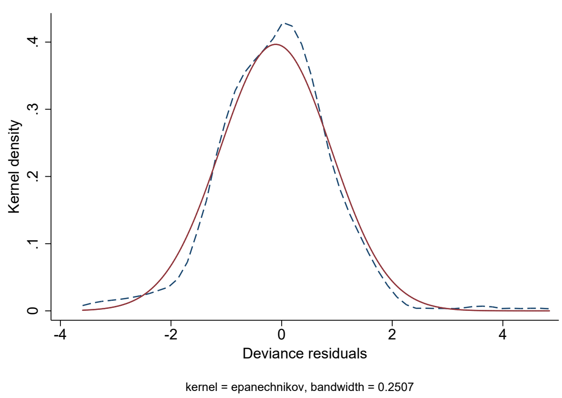
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

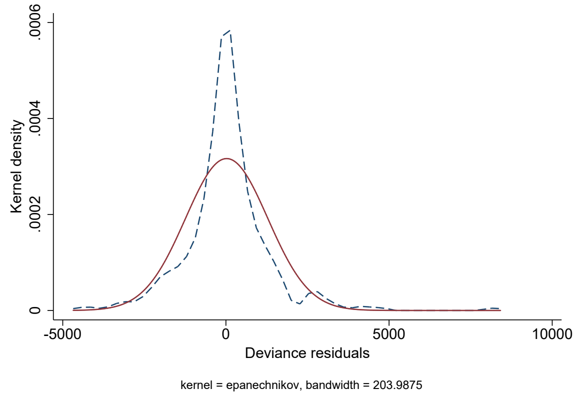
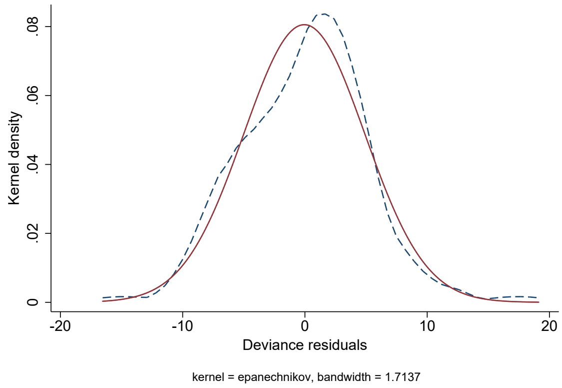
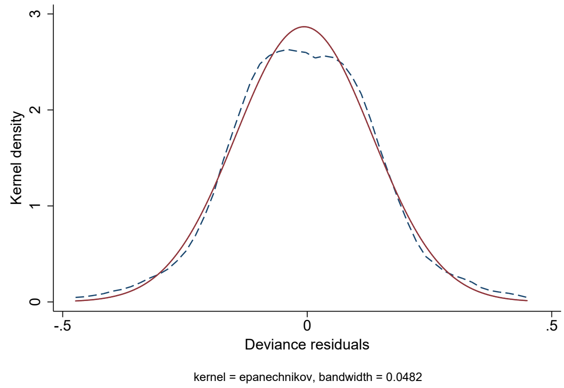


Figure 25: Robustness checks: GLMs estimators for Latin American countries: Density of deviance residuals.

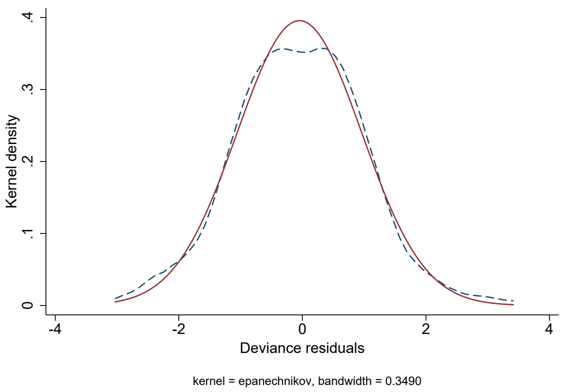
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

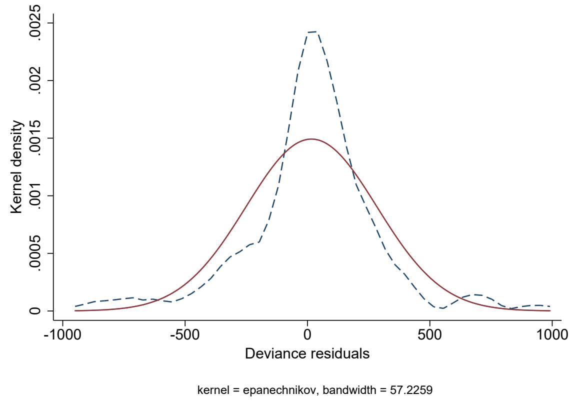
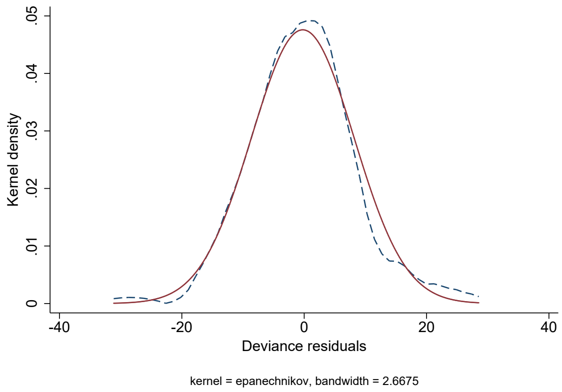
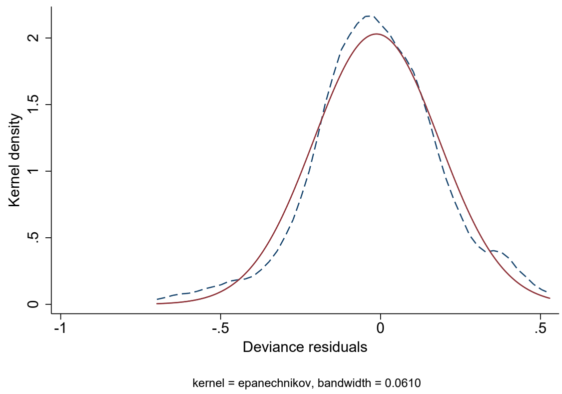


Figure 26: Robustness checks: GLMs estimators for Asian countries: Density of deviance residuals.

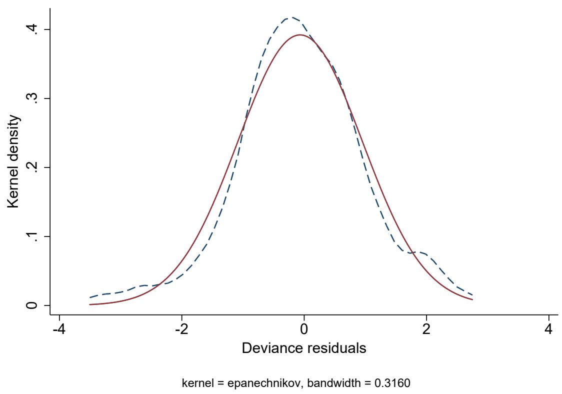
(a) PPML



(b) GPML



(c) NBPML



(d) Gaussian GLM

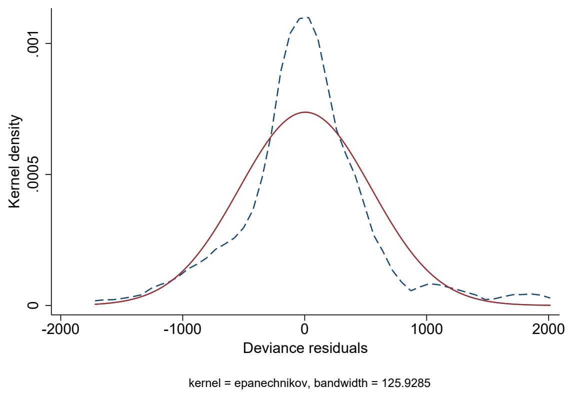


Figure 27: Robustness checks: GLMs estimators for Core EU countries: Density of deviance residuals.

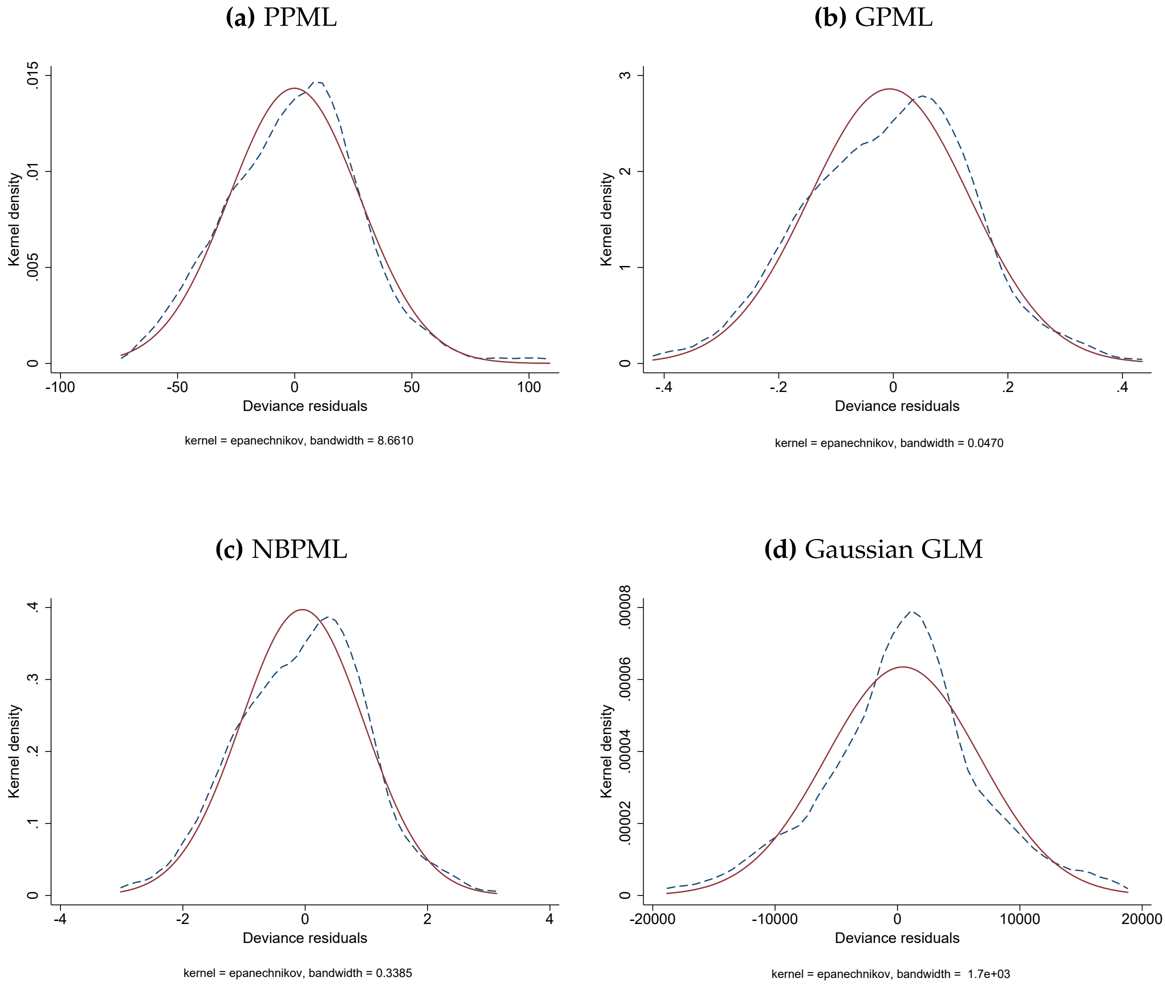


Figure 28: Robustness checks: GLMs estimators for Peripheral EU countries: Density of deviance residuals.