Comput Stat (2015) 30:865–884
DOI 10.1007/s00180-015-0556-z

ORIGINAL PAPER

Modelling bilateral intra-industry trade indexes with panel data: a semiparametric approach

Isabel Proença · Horácio C. Faustino

Received: 30 December 2013 / Accepted: 14 January 2015 / Published online: 1 February 2015

Abstract This paper focuses on the modelling of bilateral intra-industry trade indexes with panel data, applying a semiparametric approach. This extends the work of Papke and Wooldridge (J Econom 145:121–133, 2008) for fractional responses, by introducing a nonparametric component to control for unobserved heterogeneity associated with the regressors. The proposed approach is based on the semi-mixed effects generalised linear model of Lombardía and Sperlich (Comput Stat Data Anal 56:2903–2917, 2012), introduced in the context of small area statistics, and the semiparametric gravity model of Proença et al. (Empir Econ doi:10.1007/s00181-014-0891-x, 2014). The resulting nonlinear semiparametric model serves to explain the bilateral intra-industry trade indexes between Portugal and the European Union, the BRIC emerging economies, and the five Portuguese-speaking African countries.

Keywords Fractional data · Penalised splines · Intra-industry trade · Panel econometrics

1 Introduction

In economics and statistics, there is a frequent need to explain variables limited to the interval between zero and one, which are also known as fractional variables. Examples from economics are indexes, e.g. as intra-industry trade; or proportions, e.g. partic-

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

H. C. Faustino
ISEG, School of Economics and Management,
Universidade de Lisboa and SOCIUS, Lisbon, Portugal

ipation rates in voluntary pension plans, company capital structures, student failure rates and the proportion of income spent on medicines, amongst many others.

The particular difficulty in modelling these types of variables stems from finding models that retain the fitted values within the unit interval and thus exclude the popular linear model. Therefore, the traditional procedure was based on applying the log-odds transformation to the fractional variable, in order to ensure that the transformed variable may be subject to linear regression. However, this procedure cannot be applied whenever the variable assumes values at the corners of zero and one, and is undesirable whenever the objective involves the inference about the conditional expectation of the original fractional variable. In particular, while estimating the partial effects on the conditional expectation of the transformed variable is not difficult, recovering the partial effect estimates on the conditional expectation of the fractional variable proves anything but obvious. Therefore, recent approaches directly model the former through recourse to nonlinear models and estimating the unknown parameters by applying maximum quasi-likelihood methods, based on the Bernoulli distribution, such as the seminal work of Papke and Wooldridge (1996) for cross-sectional data. Later, Papke and Wooldridge (2008) also extended this approach to panel data.

One advantage of working with panel data is the possibility to control the unobserved heterogeneity specific to the individual unit and constant in time. This fact proves especially important when that unobserved term is correlated with the model's other explanatory variables and therefore potentially leads to inconsistency in estimation. For linear models, the fixed effects transformation, provides consistent estimates by eliminating all variables constant over time, even when the unobserved individual heterogeneity term is associated with model regressors. However, the coefficients of time invariant variables cannot be estimated. With nonlinear models, there are no transformations of the dependent variable removing the constant effects (observed and unobserved), such as the fixed effects or first difference transformations. Furthermore, the usual approach relies on setting parametric assumptions that describe the way which the individual unobserved random effects (unobserved heterogeneity) relate to the model's other explanatory variables. Here, the parametric approaches of Chamberlain (1982, 1984) and Mundlak (1978) are mainly applied, with one example being Papke and Wooldridge (2008). Problems, however, may stem from these parametric assumptions proving too restrictive and thus they can lead to significant biases in estimations. One alternative involves turning to nonparametrics as a way of specifying more general and flexible models. In the context of small area statistics, Lombardía and Sperlich (2008, 2012) introduce a semiparametric mixed effects model, where the unobserved random term is modelled nonparametrically and kernel methods are deployed in its estimations. Proença et al. (2014) extend their procedure to panel data, in order to explain trade flows by the gravity model, but opt for penalised regression splines, which have advantages in inference, namely by obtaining confidence bands for the unknown functions estimated.

This paper applies the semiparametric approaches of Lombardía and Sperlich (2012) and Proença et al. (2014) to the modelling of fractional responses with panel data. The proposed procedure is applied to an empirical example that aims to explain the determinants of Portugal's bilateral intra-industry trade (IIT) with 37 countries. These comprise all the European Union partner-states (EU-27), with Belgium and Luxembourg included as a single entity, the five Portuguese-speaking African countries (hereafter known as PALOPs) i.e. Angola, Cape Verde, Guinea-Bissau, Mozambique and Sao Tome and Principe, as well as Brazil, Russia, India and China, the USA, Moldova and Ukraine. In 2006, these countries not only accounted for 89 % of the immigrants in Portugal, but also 83 % of the country's trade in goods. The inclusion of the PALOP countries in the sample addresses the empirical question of whether the evolution of intra-industry trade is a result of special ties originating from a common language, together with the fact that these countries were former Portuguese colonies.

Our motivation is twofold. Firstly, it is to understand the determinants of Portuguese intra-industry trade, and specifically whether the degree of economic integration helps foster trade. Secondly, it is to examine whether there are advantages in applying the more general and flexible semiparametric procedure proposed in this study relative to the recent parametric procedure applied to modelling fractional data, and specifically whether semiparametric methods can uncover misspecification in parametric assumptions and what is the extent of differences in the estimates returned by the two methodologies. The advantages of semiparametric modeling are well documented in the literature. See, for example, Härdle et al. (2004) for a methodological perspective. There are many applications in a variety of fields where semiparametric models prove valuable. To identify those few that fit a survival model with a random effect, see Slama et al. (2003), and also Shen (2011) for examples of semiparametric transformation models that are used to estimate the distribution function of duration time (the elapsed time between two consecutive events), when data is left-truncated and right-censored. Nott (2006) provides flexible methods to estimate mean and variance of heterogeneous Gaussian data and overdispersed or underdispersed count data, based on penalised splines and Yoshida et al. (2010) introduce a computationally efficient generalised information criteria for model selection of generalised linear mixed models estimated with penalised splines. Finally, Xiaa and Härdle (2006) consider examples of semiparametric modeling applied to credit scoring and environment statistics.

The paper structure is as follows. Section 2 introduces the semiparametric approach to modelling fractional responses, while Sect. 3 focuses on the intra-industry trade problem, including a survey of existing works. Section 4 reports and discusses the results from the empirical application of the models to the intra-industry trade between Portugal and the aforementioned countries. Section 5 concludes.

2 The semiparametric model for fractional responses with panel data

We consider a panel data set constituted by a random sample of N units, where unit i is observed repeatedly over T_i periods of time. For unbalanced panels, we assume that the observations missing do not entail sample selection. For unit i and period t we observe y_{it} with $0 \leq y_{it} \leq 1$. Following Papke and Wooldridge (2008), we thus assume for unit i that,

$$E\left(y_{it}\,|\mathbf{x}_{it}\,,\mathbf{z}_i,\eta_i\right)=\Phi\left(\mathbf{x}_{it}\boldsymbol{\beta}+\mathbf{z}_i\boldsymbol{\gamma}+\eta_i\right),\quad t=1,\ldots,T_i \tag{1}$$

with $\Phi(\cdot)$ the standard normal cumulative distribution function (cdf), $\mathbf{x}_{it}$ a $1 \times k$ vector of explanatory variables that vary in time, $\mathbf{z}_i$ a $1 \times p$ vector of explanatory variables that remain constant in time, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are respectively $k \times 1$ and $p \times 1$ vectors of unknown coefficients, and η_i a heterogeneous unobserved effect. We would point out that (1) is a single index model with a Probit link. The Probit choice essentially stems from estimation convenience on the assumption that the unobserved heterogeneity term is normally distributed according to,

$$\eta_i = \alpha + a_i \quad \text{with} \quad a_i \,|(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, \mathbf{z}_i) \sim N\left(0, \sigma_a^2\right). \tag{2}$$

Then,

$$E\left(y_{it} \,|\mathbf{x}_{it}\,, \mathbf{z}_i\right) = \Phi\left[\left(\alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\gamma}\right)\frac{1}{\sqrt{1+\sigma_a^2}}\right], \quad t = 1, \ldots, T_i \tag{3}$$

with α and σ_a being unknown parameters requiring estimation. We would bring to attention the fact that (3) is the well-known Probit with random effects. However, as the response is fractional, and not binary, estimation of the unknown parameters in (3) should be performed by the maximum quasi-likelihood method.

Unfortunately, Hypothesis (2) proves hard to verify in practice, because η_i is often correlated with the explanatory variables, resulting in inconsistent estimates of the parameters in (3). For example, if we are modelling the corporate capital structure, which is the proportion of capital due to debt or to equity, unobserved effects follow from the socio-cultural company management environment which may depend on firm size and profitability, both of which are determinants of capital structure. In explaining intra-industry trade though, there is limited knowledge of the nature of the unobserved effects and we suspect that it may include aspects of product differentiation, differences in industry technologies and in factor endowments, which may be correlated with the explanatory variables GDP and trade imbalance.

Papke and Wooldridge (2008) make assumptions as to the way η_i depends on the explanatory variables $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}$, based on Mundlak (1978) and resulting in:

$$\eta_i = \alpha + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i \quad \text{with} \quad a_i \,|(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, \mathbf{z}_i) \sim N\left(0, \sigma_a^2\right) \tag{4}$$

where $\bar{\mathbf{x}}_i = (1/T_i)\sum_{t=1}^{T_i}\mathbf{x}_{it}$ is the $1 \times k$ vector of time means of the time varying explanatory variables, and $\boldsymbol{\xi}$ a $k \times 1$ vector of unknown coefficients. Based on (4), we obtain the following random effects Probit model:

$$E\left(y_{it} \,|\mathbf{x}_{i1}\,, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, \mathbf{z}_i\right) = \Phi\left[\left(\alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\gamma} + \bar{\mathbf{x}}_i\boldsymbol{\xi}\right)\frac{1}{\sqrt{1+\sigma_a^2}}\right], \quad t = 1, \ldots, T_i \tag{5}$$

Once again, the unknown α, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{\gamma}$ and σ_a^2 parameters are susceptible to estimation by the maximum quasi-likelihood method.

Nevertheless assumption (4) may prove too restrictive for modelling the dependence between the unobserved heterogeneity and the variables. Proença et al. (2014) addressed this problem by proposing a semi-mixed effects gravity model for panel data. Their approach is based on the work of Lombardía and Sperlich (2008), and Lombardía and Sperlich (2012), who consider a generalised partially linear mixed effects model for a cross-section of observations of small geographical areas and estimation with kernels and profiled likelihood based methods. Proença et al. (2014) consider the particular case of a generalised partially linear mixed effects model with an exponential link in the context of longitudinal data and use quasi-likelihood estimation with penalised splines. The model addressed in this study is another special case of the generalised partially linear mixed effects model for the Probit link and as in Proença et al. (2014), estimation will be performed with panel data using quasi-likelihood methods and penalised splines.

The key assumption in the approach included in the abovementioned studies stems from the existence of a set of proxy variables, $\mathbf{w}_i$, time invariant and continuous, and an unknown function $\psi(\cdot)$, such that $\psi(\mathbf{w}_i)$ is able to filter all the dependency between the unobserved heterogeneity term and the explanatory variables. Hence:

$$\eta_i = \alpha + \psi(\mathbf{w}_i) + a_i \quad \text{with} \quad E(a_i \,|\mathbf{w}_i \,, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, \mathbf{z}_i) = E\,(a_i \,|\mathbf{w}_i\,) = 0 \tag{6}$$

We would note that because $\psi(\cdot)$ remains unknown and the proxy variables are continuous, (6) may accommodate a great variety of situations, especially whenever many proxies are incorporated into the study. However, this approach poses two additional sorts of difficulties. One involves identifying the observable proxy variables, $\mathbf{w}_i$, whilst the other relates to an eventual difficulty in estimating $\psi(\cdot)$, when $\mathbf{w}_i$ is precisely high-dimensional. Concerning the first issue, in order to find out the proxy variables, it is necessary to know very well the source of endogeneity due to the unobserved heterogeneity. In many applications that source is known, but there are no observable proxies to capture it. This is the case, for example, when unobservable heterogeneity is due mainly to a firm's specific management culture environment, because these are socio-cultural features that are hard to quantify. In the empirical problem addressed in this study concerning intra-industry trade, as the sources of unobserved heterogeneity are not well known, although we conjecture that it may include aspects of product differentiation and differences in industry technologies and in factor endowments (as was mentioned above), it is hard to find variables $\mathbf{w}_i$. However, a simple and convenient solution can be found extending the Mundlak (1978) reasoning, which results in recurring to the average of the covariates varying in time, $\bar{\mathbf{x}}_i$, to fulfill the role of the proxies $\mathbf{w}_i$. Given that our proposed method uses a nonparametric filter $\psi(\bar{\mathbf{x}}_i)$, it is more general than the linear specification commonly used in parametric models, and consequently it is expected to lead to less biased estimates. With respect to the drawback in nonparametric estimation due to the curse of dimensionality, specification of $\psi(\cdot)$ can be restricted to an additive model. Following these ideas, we assume in this study that:

$$\eta_i = \alpha + G_1(\bar{x}_{i1}) + G_2(\bar{x}_{i2}) + \cdots + G_k(\bar{x}_{ik}) + a_i$$
$$\text{with} \quad a_i \,|(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) \sim N\left(0, \sigma_a^2\right) \tag{7}$$

with $\bar{x}_{ij}$ being the mean in time of covariate j, that is $\bar{x}_{ij} = (1/T_i)\sum_{t=1}^{T_i} x_{itj}$, and $G_l(\cdot)\;\; l = 1, \ldots, k + L$ are the unknown functions. Empirical applications require some caution when applying this approach, as the covariates may be highly correlated with the respective time means, which may cause inflated standard errors in the fitted semiparametric model.

Incorporating (7) into (1), we obtain a generalised additive mixed model for the conditional expectation of the response fractional variable. Nowadays, generalised additive mixed models are easily estimable through penalised splines. Due to the nature of the dependent variable, once again, the maximum quasi-likelihood method is best recommended, with routines in R available for this purpose. See Wood (2006) for details.

3 Intra industry trade

After the pioneering theoretical work of Krugman (1979, 1980) and Lancaster (1980), empirical analysis of intra-industry trade (IIT) has taken three main paths: (i) the measurement of IIT, by total and by types, where the IIT measure captures two types of trade: horizontal IIT (HIIT), which is trade in differentiated products with similar price ranges; and vertical IIT (VIIT), which is trade in differentiated products, distinguished by quality and price. HIIT may occur between countries with similar factor endowments and is explained by economies of scale and horizontal product differentiation (e.g., different varieties of the same product with similar quality levels), whereas VIIT may occur between countries with different relative factor endowments and is explained according to the Heckscher–Ohlin (HO) theory; (ii) The determinants of IIT, HIIT and VIIT within the framework of combining the hypotheses proposed by the different theories within the same econometric model, namely: the new trade theory (e.g., Krugman 1979, 1980; Lancaster 1980; Ethier 1982; Eaton and Kierzkowski 1984; Helpman and Krugman 1985; Falvey 1981; Flam and Helpman 1987; Falvey and Kierzkowski 1987), the Ricardian theory of comparative advantage (e.g. Davis 1995), the new economic geography (e.g. Krugman 1991a, b), the Linder (1961) theory of representative or overlapping demand (e.g. Bergstrand 1990), the product cycle theory of Vernon (1966) and the fragmentation theory of Jones and Kierzkowski (2001, 2004), and; (iii) the relationship between intra-industry trade and labour adjustment costs, deploying the concept of marginal intra-industry trade to test the smooth adjustment hypothesis: IIT induces a reallocation of production factors that cost less than inter-industry trade (e.g. Brulhart 1994; Brulhart et al. 2006).

The relevance of intra-industry and intra-firm trade in international trade is stressed by the OECD (2002, p. 159), where it is referred to that "the intra-industry trade share of manufacturing trade has increased significantly since the late 1980s across many OECD countries", and that there is also an increasing trend in intra-industry trade for all the major OECD economies between 1970 and 1990. According to the smooth adjustment hypothesis, intra-industry adjustment is costless in terms of labour reallocation and unemployment (Brulhart 1994). Hence, IIT represents a less disruptive means of adjustment more acceptable by trade union negotiators, for example, when discussing the effects of trade on employment.

In order to measure intra-industry trade by types, three alternative measures are considered: the total intra-industry trade index (IIT), the vertical intra-industry trade index (VIIT) and the horizontal intra-industry trade index (HIIT). These are the three variables explained in the empirical study. For industry l, and for partner country i, we thereby attain:

$$IIT_{li} = HIIT_{li} + VIIT_{li} \tag{8}$$

3.1 The Grubel and Lloyd IIT index

The Grubel and Lloyd index (1975) serves to measure the intra-industry trade between Portugal and its trading partner, country i. At the industry level (2 or 3-digit level), there are products that are produced according to different: technologies, factor proportions, economies of scale and product differentiation. As IIT is a two-way trade in similar finished products, we need to carry out the analysis at the product level (at least at a 5-digit level). In the first empirical studies, some authors deployed an 8-digit product level. However, the difference between 8-digit level indices and 5-digit level indices did not attain statistical significance, and the 5-digit SITC (Standard International Trade Classification) level has since been adopted (see, Greenaway et al. 1994).

In order to avoid statistical aggregation issues, the index calculations contain the maximum statistical disaggregation enabled by INE—the Portuguese Institute of Statistics' (five digit) database:

$$IIT_{it} = \frac{\sum_{j=1}^{J}\left(X_{ijt} + M_{ijt}\right) - \sum_{j=1}^{J}\left|X_{ijt} - M_{ijt}\right|}{\sum_{j=1}^{J}\left(X_{ijt} + M_{ijt}\right)} \tag{9}$$

in which X_{ijt} and M_{ijt} are the bilateral exports and imports respectively of Portugal with partner i, regarding the 5-digit product level j (of the CAE-Economic Activities Classification) at time t. This clearly yields $0 \leq IIT_{it} \leq 1$. Fontagné and Freudenberg (1997) propose an alternative measure. Despite the problems in measuring IIT, empirical studies generally apply the Grubel and Lloyd index (e.g. Lloyd and Lee 2002; Lloyd and Grubel 2003). Bilateral indices between Portugal and each country partner account for the weighted averages of the indices calculated at the 5-digit level, with their weightings attributed in accordance to the share of the product trade over total manufacturing trade.

3.2 The HIIT and VIIT indexes

HIIT and VIIT simultaneously define differentiated product imports and exports (5-digit product level). HIIT represent the trade flows of goods produced with similar factor proportions and is not significantly differing in quality, whereas VIIT are trade flows of goods produced with different factor proportions, and hence differ significantly in quality. Thus, empirical studies require a methodology that is able to separate HIIT from VIIT.

This paper applies the Abd-el-Rahaman (1991) and Greenaway et al. (1994) methodology to separate horizontal from vertical intra-industry trade. The unit values of exports (UV^X) related to the unit values of imports (UV^M), defined as the value of trade by tonne, serve to distinguish between VIIT and HIIT. As is common practice, we use a dispersion value of 15 %. Fontagné and Freudenberg (1997) also propose a different method for disengaging HIIT from VIIT. Despite the difference between the IIT definitions, both methodologies adopt a high correlation between prices (export and import unit values) and the quality of traded products. This assumption proves realistic whenever the statistical aggregation is minimised. Hence, HIIT and VIIT are calculated to a five digit disaggregation. We first make the calculations for the 5-digit product categories before the results are then aggregated to make up each 3-digit industry category.

HIIT satisfies the condition for all t:

$$1-\alpha \le \frac{UV^X_{jli}}{UV^M_{jli}} \le 1+\alpha \tag{10}$$

whilst VIIT satisfies the condition for all t:

$$\frac{UV^X_{jli}}{UV^M_{jli}} < 1-\alpha \quad \text{or} \quad \frac{UV^X_{jli}}{UV^M_{jli}} > 1+\alpha \tag{11}$$

where j denotes a 5-digit product, l denotes an industry and i is a trading partner of Portugal. The constant αcan take on any value between 0 and 1.[1]

4 Empirical application

4.1 The variables

The dependent variables are the intra-industry indexes IIT, VIIT and HIIT as defined in the previous section.

The full sample model includes the following control variables:

– ABSDYPC is the absolute difference between per-capita GDP (PPP, in current international 10^3 dollars) of Portugal and the per-capita GDP (PPP) of the respective trading partner. According to the Linder (1961) hypothesis of overlapping demand, this foresees a negative sign for the coefficient of this variable in the IIT equation. The greater the difference between the countries, the lesser is the

[1] Economically, it is reasonable to set the value of this constant equal to 0.15, and this is the option taken in this study. This is the option that is also followed in the large majority of studies.

level of IIT. According to HO theory and VIIT theory, VIIT increases whenever ABSDYPC increases (see, for example, Falvey 1981; Falvey and Kierzkowski 1987; Flam and Helpman 1987). Thus, this VIIT model variable coefficient is forecast as positive. On the other hand, the more similar the countries are in terms of per-capita GDP, the larger is the level of HIIT (Helpman and Krugman 1985; Greenaway et al. 1994, 1995). Therefore, the HIIT model expects a negative result for this variable's coefficient. Thus, considering how IIT encompasses both HIIT and VIIT, should VIIT predominate, then the variable coefficient in the IIT model may turn positive: this thereby becomes a question of the empirical evidence;

- ABSDPOP represents the absolute difference between the populations of Portugal and the foreign country (in thousands). This serves as a proxy for market size (Frankel 1997). This represents a variable that is commonly applied both in empirical studies on intra-industry trade and also in the gravity model. This expects all types of trade to increase, in keeping with increases in population numbers, as higher populations are associated with greater levels of specialisation. However, some authors consider that the signs of population variables are ambiguous, a priori, because we do not know the actual effect on specialisation (inter-industry or intra-industry) when the size of the population varies (see, for example, Gould 1994; Murat and Pistoresi 2009). Therefore, the sign of the ABSDPOP coefficient also proves ambiguous as a matter of empirical evidence.
- ABSCEE becomes a proxy for differences in physical capital endowments, and equals the absolute difference in electric power consumption (10^3 Kwh per capita) between Portugal (CEE) and its international partner (CEEK). The VIIT model expects a positive sign for this variable's coefficient, given that VIIT trade mainly involves components as explained by the HO theory of comparative advantages (e.g. Helpman and Krugman 1985; Deardorff 1998; Jones and Kierzkowski 2001; Zhang et al. 2005). The HIIT model foresees a negative sign to this variable's coefficient, as countries with similar capital endowments benefit from economies of scale (e.g. Helpman and Krugman 1985). In the IIT model, the sign of the coefficient becomes a matter of empirical evidence. Should the VIIT prove predominant, a positive sign is expected for this variable's coefficient. However, Hummels and Levinshon (1995), applying factor ratios, estimated a negative relationship between factor endowment differences and IIT;
- LDIST conveys the logarithm of geographic distance, measured in kilometres, between the capital cities of the trading partners. This provides a proxy variable for transportation costs. Hence, this variable coefficient is forecast to return a negative sign. The same notion of a negative effect becomes reinforced when empirical studies deploy the gravitational equation to explain bilateral trade (see Bergstrand 1985, 1989; Matthews 1998; Clark 2006).
- TIY represents the trade imbalance weighting in the GDP of each trading partner. This variable controls the effects of trade imbalances on all IIT types. Grubel and Lloyd (1975) posit the negative influence of trade imbalances on IIT. The IIT index would be biased downwards whenever there is a trade imbalance. Similarly, Aquino (1978) proposes an adjusted measure for the IIT index, in order to take the trade imbalance into account. The means of correcting the trade imbalance effect on intra-industry trade involves introducing it as a control variable in the

econometric specification. Thus, the expected sign of this variable turns negative for all IIT models.

In addition to these quantitative variables, we also introduce qualitative dummy variables to reflect the impact of countries belonging to the European Union prior to the 2004 enlargement (EU-15), the emerging economies of Brazil, Russia, China and India (BRICS) and also the five Portuguese-speaking African countries (PALOPS).

- EU15 is a dummy variable that assumes the value 1 when the trading partner is an EU15 member and zero otherwise. We expect a positive sign for this variable's coefficient in all equations, as the integration process reinforces the share of intra-industry trade (Verdoorn 1960; Balassa 1966);
- BRICS is a dummy variable that assumes the value 1 when the country is Brazil, Russia, India or China and zero otherwise. This foresees a positive coefficient for this variable in all equations, because IIT is the kind of trade prevailing in the most developed countries, and these emerging economies verge on being considered able to join this club;
- PALOPS is a dummy variable that assumes the value 1 when the trading partner is one of the five Portuguese-speaking African countries considered in this study, and zero otherwise. The expectation is that the same cultural relationship and similar preferences proxied by common language positively influences intra-industry trade, both by total and by types.

Furthermore, there are still other factors that likely influence intra-industry trade that were not observed, and therefore were included in the model as an unobserved random term. According to the reasoning presented in Sect. 2, to control for an eventual dependence of this unobserved term with other explanatory variables, we opt to include the mean of time varying regressors as proxies in the nonparametric control function. These are the same proxies that are also considered by Papke and Wooldridge (2008) based on Mundlak (1978), although they are included in a linear parametric setting. This procedure leads to the following variables:

- MYPCK is the mean for the panel time period for the trading partner's per-capita GDP. This serves as a proxy used in intra-industry trade models to represent dimension. The expected sign for this variable's coefficient is positive in all equations. The underlying hypothesis states that the larger the trading partner is in economic terms, the larger the intra-industry trade should be (Hummels and Levinshon 1995);
- MPOPK is the mean for the panel time period of the trading partner's population (in thousands);
- MCEEK is the mean for the panel time period of the variable CEEK (already defined above);
- MTIY is the mean for the panel time period of the variable TIY (already defined above).

4.2 The data and descriptive statistics

The data set applied in the estimation contains an unbalanced panel of 38 countries (Angola, Austria, Belgium and Luxembourg (the data is combined for these two

Table 1 Descriptive statistics: full sample

Variable	Mean	Std. Dev.	Min	Max	NT
IIT	0.171	0.160	0.001	0.620	427
HIIT	0.034	0.052	0	0.287	430
VIIT	0.137	0.125	0.001	0.494	426
YPCPT	18.500	2.311	14.439	21.943	444
YPCK	15.761	11.496	0.568	51.980	440
ABSDYPC	10.127	5.464	0.278	30.037	440
TIY	−0.042	0.134	−0.795	0.402	413
MTIY	0.087	0.110	0.000	0.795	413
POPPT	10,286.930	191.521	10,027.000	10,589.650	444
POPK	93,275.810	259,305.500	127.508	1,311,798.000	444
ABSDPOP	89,137.900	257,249.800	14.832	1,301,208.000	444
CEEP	3.832	0.494	3.077	4.526	370
CEEK	5.175	3.819	0.032	16.780	340
ABSDCEE	2.808	2.913	0.011	12.391	340
DIST	3,433.749	2,022.810	503.000	9,986.000	444
BRICS	0.108	0.311	0	1	444
PALOPS	0.135	0.342	0	1	444
EU15	0.351	0.478	0	1	444

countries), Brazil, Bulgaria, Cape Verde, China, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Guinea-Bissau, Hungary, India, Ireland, Italy, Latvia, Lithuania, Malta, Moldova, Mozambique, the Netherlands, Poland, Portugal, Romania, Russia, Sao Tome and Principe, Slovakia, Slovenia, Spain, Sweden, UK, Ukraine and USA) observed during the years 1995–2006. Descriptive statistics for the data are set out in Table 1.

As observed in the mean results, the IIT index mainly comprises the VIIT type (IIT = 0.171; HIIT = 0.034; VIIT = 0.137).

Observations with missing values in at least one variable were excluded. Therefore, the unbalanced panel ended up with 329 observations.

Since the set of countries that are in the sample is very heterogeneous (European countries, USA, China, African countries and Brazil) variables such as GDP *per capita* (YPCK), trade imbalance (TIY), population (POPK), electric power consumption (CEEK) and distance, all return significant variations across the sample, as detailed in Table 1. This furthermore helps identify the effects of the explanatory variables on the models estimated.

4.3 Estimation results

This section includes the results of estimating the semiparametric mixed effects model for fractional data, as introduced in Sect. 2, to explain intra-industry trade indexes.

Estimations are obtained by maximising the penalised quasi-likelihood function, using the *gamm* procedure of the *mgcv* R package. Here, the nonparametric estimation is performed by penalised cubic regression splines, following Wood (2006).

Parametric regression is a random effects Probit following Papke and Wooldridge (2008). Here, dependent unobserved heterogeneity is controlled by the Mundlak (1978) device, which includes the mean of the time varying variables as additional regressors. This procedure leads to the following independent variables: MYPCK, MCEEK and MTIY, which are all equal to the mean for the time period of the following respective trading partner variables: GDP per-capita (MYPCK), electric power consumption (MCEEK) and trade imbalances in GDP (MTIY). The average population size (MPOPK) was discarded, due to the lack of variation over these timeframes for all the countries considered, which causes the mean to be very closely correlated with the explanatory variable, detailing the absolute difference in populations between Portugal and the trading partner (ABSDPOP), as set out in Table 3 in the Appendix.

The semiparametric regression is a mixed effects model with a Probit link, which captures the unobserved heterogeneity, by including unknown functions in the indexes returned by the above mentioned mean variables. However, due to numerical problems in estimation caused by concurvity,[2] the MCEEK function was not included in the semiparametric regression for the horizontal intra-trade index. Table 3 in the Appendix shows that MCEEK and ABSDCEE exhibit a correlation of 0.77. In all regressions, we included a time trend to control for fixed time effects.

Table 2 provides the estimation results and Figs. 1, 2 and 3 report the nonparametric estimates, together with the respective confidence bands obtained as collections of pointwise confidence intervals. The aforementioned figures patently display how the impact of the TIY mean is nonlinear and induces misspecification of the parametric model to control for unobserved heterogeneity. On the other hand, the goodness of fit measures indicate that semiparametric regression is a better fit for all indexes, apart from HIIT, even while remaining very close in this latter case. Moreover, the estimated random effect variance is significantly smaller in the semiparametric fit for all indexes. Therefore, the semiparametric fit seems most appropriate for estimating these models, and we consequently base the interpretation of the empirical findings on these respective results.

According to the results in Table 2, the hypothesis of overlapping demand is verified, as the effect of absolute difference in per-capita GDP (ABSDGDP) is negative and statistically significant for IIT and VIIT, and unexpectedly not statistically relevant for determining HIIT. Consequently, the HO theory receives no empirical confirmation from this application. In addition, the absolute difference in the size of the trading partners' populations does not have a significant impact on the intra-industry trade indexes. With regards to the proxy for differences in capital endowment, this does

[2] Concurvity is the nonparametric analogue of multicollinearity and it occurs in an additive model when a nonparametric function estimate is well approximated by the other nonparametric function terms. For more details see Härdle et al. (2004).

Table 2 Parametric and semiparametric panel regressions explaining intra-industry trade indexes

	Parametric IIT			Semiparametric IIT			Parametric VIIT			Semiparametric VIIT			Parametric HIIT			Semiparametric HIIT		
	Coeff.	SE	*p* val.	Coeff.	SE	*p* val.	Coeff.	SE	*p* val.	Coeff.	SE	*p* val.	Coeff.	SE	*p* val.	Coeff.	SE	*p* val.
Intercept	1.5285	1.6414	0.353	1.3856	1.3850	0.318	0.5147	1.5841	0.746	0.5591	1.4019	0.690	0.4946	1.1342	0.663	−0.0977	0.8284	0.906
ABSYPC	−0.0307	0.0092	0.001	−0.0260	0.0089	0.004	−0.0262	0.0097	0.007	−0.0224	0.0095	0.019	−0.0319	0.0115	0.006	−0.0111	0.0091	0.224
ABSPOP	0.0000	0.0000	0.112	0.0000	0.0000	0.549	0.0000	0.0000	0.119	0.0000	0.0000	0.462	0.0000	0.0000	0.551	0.0000	0.0000	0.873
ABSCEE	0.0765	0.0329	0.021	0.0839	0.0321	0.009	0.0460	0.0353	0.194	0.0531	0.0348	0.128	0.1234	0.0463	0.008	0.0365	0.0159	0.022
LDIST	−0.9492	0.4831	0.060	−0.7909	0.4003	0.059	−0.6720	0.4659	0.161	−0.5506	0.4059	0.187	−0.8977	0.3297	0.011	−0.6713	0.2391	0.009
TIY	−0.8974	0.3691	0.016	−0.9077	0.3698	0.015	−1.1187	0.4055	0.006	−1.1247	0.4063	0.006	0.7097	0.7309	0.332	0.6005	0.7609	0.431
BRICS	0.2559	0.3975	0.525	0.3345	0.3231	0.310	0.1793	0.3844	0.645	0.2311	0.3297	0.490	0.4567	0.3183	0.163	0.4468	0.2427	0.077
PALOPS	−0.0388	0.4378	0.930	0.0369	0.3641	0.920	−0.0357	0.4286	0.934	0.0143	0.3744	0.970	−0.2959	0.4312	0.499	0.0713	0.3525	0.841
EU15	0.0315	0.2447	0.899	0.1692	0.2138	0.436	−0.0401	0.2361	0.866	0.0557	0.2161	0.799	0.2601	0.1721	0.143	0.5123	0.1369	0.001
t	0.0257	0.0051	0.000	0.0248	0.0050	0.000	0.0205	0.0055	0.000	0.0198	0.0055	0.000	0.0295	0.0084	0.001	0.0207	0.0081	0.011
MYPCK	0.0556	0.0155	0.001	np			0.0513	0.0150	0.002	np		0.003	0.0422	0.0119	0.002	np		
MCEEK	−0.0888	0.0440	0.054	np			−0.0665	0.0448	0.149	np		0.540	−0.1068	0.0491	0.039			
MTIY	−0.1784	1.2953	0.892	np			0.4229	1.2718	0.742	np			−2.6867	1.2506	0.041	np		
NT	329			329			329			329			329			329		
AIC	97.6			95.3			149.52			150.42			474.8			484.5		
logLik	−33.8			−29.6			−59.76			−57.21			−222.4			−226.3		
D rand	0.3696			0.2881			0.3535			0.2927			0.2266			0.1312		

The parametric regression is a Probit with random effects estimated by quasi-likelihood with R. Variables with acronyms starting with M refer to the mean in time of the variable identified by the acronym's remaining letters. The former are used to control for dependent unobserved heterogeneity, according to the Mundlak (1978) approach. The semiparametric regression results from the index, considering the Probit with random effects additive nonparametric functions of the mean variables identified in the table with NP. These nonparametric components are estimated by penalised splines. Intercepts of the parametric and semiparametric regressions cannot be compared

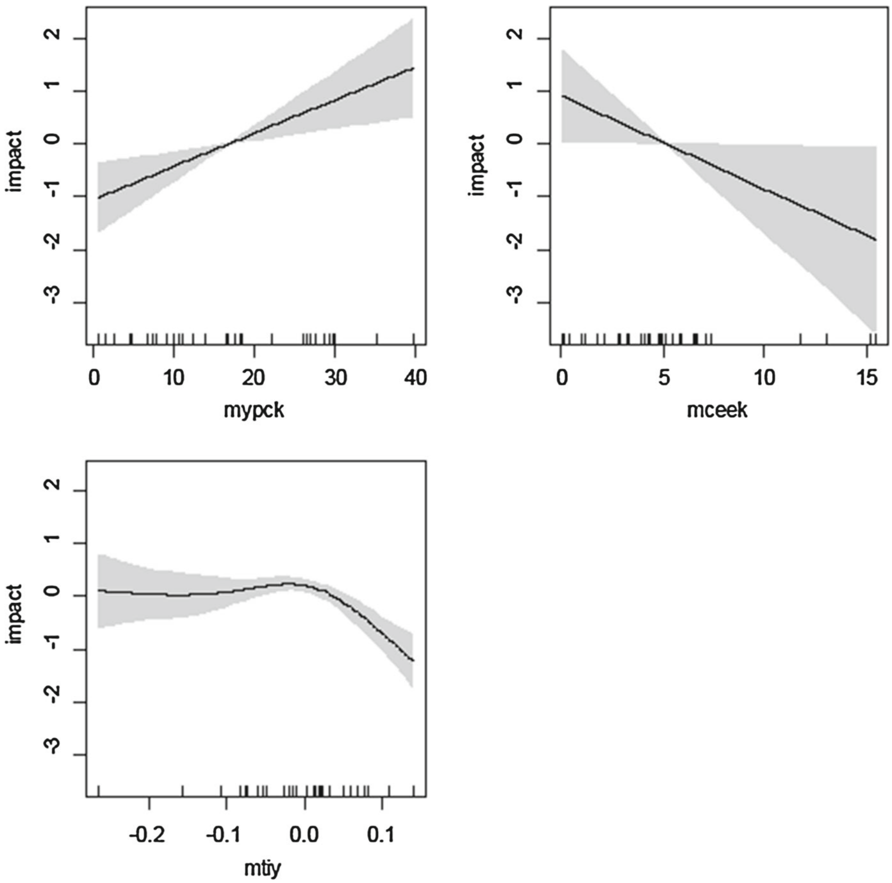

Fig. 1 Estimates of G_1 (*MYPCK*), G_2 (*MCEEK*) and G_3 (*MTIY*) for IIT *Notes* The plots include the respective 95 % confidence bands. The x-axes indicate the automatically chosen spline knots. Observe how the frequency of knots is proportional to that of the underlying observations

not significantly impact on VIIT, which proves a surprising result, as the proxy does impact positively on the IIT and HIIT indexes. The last result is different to that which was forecast by the theory. The coefficient of distance has the expected negative sign, even if it is not statistically significant for VIIT. The higher the trade imbalance (TIY), the smaller the intra-industry trade, except for HIIT, where the effect does not attain statistical significance.

The positive effect of economic integration, measured by the coefficient estimate of the dummy EU15, is only confirmed in horizontal intra-industry trade with the same holding for the BRICS grouping. These results prove to be interesting, as they imply that the effects of intra-industry trade integration only become visible for products of similar quality, whilst the effect of the similitude in economic structure between the emerging economies and Portugal only impacts on products incorporating quality.

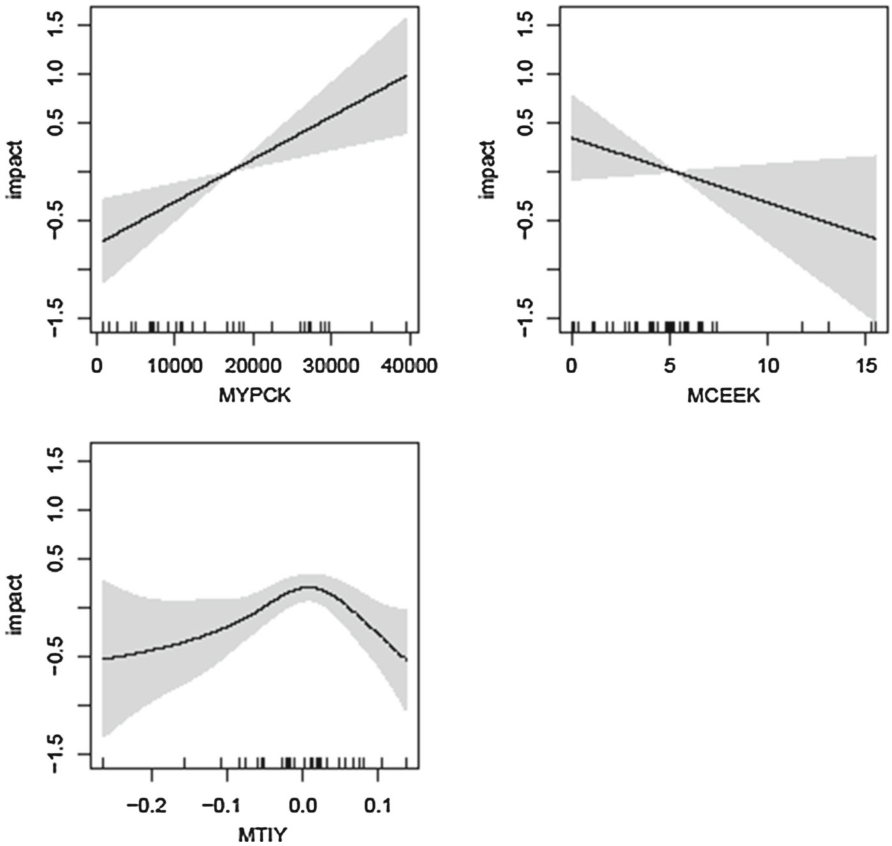

Fig. 2 Estimates of G_1 (*MYPCK*), G_2 (*MCEEK*) and G_3 (*MTIY*) for VIIT *Notes* The plots include the respective 95 % confidence bands. The x-axes indicate the automatically chosen spline knots. Observe how the frequency of knots is proportional to that of the underlying observations

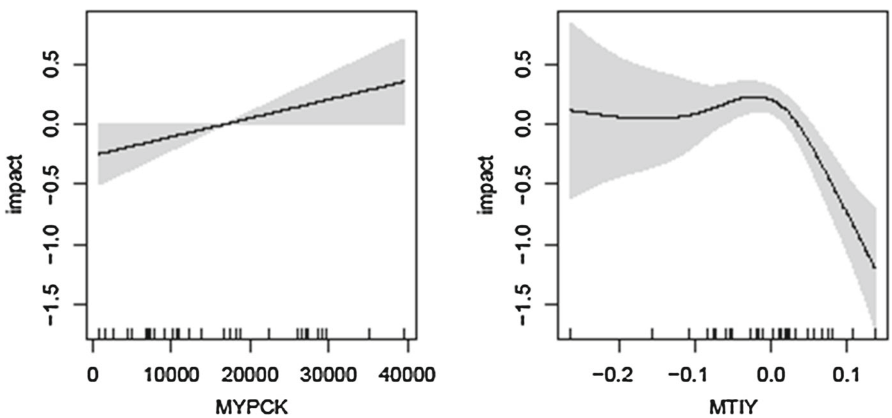

Fig. 3 Estimates of G_1 (*MYPCK*) and G_3 (*MTIY*) for HIIT *Notes* The plots include the respective 95 % confidence bands. The x-axes indicate the automatically chosen spline knots. Observe how the frequency of knots is proportional to that of the underlying observations

In general, there are no important differences between the semiparametric and parametric estimates.[3] However, there are some disparities that are worthwhile noting in that they mainly concern estimates of coefficients of variables constant in time. Proença et al. (2014) found that the more relevant discrepancies in estimation for the coefficients of the same type of variables and simulations presented by these authors show that the superiority in estimation of the semiparametric method was more noticeable for those sorts of coefficients. For the present study, Table 2 shows that, overall, the log distance coefficient is smaller in absolute value for the semiparametric fit than for the parametric one (almost 20 % less for IIT and VIIT and nearly 25 % less for HIIT). Therefore, one may conclude that estimates based on the parametric fit tend to exaggerate the negative impact of the distance between countries for intra industry trade. For horizontal intra-trade, HIIT, the semiparametric fit is able to detect the interesting abovementioned regional effects, contradicting the results of the parametric regression. The fact that there are no considerable differences between the estimates obtained by the parametric and the semiparametric fits for the majority of the coefficients in this application, does not necessarily make the semiparametric approach worthless, because it does provide robustness to the conclusions induced from the estimation.

5 Conclusions

This work proposes a semiparametric procedure for modelling fractional responses with panel data, based on the previous work of Lombardía and Sperlich (2012), and Proença et al. (2014). This consistently estimates the conditional expectation of the fractional variable when the individual random unobserved heterogeneity term correlates with the explanatory variables of the model in a more general and flexible context than the existing parametric procedures, such as that recently proposed by Papke and Wooldridge (2008). This approach's main idea involves filtering the unobserved random term component associated with the model's explanatory variables by the sum of unknown functions of a set of proxy variables, and also estimating the resulting model by maximum quasi-likelihood with penalised splines, in accordance with Wood (2006). The proxy variables depend on the prior knowledge that one gains from the nature of the unobserved heterogeneity term, and also the way this interrelates with the regressors. This knowledge is at best very limited, which makes finding these proxies difficult. In such cases, it is recommended to define them based on the approaches of Chamberlain (1984) and Mundlak (1978). The methodology proposed here is easy to apply in practice, through the existing procedures for the mgcv package of R.

With regards to its empirical application for the intra-industry trade indexes, the semiparametric procedure introduced in this study reports evidence of the misspecification of the assumptions applied in the parametric model, which potentially explains the differences registered in the results returned by both procedures. The parametric fit tends to significantly amplify the negative impact of distance in intra-industry trade,

[3] Intercepts of the parametric and semiparametric regressions cannot be compared.

and is unable to detect the effect of economic integration, in addition to the effect of trading with the emerging BRIC economies on intensifying the intra-industry trade of goods of similar quality. However the results regarding the other variables are similar between both approaches, and empirically confirm the hypothesis of overlapping demand, thus, the larger is the difference between trading partner per capita GDP, the lower is the level of intra-industry trade. The negative influence of trade imbalances is also subject to empirical confirmation, as intra-industry-trade trends become intensified by an increase in differences in the capital endowments of both partners. Whilst the results of both approaches do not differ greatly, the semiparametric procedure still proves helpful in ensuring robustness for the parametric results for a wider set of situations.

Future research may address the problem of applying the semiparametric procedure to define a hypothesis test which is able to detect inconsistencies in parametric estimation. Furthermore, an interesting issue would be to find alternative models that rely on less restrictive assumptions for the functional form of the conditional expectation of the fractional response than those of the semiparametric approach that is proposed in this paper, even if this would most probably be at the expense of greater complexity when applied in practice.

Acknowledgments The authors would like to acknowledge financial support from the FCT (Fundação para a Ciência e Tecnologia) through the PEst-OE/EGE/UI0491/2013 programme. They are also grateful for the comments of the participants of the conference on Applicable Semiparametrics, 2013, and of the seminars of the UrbanEcon group of ISEG. Special thanks are due for the comments of two anonymous referees and also those of the associate editor.

6 Appendix

See Table 3.

Table 3 Correlations between the explanatory variables

	MYPCK	MPOPK	MTIY	MCEEK	ABSDYPC	ABSDPOP	ABSDCEE	LDIST	BRICS	PALOPS	EU15
MYPCK	1										
MPOPK	−0.213	1									
MTIY	0.514	0.142	1								
MCEEK	**0.795**	−0.242	0.428	1							
ABSDYPC	−0.218	0.237	−0.215	−0.060	1						
ABSDPOP	−0.221	**0.999**	0.127	−0.244	0.239	1					
ABSDCEE	0.443	0.069	0.295	**0.773**	0.332	0.066	1				
LDIST	−0.526	0.520	−0.156	−0.300	0.430	0.523	0.162	1			
BRICS	−0.322	**0.751**	0.199	−0.295	0.179	**0.748**	−0.051	0.534	1		
PALOPS	−0.495	−0.132	−0.589	−0.333	0.456	−0.125	0.077	0.315	−0.138	1	
EU15	**0.790**	−0.185	0.413	0.565	−0.130	−0.195	0.341	−0.550	−0.256	−0.291	1

This table includes the empirical correlation coefficient between the explanatory variables applied in the estimations. Correlations above 0.7 are signaled in bold

References

Abd-el-Rahaman K (1991) Firms competitive and national comparative advantages as joint determinants of trade composition. Rev World Econ 127:83–97

Aquino A (1978) Intra-industry trade and inter-industry specialization as concurrent sources of international trade in manufactures. Weltwirtschaftliches Archiv 114(2):275–295

Balassa B (1966) Tariff reductions and trade in manufactures among industrial countries. Am Econ Rev 56(3):466–473

Bergstrand J (1985) The gravity equation in international trade: some microeconomic foundations and empirical evidence. Rev Econ Stat 67:474–481

Bergstrand J (1989) The generalised gravity equation, monopolistic competition, and the factor proportions theory in international trade. Rev Econ Stat 71:143–153

Bergstrand J (1990) The Heckscher-Ohlin-Samuelson model, the Linder hypothesis, and the determinants of bilateral intra-industry trade. Econ J 100:1216–1229

Brulhart M (1994) Marginal intra-industry trade: measurement and relevance for the pattern of industrial adjustment. Rev World Econ 130:600–613

Brulhart J, Elliott R, Lindley J (2006) Intra-industry trade and labour-market adjustment: a reassessment using data on individual workers. Rev World Econ 142:521–545

Chamberlain G (1982) Multivariate regression models for panel data. J Econom 18:5–46

Chamberlain G (1984) Panel data. In: Griliches Z, Intriligator MD (eds) Handbook of econometrics, vol. II. Elsevier Science Publisher, Amsterdam, pp 1247–1318. Chapter 22

Clark D (2006) Country and industry-level determinants of vertical specialisation based trade. Int Econ J 20:211–225

Davis DR (1995) Intra-industry trade: a Heckscher-Ohlin-Ricardo approach. J Int Econ 39:201–226

Deardorff A (1998) Determinants of bilateral trade: does gravity work in a neoclassical world? In: Frankel J (ed) The regionalization of the world economy. University of Chicago Press, IL, pp 7–32

Eaton J, Kierzkowski H (1984) Oligopolistic competition, product variety and international trade. In: Kierzkowski H (ed) Competition monopolistic, trade international. Oxford University Press, Oxford, pp 69–83

Ethier W (1982) National and international returns to scale in the modern theory of international trade. Am Econ Rev 72(3):389–405

Falvey RE (1981) Commercial policy and intra-industry trade. J Int Econ 11(4):495–511

Falvey R, Kierzkowski H (1987) Product quality, intra-industry trade and (im) perfect competition. In: Kierzkowski H (ed) Protection and competition in international trade. Essays in honor of W.M. Corden. Basil Blackwell, Oxford, United States, pp 143–161

Flam H, Helpman E (1987) Vertical product differentiation and North-South trade. Am Econ Rev 77(5):810–822

Fontagné L, Freudenberg M (1997) Intra-industry trade: methodological issues reconsidered. CEPII working paper

Frankel J (1997) Regional trading blocs in the world economic system. Peterson Institute for International Economics

Gould D (1994) Immigrant links to the home country: empirical implications for US bilateral trade flows. Rev Econ Stat 76:302–315

Greenaway D, Hine R, Milner C (1994) Country-specific factors and the pattern of horizontal and vertical intra-industry trade in the UK. Rev World Econ 130:77–100

Greenaway D, Hine R, Milner C (1995) Vertical and horizontal intra-industry trade: a cross industry analysis for the United Kingdom. Econ J 105:1505–1518

Grubel H, Lloyd P (1975) Intra-industry trade. The theory and measurement of international trade in differentiation products. The Macmillan Press, London

Helpman E, Krugman P (1985) Market structure and foreign trade. Harvester Wheatsheaf, Brighton

Hummels D, Levinshon J (1995) Monopolistic competition and international trade: reconsidering the evidence. Q J Econ 110:799–836

Härdle W, Müller M, Sperlich S, Werwatz A (2004) Nonparametric and semiparametric Models. Springer, Berlin

Krugman P (1991) Geography and trade. The MIT Press, Cambridge

Krugman P (1991) Increasing returns and economic geography. J Polit Econ 93:483–499

Krugman P (1980) Scale economies, product differentiation, and the pattern of trade. Am Econ Rev 70:950–959

Krugman P (1979) Increasing returns, monopolistic competition, and international trade. J Int Econ 9:469–480

Jones RW, Kierzkowski H (2001) A framework for fragmentation. In: Arndt SW, Kierzkowski H (eds) Fragmentation: new production patterns in the world economy. Oxford University Press, USA, pp 17–34

Jones RW, Kierzkowski H (2004) International fragmentation and the new economic geography. HEI working paper no 11/2004

Lancaster K (1980) Intra-industry trade under perfect monopolistic competition. J Int Econ 10:151–170

Linder S (1961) An essay on trade and transformation. Wiley, NY

Lloyd P, Grubel H (2003) Intra-industry trade. Edward Elgar Publishing, UK

Lloyd P, Lee H-H (2002) Frontiers of research in intra-industry trade. Palgrave, NY

Lombardía MJ, Sperlich S (2008) A new class of semi-mixed models and its application in small area estimation. J R Stat Soc B 70:913–930

Lombardía MJ, Sperlich S (2012) Semiparametric inference in generalized mixed effects models. Comput Stat Data Anal 56:2903–2917

Matthews K (1998) Intra-industry trade: an Australian panel study. J Econ Stud 25(2):84–97

Mundlak Y (1978) On the pooling of time series and cross section data. Econometrica 46:69–85

Murat M, Pistoresi B (2009) Migrant networks: empirical implications for the Italian bilateral trade. Int Econ J 23(3):371–390

OECD (2002) Economic outlook, analysis and forecasts. No. 71, issue 1, Paris. doi:10.1787/eco_outlook-v2002-1-en

Nott D (2006) Semiparametric estimation of mean and variance functions for non-Gaussian data. Comput Stat 21:603–620

Papke LE, Wooldridge JM (1996) Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. J Appl Econom 11:619–632

Papke LE, Wooldridge JM (2008) Panel data methods for fractional response variables to test pass rates. J Econom 145:121–133

Proença I, Sperlich S, Duygu S (2014) Semi-mixed effects gravity models for bilateral trade. Empir Econ. doi:10.1007/s00181-014-0891-x

Shen P-S (2011) Semiparametric analysis of transformation models with left-truncated and right-censored data. Comput Stat 26:521–537

Slama R, Werwatz A, Boutou O, Ducot B, Spira A, Härdle W (2003) Does male age affect the risk of spontaneous abortion? an approach using semiparametric regression. Am J Epidemiol 9:815–824

Verdoorn P (1960) The intra-bloc trade of Benelux. In: Robinson E (ed) Economic consequences of the size of nations. MacMillan, London, pp 327–368

Vernon R (1966) International investment and international trade in product cycle. Q J Econ 80:190–207

Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall/CRC Press, USA

Xiaa Y, Härdle W (2006) Semi-parametric estimation of partially linear single-index models. J Multivar Anal 97:1162–1184

Yoshida T, Kanba M, Naito K (2010) A computationally efficient model selection in the generalized linear mixed model. Comput Stat 25:463–484

Zhang J, Witteloostuijn A, Zhou C (2005) Chinese bilateral intra-industry trade: a panel data study for 50 countries in the 1992–2001 period. Weltwirtschaftliches Archiv 141:510–540