

## Semi-mixed effects gravity models for bilateral trade

Isabel Proença · Stefan Sperlich · Duygu Savaşçı

Received: 23 July 2013 / Accepted: 21 November 2014 / Published online: 9 December 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** In recent years, different alternatives have been suggested to specify and estimate gravity models for bilateral trade. Presently, the so-called Poisson pseudo-maximum likelihood (PPML) with log-linear index is probably the most commonly used method. A method is proposed for panel data that targets to reconcile the pros and cons of fixed and random effects models, respectively. It applies equally to two- and three-way panel models and those with country-specific time-varying effects. It allows to filter out potential correlation between observed and unobserved heterogeneity and to identify the effects of time-invariant factors. It can also be used when panels are short in time, and to other specifications than the PPML-like gamma PML, zero-inflated, or Tobit-like models. We introduce and illustrate the proposed estimator with a study of bilateral trade flows across the European Union before the recent economic crisis.

**Keywords** Gravity models · Regional trade · Panel econometrics

**JEL Classification** F10 · F15 · C14 · C23

---

I. Proença  
ISEG, School of Economics and Management, Universidade de Lisboa and CEMAPRE,  
Rua do Quelhas, 6, 1200-781 Lisbon, Portugal  
e-mail: isabelp@iseg.utl.pt

S. Sperlich (✉)  
Geneva School of Economics and Management, University of Geneva,  
Bd du Pont d'Arve 40, 1211 Geneva 4, Switzerland  
e-mail: stefan.sperlich@unige.ch

D. Savaşçı  
IEA Data Processing and Research Center, Hamburg, Germany

## 1 Introduction

The gravity model is still a standard tool when analyzing trade flows. [Niedercorn and Bechdolt \(1969\)](#) gave an economic derivation of this model, followed by [Anderson \(1979\)](#) who started out from consumer expenditure system theory. More recent theoretical contributions in this direction are provided, for example, by [Deardorff \(1998\)](#) who proved the consistency of the gravity model with the Heckscher–Ohlin trade theory, [Anderson and van Wincoop \(2003\)](#) who revisited the gravity model extending it to deal with the border puzzle, and [Anderson and Yotov \(2009\)](#) who addressed the proportions of trade costs paid by sellers and buyers.

At the same time, there has been myriad empirical research focusing on model specifications, relevant covariates, and estimation procedures that fit the data to a reasonable economic model. The work of [Bergstrand \(1985\)](#) marked an important contribution to bridge economic theory and common empirical practices. However, this discussion of correct econometric specification and estimation can differ substantially from the textbook theoretical model, see [Mátyás \(1997\)](#).

[Santos Silva and Tenreyro \(2006\)](#) have raised a problem that has been ignored so far by both the theoretical and applied communities. They argue that logarithmic transformations are inadequate to estimate elasticities. They argue that multiplicative trade models with multiplicative error, which typically exhibit heteroscedasticity, engender dependency between the error term of the transformed log-linear model and the regressors, and consequently inconsistency of the ordinary least squares estimator. As an alternative, they propose the Poisson pseudomaximum likelihood (henceforth PPML) without the use of a log-transformation. Even with the perceived popularity of the PPML, there still remains some doubt on the generality of the estimator for empirical bilateral trade models.<sup>1</sup> Based on the intensive simulation studies, [Martinez-Zarzoso \(2013\)](#) shows that in several situations feasible GLS combined with the log-transformation can have a better performance than PPML. Also, [Gomez-Herrera \(2013\)](#) performed a large simulation comparison but focused more on the handling of zero-trade. She recommended the log-linear specification with Heckman's two-step estimator.

While model specification is still at the forefront of applied gravity models, the appropriate encapsulation of unobserved heterogeneity is also important. For example, [Mátyás \(1997\)](#) uses fixed effects (FEs) for time, unobserved importer- and exporter-specific heterogeneity. [Egger and Pfaffermayr \(2003\)](#) consider a three-way model with importer- and exporter-specific, country-pair and time effects; it is identical to a two-way model with only country-pair and time effects but allows for a direct estimation of the centered country-specific effects. [Anderson and van Wincoop \(2003\)](#) estimate the unobservable countries' resistance to trade as country-specific fixed effects in a particular gravity model. [Baltagi et al. \(2003\)](#) discuss two- and three-way models with time-varying country-specific effects. But as in such a model, only a very limited set of factor impacts can directly be identified, these are typically neglected; see also [Nuroğlu and Kunst \(2013\)](#). [Anderson and Yotov \(2009\)](#) use only directional

---

<sup>1</sup> Note that like for the log-linear OLS, the consistency of PPML depends on the assumptions made about the error structure.

(i.e., importer and exporter) FEs. [Westerlund and Wilhelmsson \(2011\)](#) use a Poisson FEs estimator to study the trade effects of the 1995 European Union enlargement. Note that FE estimators yield only within-estimates and typically identify only the impact of time-varying variables.<sup>2</sup> Furthermore, the popular differencing estimator cannot be applied, resulting in an incidental parameter problem for many nonlinear models;<sup>3</sup> see [Serlenga and Shin \(2007\)](#) for a rather complex alternative. Random effect (RE) models might be an alternative, especially if the effect of time-invariant observables is of interest. They may also capture potential overdispersion, see [Lin and Zhang \(1999\)](#) for a semiparametric context.

The main problem is the independence assumption for the REs. In order to circumvent this problem, [Lombardía and Sperlich \(2012\)](#) introduce a class of semi-mixed effects models (SMEM) for cross-sectional data.<sup>4</sup> They include a nonparametric proxy in a parametric multilevel model to filter out potential mean dependency between the REs and covariates. The flexibility of the nonparametric function allows one to place SMEM between the FE and RE models when the first suffers from overparameterization (with all the mentioned problems) while the second suffers from underparameterization (leading to potential endogeneity problems). The SMEM can be interpreted as an extension of the [Mundlak \(1978\)](#) device for RE models or vice versa, since it just adds more parameters in the RE model coming thereby closer to the FE model. The inclusion of a nonparametric filter does not degrade the estimation of the effects of time-invariant variables; one might even include time or other FEs. The proposed model can be estimated with programs provided in standard software packages. Most of the semiparametric contributions to panel count data are based on REs, that is, assuming that unobserved individual heterogeneity is distribution-free but not correlated with the covariates, see, for examples, [Gurmu et al. \(1999\)](#) and [Zheng \(2008\)](#) who resort to Bayesian techniques. A distinct approach is given by [Wellner and Zhang \(2007\)](#).

Summarizing, a main motivation for using SMEM is that of overparameterization when one incorporates FEs but imposes too strong of an assumption (or underparameterization) when using a simple RE model. The SMEM is a middle ground similar to the correlated RE panel data estimator in the standard conditional mean setting. In the next section, the basic model is introduced. We then explain our modification and estimator. In Sect. 3, we study the trade flows among the 25 members of the European Union after the expansion in 2004 toward Eastern Europe. To not distort the results by the financial crises, we concentrate on the period of 2004–2007. In Sect. 4, we present a simulation study. Section 5 concludes.

---

<sup>2</sup> Except if the time-invariant covariate exhibits variation in a different dimension than the fixed effects, for example when we have covariates  $X_{ij}$  and FEs  $(\eta_i, \eta_j)$ .

<sup>3</sup> Exceptions are, for example, Logit or Poisson-distributed responses where conditioning on their mean eliminates the country-pair fixed effects.

<sup>4</sup> They actually do it in the context of small-area statistics.

## 2 A semi-mixed effects gravity model for panels

In the standard gravity model, bilateral trade  $T_{ijt}$ , say import from country  $j$  to country  $i$  at time  $t$ , is assumed to be determined by the GDPs  $Y_{it}$ ,  $Y_{jt}$  of countries  $i$  and  $j$ , by trade cost factors that might be represented by country-pair-specific information such as contiguity and distance, common language, ethnic groups or borders, and by country-specific information like landlockedness or population size. Instead of including GDP and population separately, it is common to use GDP per capita. In panel specifications, it is common to include time FEs to account for business cycles. For notational convenience, we classify the independent variables into the following groups: the GDPs  $Y_{it}$ ,  $Y_{jt}$ , the binary information  $D_{ijt}$ , non-binary but time-invariant information  $Z_{ij}$ , and the remaining set  $X_{ijt}$ . All three types of vectors,  $D_{ijt}$ ,  $Z_{ij}$ , and  $X_{ijt}$ , include variables that may vary only over  $i$ ,  $j$ , or both  $(i, j)$ , i.e., are country-specific (like the number of patents or landlockness) or specific to the relation of these countries (like common border). In the former case, we may split, for example,  $Z_{ij}$  into  $(Z_i, Z_j)$ . For  $i, j = 1, \dots, n_i, t = 1, \dots, n_t$ , often the following gravity model is considered

$$T_{ijt} = \exp[\beta' \ln(X_{ijt}) + \gamma' \ln(Z_{ij}) + \delta' D_{ijt} + \alpha_t + \eta_{ij}] Y_{it}^{\beta_1} Y_{jt}^{\beta_2} + \epsilon_{ijt} \quad (1)$$

$$= \exp[\beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta' \ln X_{ijt} + \gamma' \ln Z_{ij} + \delta' D_{ijt} + \alpha_t + \eta_{ij}] \epsilon_{ijt}, \quad (2)$$

with time FEs  $\alpha_t$ , unobserved heterogeneity  $\eta_{ij}$ , and random terms  $\epsilon_{ijt}$  (or  $\epsilon_{ijt}$ , respectively). It will become obvious that our modeling and estimation idea can be equally well used for country-specific and three-way effect models, for example if one wants to decompose  $\eta_{ij}$  into  $\eta_i + \xi_j$ . Here,  $\epsilon_{ijt}$  is a zero-mean random variable and  $\epsilon_{ijt}$  consequently a multiplicative disturbance term with a conditional mean equal to one. Further,  $X_{ijt}$ ,  $Z_{ij}$ , and  $D_{ijt}$  are column vectors,  $\beta$ ,  $\gamma$ , and  $\delta$  column vectors of unknown coefficients of corresponding size, and  $\beta_1, \beta_2$  unknown scalar coefficients. If one wants to allow for time-varying unobserved heterogeneity of countries like in [Baltagi et al. \(2003\)](#), then one has to additionally introduce  $\eta_{it} + \xi_{jt}$ . If they are defined to be FEs, then only the coefficients of variables that vary over both countries are (directly) identified. Note that some studies aim to estimate the total country-specific or the country-pair trade costs by FEs; others are more interested in estimating the coefficients of the covariates.

If the conditional expectation of  $\ln \epsilon_{ijt}$  is a function of the independent variables and/or  $\eta_{ij}$ , then the OLS estimator of the logarithmic transformation of (1), i.e.,

$$\ln T_{ijt} = \beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta' \ln X_{ijt} + \gamma' \ln Z_{ij} + \delta' D_{ijt} + \alpha_t + \eta_{ij} + \ln \epsilon_{ijt},$$

is inconsistent (no matter if  $\eta_{ij}$  refers to RE or FE), which was the motivating reason behind Santos Silva and Tenreiro's (2006) PPML estimator. For the reasons discussed, namely that FE models only capture the within effects, do not allow for the direct estimation of the impact of time-varying variables, and (potentially, depending on the model and estimation method) suffer from incidental parameter problem, we concen-

trate in the following on the problem of estimating the gravity model by a semi-mixed effects PPML, including versions with directional and time-varying effects.

## 2.1 The semi-mixed effects gravity model

There are two main alternatives to introduce the SMEM. [Lombardía and Sperlich \(2012\)](#) propose a nonparametric filter to eliminate possible dependency between random heterogeneity and the covariates. Translated to our problem, they make use of the fact that for a set of (time-invariant) variables  $W_{ij}$ , there may exist a function  $\psi_s(\cdot)$  of smoothness  $s$  such that for

$$\eta_{ij} = \psi_s(W_{ij}) + u_{ij} \quad (3)$$

$u_{ij}$  is a RE unrelated to  $Y_{it}$ ,  $Y_{jt}$ ,  $X_{ijt}$ ,  $Z_{ij}$ ,  $D_{ijt}$ , and  $W_{ij}$ . Why should a nonparametric function  $\psi_s$  offer this? For  $W_{ij}$  (containing at least one continuous variable), a nonparametric function with arbitrary smoothness  $s$  can be everything between interpolation (say  $s = 0$ ) and a constant (say  $s = \infty$ ). This implies that (3) offers us a continuum of models between the two extremes, namely the FE and the RE model. Parameter  $s$  is a slider that places our model between (if  $0 < s < \infty$ ) or on one of them (if  $s = 0$  or  $s = \infty$ ). Certainly, a poor choice of variables in  $W_{ij}$  will lead to a wiggly function approaching almost the FE model. For example, (1) becomes in any case

$$T_{ijt} = e^{\beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta' \ln X_{ijt} + \gamma' \ln Z_{ij} + \delta' D_{ijt} + \alpha_t + \psi_s(W_{ij}) + u_{ij}} + \epsilon_{ijt} \quad (4)$$

with  $u_{ij}$  being independent of  $\epsilon_{ijt}$  by assumption. If one considers directional effects  $\eta_i + \xi_j$ , decomposition (3) becomes  $\eta_i = \psi_{s1}^1(W_i) + u_i$ ,  $\xi_j = \psi_{s2}^2(W_j) + v_j$  with  $\psi^1$ ,  $\psi^2$  being two different nonparametric functions of smoothness  $s1$ ,  $s2$ , and  $u$ ,  $v$  the remaining REs; analogously, one has  $\eta_{it} = \psi_{s1}^1(W_{it}) + u_{it}$ ,  $\xi_{jt} = \psi_{s2}^2(W_{jt}) + v_{jt}$  for the time-varying version.

There are three open questions: the choice of smoothness  $s$ , composition of  $W_{ij}$ , and estimation of (4). Concerning the first question, when using smoothing methods for estimating  $\psi_s$ , the choice of  $s$  is equivalent to the question of smoothing parameter selection. For example, using kernel smoothing, this is simply bandwidth choice ([Köhler et al. 2014](#)), the penalization parameter when using P-splines ([Wood 2008](#)), or the number of knots when using regression splines ([Nie and Racine 2012](#)). Whereas smoothing parameter selection is commonly invoked based on minimizing mean squared error of the unknown function relative to the truth, here our goal was to smooth over  $W_{ij}$  to remove potential correlation between the REs and the covariates. Recalling that a trade-off between bias and variance still exists, which will serve as a platform for selecting an appropriate smoothing parameter. Note that Sections 4.5 to 4.7 of [Wood \(2006\)](#) are dedicated to the different methods of smoothing parameter selection for P-spline regression, namely the minimization of the unbiased risk estimator, the minimization of the cross validation, or the generalized cross-validation criterion. As P-splines can be rewritten as linear mixed effects models in which the smoothing parameter appears as the a priori variance of spline basis coefficients, one might also take an estimate of this variance for being our data-adaptive smoothing parameter, see [Kauermann \(2005\)](#) for details.

Concerning the second question, proper composition of  $W_{ij}$  should start from [Mundlak \(1978\)](#). Statistically speaking, he introduces additional parameters in the RE model in order to eliminate potential endogeneity, that is, [Mundlak's \(1978\)](#) approach lies between the RE and FE model. This brings us to the second way of introducing the SMEM: as an extension of his device, again regarding the RE model as an underparametrized model with the FE model as its overparametrized counterpart. He proposes to include (linearly) the temporal means of the time-varying covariates. However, for short-time panels, the inclusion of all these means (additional to the time-varying variables) may render estimation numerically less stable and may introduce covariance structures with large variances and covariances for the coefficient estimates. This holds in particular for those variables that exhibit little variation over time. Nonetheless, since [Mundlak \(1978\)](#), these temporal means are considered to be the most natural candidates for the vector of proxy variables  $W_{ij}$  in order to control for the relation of  $(Y_{it}, Y_{jt}, X_{ijt}, Z_{ij}, D_{ijt})$  with  $\eta_{ij}$ .<sup>5</sup> For some applications, one may have a clear idea of the causes of the dependency between explanatory variables and the individual unobserved heterogeneity term. For instance, if we modeled wages, this is to a good part individual ability, and we would look for corresponding proxies like IQ, c.f. [Blackburn and Neumark \(1992\)](#).

Estimation of (4) follows from existing methods for generalized partial linear mixed effects models (GPLMM). In practice, given the dimensionality of  $W_{ij}$ , one typically would model  $\psi_s$  additively resulting in a generalized additive mixed effects model (GAMM). This simplifies estimation and mitigates the curse of dimensionality. Estimation of the GPLMM and GAMM is well studied, see, for example, [Lin and Zhang \(1999\)](#), [Wood \(2006\)](#), [Lin and Carroll \(2006\)](#), [Lombardía and Sperlich \(2008\)](#), and references therein. The first two apply P-splines, whereas the latter two apply kernel smoothed profiled likelihood. Implementations are provided in different software packages, typically based on splines; they can be found as extensions of generalized additive models (see, e.g., `mgcv` in **R**) or in the context of Bayesian statistics (see, e.g., `BayesX` in **R**).

## 2.2 Estimating SMEMs by penalized splines

Before we introduce the implementation of SMEMs via P-splines,<sup>6</sup> we should briefly comment on this choice because often GPLMMs are estimated via kernel smoothing. Both methods (P-spline and kernel smoothing) maximize a pseudo- or quasi-likelihood.<sup>7</sup> In the kernel approach, this objective function is smoothed by kernels

<sup>5</sup> When including temporal means of covariates, one has to be careful for the coefficients' comparison: for example, including  $\beta \ln x_{ijt}$  and  $\xi \cdot \frac{1}{n_t} \sum_t \ln x_{ijt}$ , the total impact of  $\ln x_{ijt}$  is  $\beta + \frac{1}{n_t} \xi$ . One can therefore find both, studies looking only at  $\hat{\beta}$  and those looking at  $\hat{\beta} + \frac{1}{n_t} \hat{\xi}$ .

<sup>6</sup> The notation is not unique in the literature, while many use it as a synonym for *penalized splines* in general, others refer exclusively to its implementation with B-spline basis.

<sup>7</sup> [Santos Silva and Tenreiro \(2006\)](#) use the notation of *pseudo* referring to [McCullagh and Nelder \(1989\)](#) who speak of *quasi*-likelihood. [Nelder \(2000\)](#) differentiates between them. Along his classification, one might be more interested in the quasi-maximum likelihood methods, but in abuse of notation, we keep the abbreviation PPML.

when calculating the nonparametric part; in the P-spline approach, a regression spline is used inside the quasi-likelihood, and a term is added to the objective function that penalizes for the wiggleness of the regression splines. Commonly attributed advantages of the latter are the simpler implementation, its computational speed, the easier imposition of additional structure like additivity, and the availability of different smoothing parameter selectors. For these reasons, we opted for splines without generally judging one smoother to be superior to the other.

For the sake of notation but without the loss of generality, let us explain the idea and implementation of penalized splines for the reduced set of covariates  $D_{ijt}$ ,  $W_{ij}$  with  $W_{ij} \in \mathbb{R}^{d_w}$ . For presentation, it is easier to work with an additive nonparametric function, i.e.,

$$\psi_s(W_{ij}) = f_0 + f_1(W_{ij1}) + \dots + f_d(W_{ijd_w}) \tag{5}$$

where  $f_0$  is a constant and  $f_k$ ,  $k = 1, \dots, d_w$  are one-dimensional nonparametric functions, centered to zero for identification, and typically assumed to be twice continuously differentiable. The nonadditive case is discussed below. Each function  $f_k$  is approximated by a spline (piecewise polynomials on intervals, separated by so-called knots, that cover the support of the covariate). One speaks of penalized splines if in the moment of estimation for each nonparametric  $f_j$  an additional term is added that penalizes for its wiggleness. As a spline basis, B-splines are the most popular ones for regression splines, simple cubic splines are also often used for penalized splines. For a cubic regression spline, we can write for each nonparametric additive component

$$f_k(\omega) = \alpha_0 + \alpha_1\omega + \alpha_2\omega^2 + \alpha_3\omega^3 + \sum_{l=1}^L \theta_l \omega^3 \mathbb{I}\{\tau_{l-1} < \omega \leq \tau_l\} \tag{6}$$

with  $\tau_l$  being the knots,  $\tau_0 < \min\{W_{ijk}\}$ ,  $\tau_L \geq \max\{W_{ijk}\}$ , and  $\sum_{l=1}^L \theta_l = 0$ . It is obvious that this has continuous second derivatives everywhere inside the support, and that the smoothness can be controlled via the  $\theta_l$ . We disregard here smoothing splines which consider each observation as being a knot. To reach consistency, the number of knots is assumed to converge to infinity for increasing sample size. For a given sample size, there are a finite number of knots and therefore (5) becomes an expression of several parametric polynomials with a finite number of parameters that can be estimated by parametric methods.

Once the choice of the basis is made, the challenge for estimation is to fix the number and placement of knots. While the number could be determined by a data-driven criterion like cross validation, the placement is a much harder problem. The P-splines is a compromise between regression and smoothing splines. In fact, depending on the knots-sample-size ratio, P-splines converge to the same limit as either regression or smoothing splines do, see [Krivobokova et al. \(2010\)](#). The key idea is to be generous with the number of knots such that their placement becomes less important. In order to avoid overfitting, penalizing terms are added to the objective function (i.e., the quasi-likelihood) to control the variation in the  $\theta_l$  in (6). Numerically, this is equivalent to the estimation of a generalized linear mixed effects model (GLMM) with the  $\theta_l$  being random coefficients with variance  $\sigma_\theta^2$ . Standard methods from parametric GLMM can

be used to estimate  $\sigma_\theta^2$  providing a data-adaptive mean squared prediction error minimizing smoothing parameter for the nonparametric functions. So there is an interplay between the number of knots and the penalizing parameter. Therefore, nonparametric estimates based on different number of knots are virtually identical as long as a reasonably large number of knots is chosen. In practice, every 10th data point of the ordered sample is taken as a knot.

Clearly, to this GLMM, we can add now  $\delta' D_{ijl}$  and REs  $\eta_{ij}$  without changing the estimation procedure. So far, only approximate inference is available, for example using the penalized quasi-likelihood approach of [Breslow and Clayton \(1993\)](#) adapted to GAMMs as outlined in [Lin and Zhang \(1999\)](#).<sup>8</sup> For PPML, variances were assumed to be proportional to the mean. Although in practice this is often the case, one could calculate the standard deviations by resampling methods like subsampling or wild bootstrap as in [Lombardía and Sperlich \(2008\)](#) at high computational costs. Further research should be devoted to the estimation and efficient implementation of standard errors that are robust against more complex heteroscedasticity and autocorrelation. Note that the classical naive bootstrap is typically inconsistent in this context.

If data sets are large enough or interactions between some proxies are very likely to be important for the filtering, then one should relax the additivity assumption. However, for  $d > 3$ , the curse of dimensionality renders  $\sqrt{n}$ -estimation of the parameters of interest infeasible or at least very hard with the need of imposing strong smoothness conditions on  $\psi_s$ . This would also contradict the filtering idea for which the possibility of choosing a small  $s$  is crucial. Therefore, we recommend to include interaction terms separately in an additive fashion (see [Sperlich et al. 2002](#)), e.g.,

$$\psi_s(W_{ij}) = f_0 + f_{1,2}(W_{ij1}, W_{ij2}) + f_3(W_{ij3}) + \dots + f_d(W_{ijd_W}). \quad (7)$$

More generally, we can think of (5) with some of the  $W_{ijk}$  being vectors of dimension  $\leq 3$ . These interaction terms are typically estimated with thin-plate splines ([Green and Silverman 1994](#)) or simply by using tensor products of the corresponding one-dimensional splines, e.g., in (7) you model  $f_{1,2}$  by the tensor product of the basis functions of  $f_1$  and  $f_2$  in (5). Smoothing parameter selection remains the same as before.

We conclude this section with two remarks. First, in the moment of estimation, we may face the problem of multifunctionality (the nonparametric analogue to multicollinearity). There are basically two ways to handle this: variable selection or imposing some structure on  $\psi_s$  as will be seen in our application. The problem of testing and model selection in GPLMMs has recently been studied in different articles, see, for example, [Sperlich and Lombardia \(2010\)](#). Model selection can be based on different existing criteria including, for example, AIC. Although we consider a quasi-likelihood, the likelihood is still the objective function whose maximization we take as the criterion for estimation.

Note finally that an extension of our model with nonparametric filter to the more complex semi- and nonparametric world is straightforward and can even be performed

<sup>8</sup> For the implementation, see [Venables and Ripley \(2002\)](#) concerning GLMM and the modifications of [Wood \(2011\)](#) for the particular case of P-splines for GAMMs.



with the same estimation method and software.<sup>9</sup> More specifically, one might want to extend model (4) to a GAMM of the form

$$T_{ijt} = \exp[\beta_1(Y_{it}) + \beta_2(Y_{jt}) + \beta(X_{ijt}) + \gamma(Z_{ij}) + \delta' D_{ijt} + \alpha_t + \psi_s(W_{ij}) + u_{ij}] + \epsilon_{ijt} \quad (8)$$

with  $u_{ij}$  being REs,  $\epsilon_{ijt}$  random errors,  $\beta_1$  and  $\beta_2$  univariate nonparametric functions,  $\beta$ ,  $\gamma$ ,  $\psi_s$  nonparametric (additive) functions, and  $\alpha_t$  time FEs.

### 3 Trade flows inside the European Union after the big eastward enlargement

We study trade flows among the EU25 countries from 2004 through 2007 by the use of our method inside the PPML framework. Note that this was the period from the opening of the EU toward Eastern Europe until the financial crises in 2008.

#### 3.1 Data and variables

The dependent variable of interest will be the import flows  $T$ , given that countries often tend to monitor their imports more carefully than their exports. Regarding the independent time-variant variables, we use as proxies for the overall economic mass of the countries the logarithm of GDPs (MGDP and XGDP for importer and exporter countries, respectively) and the logarithm of populations (MPOP and XPOP for importer and exporter countries). We expect a positive effect for MGDP and XGDP while for MPOP and XPOP, the literature documents an ambiguous effect. As proxies for bilateral trade costs, we have the distance of the two trading partners (logarithm of kms between the capitals), DIST, and a dummy to capture contiguity, NEIGH, which is equal to one if both trading countries share a land or sea border. We expect a negative effect on trade from DIST and a positive one from NEIGH. Further variables of interest are whether countries are landlocked, MLOCK, and XLOCK, which are equal to one if the importer (respectively, exporter) country has no direct connection to the sea; COMLANG which is equal to one if both trading partners share the same official language; ETHNIC which is equal to one if in the importer country there is an ethnic minority of the exporter country; and EU15 which is equal to one if both countries belonged to the European 15. While for landlockedness we expect negative effects on trade, the effect of all the other dummy variables should be positive. Finally, we introduce time dummies, denoted by 2005, 2006, and 2007.

Table 1 reports some descriptive statistics of the variables of interest. Further details about the variables can be found in the ‘‘Appendix.’’ Although we observe some zeros for trade ( $T$ ), zero inflation is not an issue given the limited occurrences in our data set. Note that GDP and population are highly correlated ( $|\rho| > 0.8$ ). Being landlocked is strongly correlated with population size ( $|\rho| \approx 0.3$ ) and GDP ( $|\rho| \approx 0.25$ ). As COMLANG and ETHNIC are insignificant with  $p$  values clearly above 0.5 in all

<sup>9</sup> Henderson and Millimet (2008) state that the added value of nonparametric gravity models is poor. However, they start out from the more restrictive model of Anderson and van Wincoop (2003). Their conclusions do not necessarily hold for the more flexible gravity models considered here.

**Table 1** Descriptive statistics

	Variable	Mean	SD	Min	Max
	T	37.55	95.48	0.00	920.77
	GDP	12.00	1.52	9.17	14.70
	POP	16.09	1.27	13.03	18.23
	DIST	7.01	0.64	4.04	8.12
	NEIGH	0.13	0.34	0	1
All variables except $T$ and the dummies are given in logarithms. We observed 25 countries over 4 years giving 2,400 observations of bilateral trade flows	COMLANG	0.03	0.17	0	1
	ETHNIC	0.02	0.13	0	1
	EU15	0.35	0.48	0	1
	LOCK	0.20	0.40	0	1

our models, we opt to give below only estimation results with models where these regressors are excluded.

For the proxies,  $W_{ij}$ , that shall filter possible dependence between the unobserved heterogeneity and the regressors, we consider variables that reflect the size of a country. This goes back to the idea that multilateral resistance is associated with country size, see, for example, (Anderson and Yotov 2009). As GDP and population are already included, further possible proxies are the area of a country, MAREA and XAREA, reflecting the physical size, and the number of patents (MPAT, XPAT), reflecting the technological size. Note that the AREA variables are highly correlated with the corresponding population sizes ( $|\rho| \approx 0.8$ ) and the corresponding GDPs ( $|\rho| \approx 0.6$ ). The number of average patents have a correlation of more than 0.8 with average GDP, and of about 0.6 with population. Note finally that the selection of proper proxies for  $W$  depends on the context; our proposal here is not necessarily a general suggestion.

Most of the results we present below have been obtained by the `gamm` procedure in the software package **R**, see Wood (2006). For the models for which alternative software existed in **R** or `stata`, we counterchecked the results. Where offered by the routines, we tried different specifications of the distribution of the random effects. Estimates did not vary significantly along the different specifications.

### 3.2 Estimation and comparison of models

Tables 2 and 3 provide coefficient estimates, excluding intercepts, for a variety of different models discussed above. Function  $\psi_s$  is composed of additive P-spline functions with data-adaptive smoothing parameter and knots as described above. We start with the estimation of parametric FE panel models with country-pair ( $\eta_{ij}$ , Model 1a) and directional ( $\eta_i + \xi_j$ , Model 1b) FEs, respectively. Models 2a and 2b are the SMEM counterparts of Models 1a and 1b where the temporal averages of GDP and population are included in  $\psi_s$ .

We neglect here the possibility of two-step estimation of the impact of covariates whose coefficients are not identified in the FEs model. Consequently, in the first model, we see only estimates for covariates that vary over time, while for the second we get also estimates for pair-specific effects that are time-invariant. In the appendix

**Table 2** Estimates of coefficients and random effects standard deviations

	Model 1a FE $\eta_{ij}$	Model 1b FE $\eta_i + \xi_j$	Model 2a $\psi_{ij} + u_{ij}$	Model 2b $\psi_i^1 + u_i + \psi_j^2 + v_j$
MGDP	0.823 (0.091)**	0.820 (0.212)**	0.822 (0.042)**	0.819 (0.207)**
XGDP	0.228 (0.110)*	0.409 (0.092)**	0.227 (0.026)**	0.232 (0.131)
MPOP	0.000 (0.003)	0.000 (0.027)	0.000 (0.004)	0.001 (0.019)
XPOP	-0.093 (0.061)	0.022 (0.121)	-0.092 (0.046)*	-0.068 (0.201)
DIST	-	-0.001 (0.000)**	-1.228 (0.056)**	-0.722 (0.023)**
NEIGH	-	0.375 (0.034)**	0.284 (0.093)**	0.348 (0.026)**
EU15	-	0.680 (0.067)**	0.505 (0.094)**	0.769 (0.066)**
MLOCK	-	-	-0.157 (0.094)	0.271 (0.143)
XLOCK	-	-	0.111 (0.114)	0.354 (0.259)
2005	0.018 (0.010)	0.010 (0.039)	0.018 (0.005)**	0.018 (0.026)
2006	0.072 (0.018)**	0.054 (0.044)	0.072 (0.007)**	0.072 (0.034)*
2007	0.087 (0.026)**	0.058 (0.058)	0.087 (0.009)**	0.087 (0.046)
mM/XGDP	-	-	n.p.	n.p.
mM/XPOP	-	-	n.p.	n.p.
$\sigma_u$ or $\{\sigma_u, \sigma_v\}$	-	-	0.531	{0.230, 0.372}

Temporal means of time-varying regressors are denoted by an “m” in front of the original name of the variable. Standard errors are in (·) brackets

*n.p.* nonparametric estimated; the corresponding estimates are shown in Figs. 3 and 4

\* Significance at the 5%, \*\* at the 1% level

are also provided the regression results for a PPML without modeling unobserved heterogeneity, i.e., pooled regression without random or fixed effects.

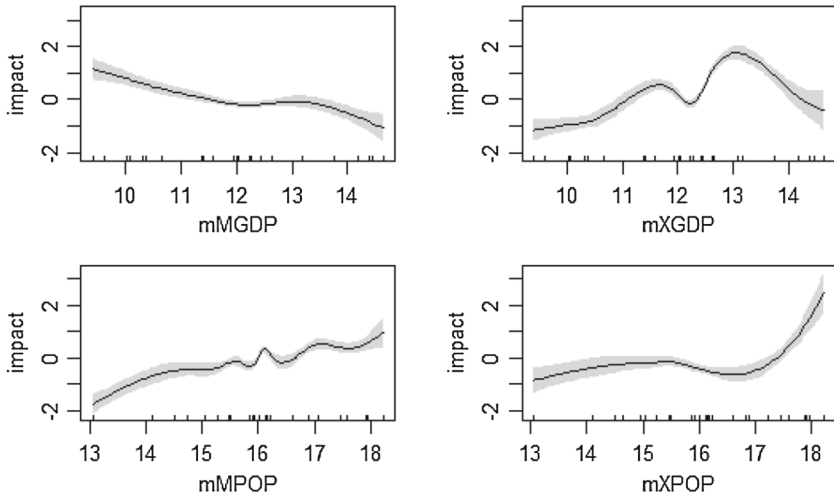
Considering the results in Table 2 for the PPML, the estimated elasticity of exporter GDP has size smaller than the usual estimates obtained with log-linear models, except for Model 1b. This is a tendency depicted also in other studies that apply PPML instead of estimating a log-linear model, see, for example, the cross-sectional study of [Proença et al. \(2008\)](#) considering the same country set addressed here. The population sizes seem not to be relevant, what is most likely due to their high correlation with GDP. If countries share borders or if both trading partners belonged to the EU15, one finds

**Table 3** Estimates of coefficients and random effects standard errors

	Model 3a $\psi_{ij} + u_{ij}$	Model 3b $\psi_i^1 + u_i + \psi_j^2 + v_j$	Model 3c $\psi_{it}^1 + u_{it} + \psi_{jt}^2 + v_{jt}$
MGDP	0.823 (0.042)**	0.824 (0.208)**	0.937 (0.042)**
XGDP	0.228 (0.026)**	0.260 (0.142)	0.097 (0.026)**
MPOP	0.000 (0.004)	0.001 (0.019)	0.009 (0.004)**
XPOP	-0.091 (0.045)*	-0.055 (0.167)	-0.010 (0.046)*
DIST	-1.185 (0.059)**	-0.722 (0.023)**	-0.708 (0.056)**
NEIGH	0.319 (0.100)**	0.346 (0.026)**	0.341 (0.093)**
EU15	0.473 (0.096)**	0.749 (0.065)**	0.734 (0.094)**
MLOCK	-0.039 (0.084)	0.181 (0.138)	0.329 (0.094)
XLOCK	0.033 (0.101)	0.366 (0.220)	0.227 (0.114)
2005	0.018 (0.005)**	0.016 (0.026)	0.020 (0.005)**
2006	0.072 (0.007)**	0.068 (0.034)*	0.083 (0.007)**
2007	0.087 (0.009)**	0.081 (0.047)	0.100 (0.009)**
mMGDP	-0.355 (0.066)	-0.502 (0.217)*	-0.621 (0.369)
mXGDP	0.348 (0.060)	0.225 (0.178)	0.271 (0.204)
mMPOP	0.283 (0.066)	0.355 (0.106)**	0.441 (0.078)**
mXPOP	0.453 (0.087)	0.333 (0.222)	0.530 (0.192)**
M/XAREA	n.p.	n.p.	n.p.
mM/XPAT	n.p.	n.p.	-
M/XPAT	-	-	n.p.
$\sigma_u$ or $\{\sigma_u, \sigma_v\}$	0.573	{0.234,0.372}	{0.156,0.225}

Temporal means of time-varying regressors are denoted by an “m” in front of the original name of the variable. W refers to the vector of size proxies (area and patents). Standard errors are in (·) brackets *n.p.* nonparametric estimated; some of the corresponding estimates are shown in Figs. 3 and 4

\* Significance at the 5%, \*\* at the 1% level

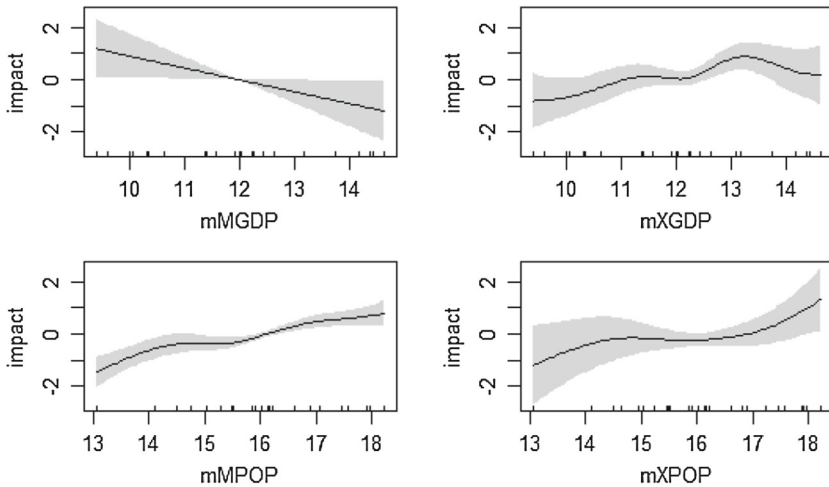


**Fig. 1** Estimates of  $f_1(mMGDP)$ ,  $f_2(mXGDP)$ ,  $f_3(mMPOP)$ , and  $f_4(mXPOP)$  of  $\psi_s$ , see Eq. (5), for Model 2a with 95% confidence bands

significant positive effects. As expected, distance has a significant negative impact, whereas landlockedness has no significant effect inside the EU. Time effects exhibit a positive trend but are only clearly significant for models accounting for country-pair effects.

Comparing Model 1a with 2a, we see that the coefficient estimates stay basically unchanged. As expected, the estimated standard errors are substantially smaller for the SMEM. This is different when comparing Model 1b with 2b where several coefficients have changed quite a bit, and not all standard errors are smaller in 2b. When we compare Model 2a with 2b, we see that their coefficients are much closer to each other than those of their FE counterparts 1a and 1b. Also, the additive components of their corresponding  $\psi$  functions are quite similar, cf. Figs. 1, 2, though the confidence bands in Model 2a are much smaller than those in 2b. At first glimpse, it seems that the model with country-pair effects has smaller standard deviations for most covariates. Note, however, that exactly those covariates that are country-pair-specific, namely DIST, NEIGH, and EU15 have larger standard deviations in the country-pair SMEM, while all importer-/exporter-specific covariates have larger standard deviations in the importer/exporter SMEM. This is to be expected as these covariates try to capture the same variation as the respective REs. Looking at the last line of Table 2, we observe that the variance of the country-pair effects is smaller than the sum of the variances of the directional REs.

The FE Models 1a and 1b identify only the linear within effects of our covariates. In Models 2a, 2b, the *between effects* of GDP and population are included nonparametrically (via their temporal means). One may prefer to include these temporal means linearly, perhaps for ease of interpretation. This stops smoothing parameter  $s$  from being a data-adaptive slider to place our model between the FE and the RE model. However, Chamberlain (1982) showed even for much simpler models that the linear inclusion of all temporal means can easily fail to control for potential endogeneity.

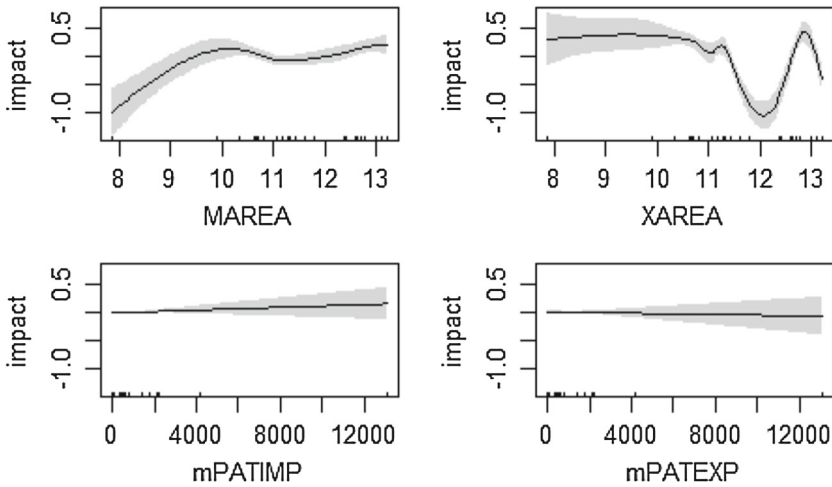


**Fig. 2** Estimates of  $f_1(mMGDP)$ ,  $f_2(mXGDP)$ ,  $f_3(mMPOP)$ , and  $f_4(mXPOP)$  of  $\psi_s$ , see Eq. (5), for Model 2b with 95% confidence bands

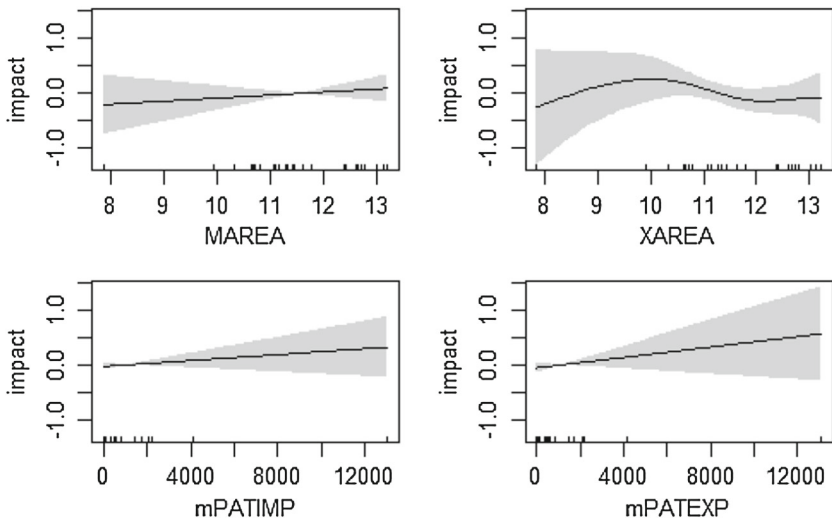
Therefore, we need additional proxies when modeling the impact of the temporal means of GDP and population linearly. In order to do so, we include the size proxies log-area and number of patents in  $\psi_s$ , see Models 3a, 3b, and 3c. These models contain the same variables as above, but with mMGDP, mXGDP, mMPOP, and mXPOP entering these models linearly. As before, the difference between Model 3a and 3b is that the former one has a country-pair RE  $u_{ij}$ , whereas the latter includes the importer and exporter REs  $u_i + v_j$ . We first note that from Model 2a to 3a, there is hardly an important change in the coefficient values except for NEIGH which increases by about 10%. Further, the coefficient estimates on MLOCK and XLOCK change greatly, but remain statistically insignificant. Similarly, the changes from Model 2b to 3b seem to be minor; only the coefficient of XPOP changes by more than 10%, and landlockedness remains insignificant. The variation in the REs  $u_{ij}$  and  $(u_i, v_j)$  across different models is minimal. Not surprisingly, the coefficients of country-pair and directional effects models come closer to each other when new covariates are added to the model.

The nonparametric components of Models 3a, 3b are plotted in Figs. 3 and 4. As before, the functional shapes are similar and the importer/exporter REs model exhibits much larger standard errors (in terms of broader confidence bands, compare the scale of the vertical axes).

Next, we consider models with time-varying country effects. A time-varying country-pair effect collapses with the error term and is therefore skipped. Models 3c and 4c contain time-varying directional effects  $u_{it} + v_{jt}$ . In order to filter potential correlation, we need here proxies  $W$  that are also time-varying. Therefore, in Model 3c, 4c, the number of patents enter  $\psi_s$  with their yearly observations, not aggregated as temporal means. In Model 3c, the coefficient estimates change moderately compared to Model 3b with time-invariant REs. The standard deviations of Model 3c are similar to those of 3a but the coefficient estimates are still closer to those of 3b. Note that the variance (squared standard deviations) of the time-varying REs is a bit larger than



**Fig. 3** Estimates of  $f_1(MAREA)$ ,  $f_2(XAREA)$ ,  $f_3(mMPAT)$ , and  $f_4(mXPAT)$  of  $\psi_s$ , see Eq. (5), for Model 3a with 95% confidence bands



**Fig. 4** Estimates of  $f_1(MAREA)$ ,  $f_2(XAREA)$ ,  $f_3(mMPAT)$ , and  $f_4(mXPAT)$  of  $\psi_s$ , see Eq. (5), for Model 3b with 95% confidence bands

those of Model 3b divided by four ( $= n_t$ ). This indicates that the REs ( $u_{it}, v_{jt}$ ) do not vary much over time. If we compare the results in Table 3 with those obtained for FE Models 1a and 1b, then we see that the coefficient estimates for Model 3c differ most severely.

So far we have only estimated models with additively separable control functions  $\psi$ ,  $\psi^1$ , and  $\psi^2$ . When estimating the model with a four-dimensional nonparametric  $\psi$ , the program stops due to multicollinearity for both estimation procedures, thin plates and tensor products. When including two-dimensional interactions instead, recall (7),

one has still multicollinearity or to impose strong smoothness on  $\psi$ ,  $\psi^1$ , and  $\psi^2$  so that their estimates become in fact linear functions. All selection criteria give worse values than one obtains for additive separable control functions. This certainly is not a general finding but depends always on the data sample at hand.

The last study in this section is the extension of the above models toward semiparametric alternatives in which the use of nonparametrics is not limited to the filtering of potential dependences between covariates and REs. Endogeneity can also occur due to functional misspecification of the impact of covariates. However, if one tries to estimate a PPML with a nonparametric additive index (8), then the estimation procedure suffers from multifunctionality and other numerical instability, unless each variable enters only once, and some in a restricted way. One has to skip the temporal means or the yearly observations of GDP and population. Even then, there are still numerical problems due to the strong dependency between GDP and population. One can either consider GDP per capita and skip the population covariates, or restrict the impact of them to a simple parametric form. The results for two examples are given in Table 4 and Figs. 5 and 6, where Model 4a has a country-pair RE  $u_{ij}$ , and Model 4c time-varying importer/exporter REs  $u_{it} + v_{jt}$ .

When comparing Model 4a with 3a and 4c with 3c, we see that the coefficient estimates have changed along the correlation of their covariates with MGDP and/or XGDP, see, for example, the nonlinear impact of XGDP, Fig. 6, and compare it with the changes for XPOP (and XLOCK) in Model 4a. These could indicate a problem of functional (mis-)specification in the simpler generalized linear models. The main remaining difference between Models 4a and 4c is the much larger confidence bands for the latter, and while in 4a XLOCK is significant but not MLOCK, in 4c we find exactly the contrary situation. Furthermore, in Model 4c, the variance of the REs is almost zero, i.e., once we allow for nonparametric country-specific impact functions of GDP, individual time-varying directional effects disappear. For this, many explanations are possible, so that we abstain from further discussion on this as it would result in speculation.

We conclude this section with two remarks: if you want to perform a Hausman-type test, note that the attractive decomposition of the covariance matrix of the squared difference in parameters under fixed versus random effects estimation which we have in the classical Hausman test is no longer possible. Moreover, the test statistic and its variance are influenced by the nonparametric smoother for  $\psi$ . Therefore, the distribution of the test statistic has to be estimated by resampling. We implemented such a test based on wild bootstrap for the simple case when using log-linear transformation of the model instead of PPML. It turned out that already in that case, the variance estimate for the test statistic was little reliable unless sample size was huge.

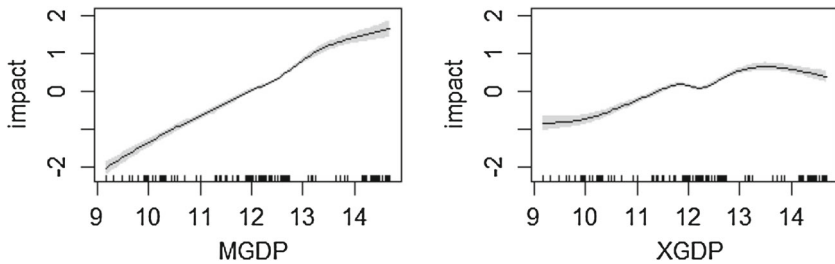
For model selection, one would need an objective (function) such as the prediction quality and likelihood maximization. Standard software provides the log-likelihood, AIC, and BIC. However, for the quasi-likelihood methods, the use of the log-likelihood as a quality measure (to compare models) is disputed. Using the popular AIC and BIC criteria, Model 4a is the best, and all of the “a” models outperforms the “b” and “c” versions. The latter two have more or less the same AIC, BIC, and log-likelihood. All in all, it is clear that the more flexible the model, the better the data fit; prediction qualities have not been studied. On the other hand, the determination of the degrees of



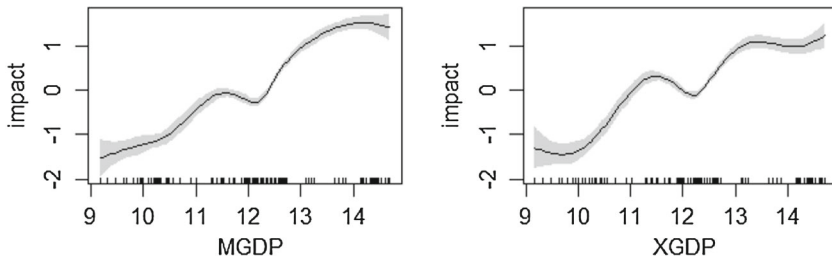
**Table 4** Estimates of coefficients and random effects standard deviations

	Model 4a $\psi_{ij} + u_{ij}$	Model 4c $\psi_{it}^1 + u_{it} + \psi_{jt}^2 + v_{jt}$
MGDP	n.p.	n.p.
XGDP	n.p.	n.p.
MPOP	0.003 (0.004)	0.022 (0.022)
XPOP	0.128 (0.039)**	0.305 (0.046)**
DIST	-1.202 (0.060)**	-0.717 (0.024)**
NEIGH	0.372 (0.101)**	0.347 (0.027)**
EU15	0.274 (0.079)**	0.748 (0.056)**
MLOCK	-0.082 (0.085)	0.235 (0.084)**
XLOCK	0.258 (0.103)*	0.093 (0.085)
2005	0.032 (0.005)**	0.042 (0.031)
2006	0.105 (0.007)**	0.064 (0.030)*
2007	0.139 (0.011)**	0.110 (0.038)*
mMGDP	-	-
mXGDP	-	-
mMPOP	-	-
mXPOP	-	-
M/XAREA	n.p.	n.p.
mM/XPAT	n.p.	-
M/XPAT	-	n.p.
$\sigma_u$ or $\{\sigma_u, \sigma_v\}$	0.578	{0.037,0.0001}

Temporal means of time-varying regressors are denoted by an “m” in front of the original name of the variable. W refers to the vector of size proxies (area and patents). Standard errors are in (·) brackets  
*n.p.* nonparametric estimated; some of the corresponding estimates are shown in Figs. 5 and 6  
 \* Significance at the 5%, \*\* at the 1% level



**Fig. 5** Estimates for the nonparametric impact of MGDP and XGDP in Model 4a with 95% confidence intervals



**Fig. 6** Estimates for the nonparametric impact of MGDP and XGDP in Model 4c with 95% c.b

freedom for non- and semiparametric estimators is still an unsettled matter that affects also the reliability of AIC and BIC.<sup>10</sup>

### 3.3 Empirical findings

In general, our results on the effects of the determinants of trade flows are consistent with theory. The increase in importer and exporter GDP leads to a significant increase in trade while the effect of population is much less important. These results indicate that it might be more appropriate to include GDP and population separately instead of using *GDP per capita*. Distance has the negative effect documented in the literature even after accounting for contiguity of trading partners. Inside the EU, it is irrelevant for trade whether the trading partners share the same official language or ethnic groups. Countries belonging to the European 15 still trade more on average (among themselves), and this effect is relatively large, although it is expected that with deepening economic integration, this gap will decrease. The effect of landlockedness is ambiguous. Countries with access to the sea trade generally more with overseas partners than landlocked countries do. For the EU, this can lead to the effect that relative to their GDP, landlocked EU members trade more with their EU partners. Fixed time effects are all positive and increasing during the considered time period.

It is interesting to note that when the effects of importer and exporter GDP are estimated nonparametrically, they exhibit significant nonlinearities. This questions the usual hypothesis of constant elasticities. While the SMEM estimates with country-pair REs are consistent with the (benchmark) FEs model, this is less the case for models with directional effects. Models with country-specific fixed and/or random effects (i.e.,  $u_i + u_j$ , not shown here) give quite different, partly implausible estimates. These findings deserve to be further investigated in the future; namely, it would be interesting to analyze how well a constant elasticity estimate approximates the average elasticity.

## 4 A simulation-based comparison study

In this section, we present a simulation study to analyze the performance of our method in a situation close to the real data we are dealing with in the empirical study. Therefore, the explanatory variables in the simulated data are fixed and equal to the values,

<sup>10</sup> See Racine and Parmeter (2014) for a method of model selection based on in-sample prediction that accommodates non- and semiparametric estimations and panel data.

respectively, observed in the data. Consequently, the sample size is fixed at 2,400. In order to simplify the calculations in the simulations, we choose as explanatory variables only a subset of the variables used in the empirical study: namely, the GDP of the importer country ( $Y_{it}$ ), the GDP of the exporter country ( $Y_{jt}$ ), and the distance between both countries ( $Z_{ij}$ ). As proxy variables for the nonparametric filter, we consider the area of the exporter country, denoted by  $W_{j1}$ , and the average number of patents of the importer country, denoted by  $W_{i2}$ .<sup>11</sup> We consider two different settings to generate the dependent variable  $T_{ijt}$ .

We are further limiting the simulation study to cases where the data include an unobserved heterogeneity term that is nonlinear in the proxy variables. The goal was to analyze the behavior of estimators that ignore the existence of this term or intend to capture it using the averages of the variables that are time-varying following [Mundlak \(1978\)](#). We compare parametric approaches where the Mundlak variables enter linearly in the index function with our more flexible semiparametric approach where we consider unknown functions of these variables using an additive model. For the sake of comparison, we also study the performance of the semiparametric model using the true proxies and we expect this will show the best performance. The reason to limit the study in this way is that in [Lombardía and Sperlich \(2012\)](#), the superiority of semi-parametric mixed effects models has already been demonstrated for the much simpler case of (log-)linear random effect models with complex correlation structure, i.e., a  $\psi(\cdot)$  that is not additive linear. It is furthermore clear that a nonparametric  $\psi(\cdot)$  can filter out potential dependency between covariates and random effects if the  $W_{ij}$  are just some available proxies for the (unobserved, maybe unknown) real drivers of this dependence. We concentrate here on the comparison of cases where differences and improvements might be less clear.<sup>12</sup> We limit our presentation here to the case where the parametric models with Mundlak device are expected to perform adequately by construction.

The first setting introduces an additive error in the gravity model, i.e.,

$$T_{ijt} = \exp[\beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) + \eta_{ij}] + \epsilon_{ijt} \quad (9)$$

where coefficients  $\beta_1$ ,  $\beta_2$ , and  $\gamma$  are set to be equal to 0.8, 0.25, and  $-1.2$  with intercept  $\beta_0 = -4.5$  such that the mean of  $T_{ijt}$  is close to the mean of the dependent variable observed in our data and the standard deviation is approximately equal to what we actually observe in the data. The unobserved heterogeneity term is constructed as

$$\begin{aligned} \eta_{ij} = & \alpha_1[0.5(W_{j1} - 11.52)^2 + 0.5(W_{j1} - 11.52)^3 + 12] \\ & + \alpha_2[0.5(W_{i2} - 5.67)^2 + 0.5(W_{i2} - 5.67)^3 + 12] + u_{ij} \end{aligned}$$

with  $\alpha_1 = \alpha_2 = 0.05$ , and  $u$  being an independent normal variable with mean zero and standard deviation 0.25. In other words, in the simulation, this drives the dependence

<sup>11</sup> This selection is arbitrary and only to simplify the simulation study; it is neither a selection based on the findings from the previous sections nor a recommendation for empirical studies.

<sup>12</sup> Note that we conducted many more simulation studies, available from the authors, which were even more in favor of the SEM approach.

between explanatory variables and REs, with a high level of correlation with  $\log(Y_{it})$  and a very small one with the other explanatory variables. The  $W$  are also correlated with the time means of  $\log$  GDP, so that the Mundlak device alone should capture well the dependence between the explanatory variables and the unobserved heterogeneity term. The error is generated by

$$0.15\xi \cdot \exp \left[ \beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) + \alpha_1 W_{j1}^* + \alpha_2 W_{i2}^* \right] \quad (10)$$

with  $\xi$  being Chi-squared distributed with one degree of freedom,  $W_{j1}^* = 0.5(W_{j1} - 11.52)^2 + 0.5(W_{j1} - 11.52)^3 + 12$ , and  $W_{i2}^* = 0.5(W_{i2} - 5.67)^2 + 0.5(W_{i2} - 5.67)^3 + 12$ . Note that  $\epsilon_{ijit}$  is heteroscedastic with a variance that depends on the explanatory variables and our proxies via the conditional mean of  $T_{ijit}$ . One might think that the motivation for the construction (10) is obvious but is typically neglected in simulation studies: we need an asymmetric error term to ensure we do not generate negative trade flows. The variance is a function of the mean because the larger the average trade flow for given covariates, the larger but less asymmetric is the error dispersion.

The other setting assumes a multiplicative error  $\epsilon_{ijit}$  leading to

$$T_{ijit} = \exp \left[ \beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) + \eta_{ij} \right] \epsilon_{ijit},$$

where coefficients  $\beta_1, \beta_2$ , and  $\gamma$  are set to the same values as before,  $\beta_0 = -4.25$ ,  $\eta_{ij}$  is generated as in the precedent setting, and  $\epsilon_{ijit}$  is drawn from a Gamma distribution with mean equal to one and variance equal to

$$\text{Var} (T_{ijit} | Y_{it}, Y_{jt}, Z_{ij}, W_{1i}, W_{2j}) = \exp[\beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) + \alpha_1 W_{j1}^* + \alpha_2 W_{i2}^* + 1].$$

This choice of the error distribution is convenient for a multiplicative but heteroscedastic error such that the variance is related to the mean function and leads to one of the primary arguments in Santos Silva and Tenreyro (2006) for the PPML. We have checked for both models that the resulting distributions for simulated  $T$  are comparable to the distribution of the observed trade numbers.

We evaluate our different estimators using 1,000 simulations for each model. For each setting, five different estimation procedures were applied

1. A random effects estimator with the Mundlak device of the log-linear model (REOLSM)

$$\log(T_{ijit}) = \beta_0 + \beta_1^* \log(Y_{it}) + \beta_2^* \log(Y_{jt}) + \gamma \log(Z_{ij}) + b_1^* \overline{\log(Y_i)} + b_2^* \overline{\log(Y_j)} + \phi_{ij} + \tau_{ijit}$$

with  $\overline{\log Y_k} = (1/4) \sum_{t=2004}^{2007} \log Y_{kt}$  for  $k = i, j$ .

2. A pooled PPML of

$$T_{ijit} = \exp \left[ \beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) \right] + \varpi_{ijit}.$$

**Table 5** Performance of parametric estimators in simulations for additive error

	REOLSM			PPML			REPPMLM		
	$\beta_1$	$\beta_2$	$\gamma$	$\beta_1$	$\beta_2$	$\gamma$	$\beta_1$	$\beta_2$	$\gamma$
Mean	0.7974	0.2531	-1.2806	1.7321	0.2283	-1.2415	0.8116	0.2439	-1.293
Bias	-0.0026	0.0031	-0.0806	0.9321	-0.0217	-0.0415	0.0116	-0.0061	-0.093
SD	0.0369	0.0359	0.0152	0.042	0.0214	0.0526	0.2304	0.1385	0.0167
RMSQE	0.0370	0.0361	0.0820	0.9330	0.0305	0.067	0.2306	0.1386	0.0945
Count	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

REOLSM, PPML, and REPPMLM stand for, respectively, random effects OLS of log-linear model with Mundlak device, pooled PPML and random effects PPML with Mundlak device

3. A random effects PPML (REPPMLM) with Mundlak device of

$$T_{ijt} = \exp \left[ \beta_0 + \beta_1^* \log(Y_{it}) + \beta_2^* \log(Y_{jt}) + \gamma \log(Z_{ij}) + b_1^* \overline{\log(Y_i)} + b_2^* \overline{\log(Y_j)} + \phi_{ij} \right] + \omega_{ijt}$$

with  $\phi_{ij} \sim N(0, \sigma_\phi^2)$ .

4. The misspecified semi-mixed effects PPML (SPPPML<sub>Miss</sub>) of

$$T_{ijt} = \exp \left[ \beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) + \psi_1 \left[ \overline{\log(Y_i)} \right] + \psi_2 \left[ \overline{\log(Y_j)} \right] + \phi_{ij} \right] + \omega_{ijt}^*$$

with  $\psi_1(\cdot)$  and  $\psi_2(\cdot)$  estimated by penalized splines and  $\phi_{ij} \sim N(0, \sigma_\phi^2)$ .

5. The correctly specified semi-mixed effects PPML (SPPPML<sub>Well</sub>) of

$$T_{ijt} = \exp \left[ \beta_0 + \beta_1 \log(Y_{it}) + \beta_2 \log(Y_{jt}) + \gamma \log(Z_{ij}) + \psi_1 (W_{j1}) + \psi_2 (W_{i2}) + \phi_{ij} \right] + \tau_{ijt}^*$$

with  $\psi_1(\cdot)$  and  $\psi_2(\cdot)$  estimated by penalized splines and  $\phi_{ij} \sim N(0, \sigma_\phi^2)$ .

The terms  $\tau_{ijt}$ ,  $\tau_{ijt}^*$ ,  $\omega_{ijt}$ ,  $\omega_{ijt}^*$  are the resulting model deviations (errors).

The results from the simulation study for the additive error setting are summarized in Tables 5 and 6 for the parametric and semiparametric estimators, respectively. They include the mean estimate, along with the biases, standard errors, and root mean squared errors of the estimators for  $\beta_1$ ,  $\beta_2$ , and  $\gamma$  of model (9) estimated with each of the above five procedures. Recall that in this setting, inconsistency in estimation may arise due to the correlation between the unobserved heterogeneity term and the explanatory variables, especially concerning  $\log(Y_{it})$ . Therefore, it is not surprising the bad performance of the pooled PPML (the only one that neglects unobserved heterogeneity) on estimating  $\beta_1$ . The random effects estimator of the log-linear model with Mundlak device shows (as expected) good performance whether the bias or the root mean squared error is concerned. The semiparametric estimator (whether using

**Table 6** Performance of semiparametric estimators in simulations for additive error

	SPPPML <sub>Miss</sub>			SPPPML <sub>Well</sub>		
	$\beta_1$	$\beta_2$	$\gamma$	$\beta_1$	$\beta_2$	$\gamma$
Mean	0.8036	0.2458	-1.1836	0.8166	0.2473	-1.1959
Bias	0.0036	-0.0042	0.0164	0.0166	-0.0027	0.0041
SD	0.2267	0.1372	0.0191	0.1048	0.0197	0.0451
RMSQE	0.2268	0.1373	0.0252	0.1061	0.0199	0.0453
Count	1,000	1,000	1,000	1,000	1,000	1,000

SPPPML<sub>Miss</sub> and SPPPML<sub>Well</sub> stand for, respectively, semi-mixed effects PPML with Mundlak variables as proxies and semi-mixed effects PPML using the true proxies

**Table 7** Performance of parametric estimators in simulations for multiplicative error

	REOLSM			PPML			REPPMLM		
	$\beta_1$	$\beta_2$	$\gamma$	$\beta_1$	$\beta_2$	$\gamma$	$\beta_1$	$\beta_2$	$\gamma$
Mean	1.5040	0.1374	-1.5403	1.7283	0.2260	-1.2384	0.7967	0.2508	-1.2921
Bias	0.7040	-0.1126	-0.3403	0.9283	-0.0240	-0.0384	-0.0033	0.0008	-0.0921
Median	1.4974	0.1477	-1.5375	1.7269	0.2250	-1.2367	0.7947	0.2508	-1.2918
SD	0.3829	0.1943	0.0391	0.0431	0.0228	0.0497	0.0579	0.0254	0.0199
RMSE	0.8014	0.2246	0.3425	0.9293	0.0331	0.0628	0.0580	0.0254	0.0942
Count	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

REOLSM, PPML, and REPPMLM stand for, respectively, random effects OLS of log-linear model with Mundlak device, pooled PPML and random effects PPML with Mundlak device

the Mundlak variables or the true proxies) shows the best results in estimating the coefficient of distance,  $\gamma$ . Also, the misspecified semiparametric estimator using the Mundlak variables as proxies instead of the true ones depicts a better performance than the random effects PPML with Mundlak device. Overall, the best performance in estimation is achieved with the semiparametric estimator with the true proxies, as expected, to control for unobserved heterogeneity.

Tables 7 and 8 show the results obtained in the simulations with the multiplicative error. Recall that in this setting, there are two possible sources of endogeneity: one due to the correlated unobserved heterogeneity term, as in the setting analyzed before, and the other induced by the log-linear transformation, concerning the dependency of the idiosyncratic error term with the explanatory variables due to the presence of conditional heteroscedasticity in the multiplicative error. This last situation explains the disastrous performance in terms of bias of the log-linear model confirming the findings of Santos Silva and Tenreiro (2006). But note also that for the first simulation study, OLS gave much better results than the PPML, although the data were generated from a model with heteroscedastic errors. The pooled PPML shows a large bias in estimating  $\beta_1$  as before. For the other estimators, the results lead to the same conclusions as obtained in the additive error setting.

**Table 8** Performance of semiparametric estimators in simulations for multiplicative error

	SPPPML <sub>Miss</sub>			SPPPML <sub>Well</sub>		
	$\beta_1$	$\beta_2$	$\gamma$	$\beta_1$	$\beta_2$	$\gamma$
Mean	0.7935	0.2516	-1.1957	0.7995	0.2486	-1.1961
Bias	-0.0065	0.0016	0.0043	-0.0005	-0.0014	0.0039
Median	0.7915	0.2516	-1.1966	0.7980	0.2490	-1.1961
SD	0.0576	0.0253	0.0208	0.0369	0.0110	0.0242
RMSE	0.0579	0.0253	0.0212	0.0369	0.0111	0.0245
Count	1,000	1,000	1,000	1,000	1,000	1,000

SPPPML<sub>Miss</sub> and SPPPML<sub>Well</sub> stand for, respectively, semi-mixed effects PPML with Mundlak variables as proxies and semi-mixed effects PPML using the true proxies

To summarize, the simulations show the good performance of the semiparametric models in both settings considered, being noticeable better in estimating the coefficient of distance. When the true proxies are used in the nonparametric part that controls the unobserved heterogeneity, the semiparametric approach shows a clear superiority. The Mundlak variables seem to be successful in controlling the potential endogeneity due to the dependence of the unobserved heterogeneity term with the explanatory variables; so that it should always be part of the proxy set  $W$ , no matter whether they enter linearly or via  $\psi_s$ . However, the parametric random effects PPML with Mundlak device are not effective to avoid bias in the estimation of the coefficient of distance. This suggests that, as expected, the semiparametric approach outperforms existing methods mitigating bias of the coefficients of time-invariant variables.

## 5 Concluding remarks

In this paper, we have introduced a methodology to estimate a bilateral gravity model for panel data. We discussed several reasons why, especially for short-time panels, FE regression is not appropriate to estimate trade flows (at least within a nonlinear model). On the other hand, RE and pooled Poisson regression ignore potential correlation between country unobserved effects and explanatories that may lead to inconsistent estimation. Our approach is a good compromise and has been shown to give sensible results. As a by-product, one can also explore and handle the problem of further model specification issues.

We propose the use of a SMEM with some proxy variables to filter nonparametrically the non-observable effects (country-pair or country-specific, etc.) that are assumed to be correlated with the covariates. Given that these proxies' impact is modeled nonparametrically, this model class nests the two extreme cases of FE and REs models. If the temporal means are included in the set of proxies, then it also nests the Mundlak device. In addition, it allows for the direct identification of coefficients of time-invariant variables as long as  $\varsigma > 0$ . The generally resulting model is a semi-MEM in the sense that it has still a residual RE component.

The SMEM can be used to extend the suggestion of Santos Silva and Tenreiro (2006) to use PPML estimation for gravity models or any other nonlinear approach without running into the so-called incidental parameter problem. Though we introduced the model class and estimator directly for the case of panel data analysis, one should emphasize that it works equally well for other kinds of data (unbalanced panels, cohorts, cross-sectional, etc.). It is known that the introduction of an additional randomness in (pseudo-)Poisson models can substantially increase the efficiency of estimation and can easily be combined with zero inflation (a generalization of the Tobit approach to treat censored responses). Finally, the introduced principle can straightforwardly be extended to random coefficients and would then also nest the correlated random coefficients models (Wooldridge 2005).

The presentation and claims are underpinned by a simulation study. The method has been applied to study the trade flows among the EU25 countries from 2004 through 2007. Results show that it lead to the estimates of the effects of time-varying variables that are very close to the FE PPML with the advantage of simultaneously providing estimates of the constant in time variables. Moreover, with our semiparametric approach, we could estimate, for example, importer and exporter GDP elasticities controlling for importer and exporter time-varying unobserved effects. Results concerning the nonparametric filters and respective confidence bands show important nonlinearities that cannot be revealed with a parametric PPML.

It should finally be emphasized that, though the idea and model class are completely new, the proposed approach is readily available as **R** and *stata* provide built-in commands that help the practitioner to estimate these kinds of models in an easy-to-handle way. However, further research is necessary to improve the construction of robust standard errors and offer endogeneity tests. Also, the development of further smoothing parameter selection procedures that are tailored toward the filtering idea could be of interest.

**Acknowledgments** Isabel Proença is grateful for the financial support received from FCT (Fundação para a Ciência e Tecnologia) through the PEst-OE/EGE/UI0491/2011 program and Stefan Sperlich for funding from the Swiss National Science Foundation 100018-140295. We thank helpful discussions with Jaya Krishnakumar, Inmaculada Martínez-Zarzoso, Marcelo Olarreaga, and Walter Zucchini, of participants of ESEM/EEA-2012, the seminars at ISEG and GSEM, an anonymous referee, and in particular the AE Christopher Parmeter who essentially contributed to the stepwise improvement of this article.

## Appendix: Addition information on used data

**Countries included in the data set:** Austria (AU), Belgium (BE), Bulgaria (BU), the Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (GR), Hungary (HU), Ireland (IR), Italy (IT), Latvia (LV), Lithuania (LH), Luxembourg (LU), the Netherlands (NE), Poland (PL), Portugal (PT), Romania (RO), Slovakia (SK), Slovenia (SV), Spain (SP), Sweden (SW), and the United Kingdom (UK).



**Table 9** Pooled PPML

	MGDP	XGDP	MPOP	XPOP	DIST	NEIGH
Coeff.	0.624	0.692	0.079	0.009	-0.623	0.509
SE	0.027**	0.038**	0.026**	0.039	0.028**	0.039**
	EU15	MLOCK	XLOCK	2005	2006	2007
Coeff.	1.57	0.068	-0.015	0.023	0.047	0.045
SE	0.058**	0.051	0.053	0.039	0.038	0.037

A pooled PPML regression without random or fixed effects: estimates of the coefficients with standard errors

### Details about the used variables:<sup>13</sup>

- T: nominal import (cif) flows in  $10^3$  euros.
- MGDP/XGDP: importer/exporter country's nominal GDP at market prices in millions of euro, from Eurostat's New Cronos Database.
- MPOP/XPOP: importer/exporter country's population, expressed in thousands of people at the end of the period, from Eurostat's New Cronos Database.
- DIST: absolute distance expressed in kilometers, the geodesic distance between capitals (in the case of the Netherlands, Amsterdam substitutes Den Haag), measured as the surface distance between two points of latitude and longitude (great circle distance) obtained from [www.wcrl.ars.usda.gov/cec/java/lat-long.htm](http://www.wcrl.ars.usda.gov/cec/java/lat-long.htm).
- NEIGH: neighboring dummy variable is equal to one if two trading partners share a land or sea border, zero otherwise. From CIA's World Factbook 2003 as published on [www.cia.gov/cia/publications/factbook/index.html](http://www.cia.gov/cia/publications/factbook/index.html).
- EU15: dummy equal to 1 if the exporting country belongs to the European 15.
- MLOCK/XLOCK: landlockedness dummy for importer/exporter country; equals one if country has no direct connection to sea.
- COMLANG: common language dummy variable is equal to one if two trading partners share the same official language, zero otherwise. From CIA's Factbook 2003 on [www.cia.gov/cia/publications/factbook/index.html](http://www.cia.gov/cia/publications/factbook/index.html).
- ETHNIC: ethnic dummy, equal to one if there is an ethnic minority of the exporter country in the importer country that represents more than 5% of total population of the latter. From CIA's The World Factbook 2003.
- MAREA/XAREA: importer/exporter country's area.
- PAT: the number of patents of importer country recorded as EPO (European patent office) patent applications (Direct EPO filings + EURO-PCT in regional phase); source OECD (Table 9).

### References

- Anderson JE (1979) A theoretical foundation for the gravity equation. *Am Econ Rev* 69:106–116
- Anderson JE, Van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. *Am Econ Rev* 93:170–192

<sup>13</sup> We thank Enrique Martínez-Galán and Eliano Marques for their help in obtaining these data.

- Anderson JE, Yotov YV (2009) The changing incidence of geography. *Am Econ Rev* 100:2157–2186
- Baltagi HB, Egger P, Pfaffermayr M (2003) A generalized design for bilateral trade flow models. *Econ Lett* 80:391–397
- Bergstrand JH (1985) The gravity equation in international trade: some micro-economic foundations and empirical evidence. *Rev Econ Stat* 67:474–481
- Blackburn M, Neumark D (1992) Unobserved ability, efficiency wages, and interindustry wage differentials. *Q J Econ* 107:1421–1435
- Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 88:9–25
- Chamberlain G (1982) Multivariate regression models for panel data. *J Econ* 18:5–46
- Deardorff AV (1998) Determinants of bilateral trade: does gravity work in a neoclassical world? In: Frankel JA (ed) *The regionalization of the world economy*. University of Chicago Press, Chicago, pp 7–22
- Egger P, Pfaffermayr M (2003) The proper panel econometric specification of the gravity equation: a three-way model with bilateral interaction effects. *Empir Econ* 28:571–580
- Gomez-Herrera E (2013) Comparing alternative methods to estimate gravity models of bilateral trade. *Empir Econ* 44:1087–1111
- Green PJ, Silverman BW (1994) *Nonparametric regression and generalized linear models*. Chapman & Hall, London
- Gurmu S, Rilstoneb P, Sternc S (1999) Semiparametric estimation of count regression models. *J Econ* 88:123–150
- Henderson DJ, Millimet DL (2008) Is gravity linear? *J Appl Econ* 23:137–172
- Kauermann G (2005) A note on smoothing parameter selection for penalized spline smoothing. *J Stat Plann Inference* 127:53–69
- Köhler M, Schindler S, Sperlich S (2014) A review and comparison of bandwidth selection methods for Kernel regression. *Int Stat Rev* 82:243–274
- Krivobokova T, Kneib T, Claeskens G (2010) Simultaneous confidence bands for penalized spline estimators. *J Am Stat Assoc* 105:852–863
- Lin X, Carroll RJ (2006) Semiparametric estimation in general repeated measures problems. *J Royal Stat Soc B* 68:69–88
- Lin X, Zhang D (1999) Inference in generalized additive mixed models by using smoothing splines. *J R Stat Soc B* 55:381–400
- Lombardía MJ, Sperlich S (2008) Semiparametric inference in generalized mixed effects models. *J R Stat Soc B* 70:913–930
- Lombardía MJ, Sperlich S (2012) A new class of semi-mixed models and its application in small area estimation. *Comput Stat Data Anal* 56:2903–2917
- Martinez-Zarzoso I (2013) The log of gravity revisited. *Appl Econ* 45:311–327
- Mátyás L (1997) Proper econometric specification of the gravity model. *World Econ* 20:363–368
- McCullagh P, Nelder JA (1989) *Generalized linear models*, 2nd edn. Chapman and Hall, London
- Mundlak Y (1978) On the pooling of time series and cross section data. *Econometrica* 46:69–85
- Nelder JA (2000) Quasi-likelihood and pseudo-likelihood are not the same thing. *J Appl Stat* 27:1007–1011
- Nie Z, Racine J (2012) The crs package: nonparametric regression splines for continuous and categorical predictors. *R J* 4:48–56
- Niedercorn JH, Becholdt BV (1969) An economic derivation of the ‘gravity law’ of spatial interaction. *J Reg Sci* 9:273–282
- Nuroğlu E, Kunst RM (2013) Competing specifications of the gravity equation: a three-way model, bilateral interaction effects, or a dynamic gravity model with time-varying country effects? *Empir Econ*. doi:10.1007/s00181-013-0696-3
- Proença I, Fontoura MP, Martínez-Galán E (2008) Trade in the enlarged European Union: a new approach on trade potential. *Port Econ J* 7:205–224
- Racine JS, Parmeter C (2014) Data-driven model evaluation: a test for revealed performance. In: Ullah A, Racine JS, Su L (eds) *Handbook of applied nonparametric and semiparametric econometrics and statistics*. Oxford University Press, Oxford, pp 308–345
- Santos Silva J, Tenreiro S (2006) The log of gravity. *Rev Econ Stat* 88:641–658
- Serlenga L, Shin Y (2007) Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *J Appl Econ* 22:361–381
- Sperlich S, Lombardía MJ (2010) Local polynomial inference for small area statistics: estimation, validation and prediction. *J Nonparametr Stat* 22:633–648

- Sperlich S, Tjøstheim D, Yang L (2002) Nonparametric estimation and testing of interaction in additive models. *Econ Theory* 18:197–251
- Venables WN, Ripley BD (2002) *Modern applied statistics with S*, 4th edn. Springer, Berlin
- Wellner JA, Zhang Y (2007) Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann Stat* 35:2106–2142
- Westerlund J, Wilhelmsson F (2011) Estimating the gravity model without gravity using panel data. *Appl Econ* 43:641–649
- Wood SN (2006) *Generalized additive models: an introduction with R*. Chapman and Hall/CRC Press, London
- Wood SN (2008) Fast stable direct fitting and smoothness selection for generalized additive models. *J R Stat Soc B* 70:495–518
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *J R Stat Soc B* 73:3–36
- Wooldridge JM (2005) Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Rev Econ Stat* 87:385–390
- Zheng X (2008) Semiparametric Bayesian estimation of mixed count regression models. *Econ Lett* 100: 435–438