A Two-Stage Pseudo-Maximum Likelihood Method for Poisson Models with Dual Binary Endogenous Explanatory Variables: An Application to Trade and Investment Agreements

Ching-mu Chen[*], Shih-Yang Lin[†], Shin-Kun Peng[‡], Wen-Jen Tsay[‡]

December 30, 2024

Abstract

This paper introduces a novel estimator for the Poisson model with dual binary endogenous explanatory variables, addressing the need to quantify the effects of preferential economic integration agreements. We develop a two-stage pseudo-maximum likelihood estimator and derive exact analytical gradient matrices as well as Hessian matrices to improve computational speed and eliminate approximation errors substantially. Applying our approach to trade data, we find that preferential trade agreements positively impact bilateral trade flows, while bilateral investment treaties have a negative effect. Additionally, the positive interaction term between these two treaties suggests that preferential trade agreements promote horizontal foreign direct investment.

[*]Department of Geography, National Changhua University of Education, Taiwan.

[†]Department of Economics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan. E-mail: d07323003@ntu.edu.tw

[‡]Institute of Economics, Academia Sinica, Taiwan.

JEL Classification Numbers: C21, F13, F14, F15

Keywords: Binary endogenous explanatory variables; Pseudo-maximum likelihood; Poisson regression; Preferential economic integration agreement; Gravity model of international trade.

I. Introduction

Preferential economic integration agreements (PEIAs), such as preferential trade agreements (PTAs), bilateral investment treaties (BITs), or double-taxation treaties (DTTs), serve as vital policy tools for countries to enhance trade and welfare. Consequently, assessing the effects of PEIAs on bilateral trade values has continued to be a central topic in international trade literature for over six decades, dating back to the work of Tinbergen (1962). Presently, Tinbergen's gravity equation, combined with the Poisson pseudo maximum likelihood (PPML) estimator and incorporating both exporter and importer fixed effects (two-way fixed effects), serves as the prevailing method for this analysis.[1]

The econometric challenge in this task stems from the fact that PEIAs are not exogenous variables (Baier and Bergstrand, 2007) and solutions for addressing endogeneity in PEIAs are still under development.[2] Addressing endogenous variables in non-linear models such as Poisson models poses a well-known challenge, especially when dealing with binary endogenous explanatory variables (Wooldridge, 2014). Notably, the pioneering instrumental variable technique employed by Egger et al. (2011) can only tackle the endogeneity arising from self-selection into one specific PEIA, leaving the endogeneity in other PEIAs unresolved.[3]

This study thus develops a Poisson model that incorporates two binary endogenous explanatory variables (BEEVs) when two-way fixed effects are included. Specifically, we adopt the approach of Terza (1998) to develop a two-stage PPML estimator, hereafter referred to as 2SPPML.[4] Specifically, our econometric model includes an outcome equation with a structural-causal interpretation (Eq. (2)), along with additional equations that capture the generating processes of the endogenous binary variables also termed treatment variables (Eqs. (3)–(4)).

We choose this method due to its efficiency when the econometric model is correctly specified and its ability to examine both the determinants and effects of PEIAs. Consequently, our econometric model unifies two strands of literature within a single framework: (1) the economic determinants of the formation of PEIAs, as explored by Baier and Bergstrand (2004) and Bergstrand and Egger (2013), and (2) the effects of PEIAs on trade values, as studied by Heid and Vozzo (2020) and Egger and Tarlea (2021). In contrast, if the focus is solely on the effects of PEIAs, then the instrumental-variable (IV) approach can be an alternative since it requires few assumptions. In the literature, Mullahy (1997), Windmeijer and Silva (1997), and Jochmans and Verardi (2022) propose different IV estimators based on distinct orthogonality conditions.[5]

Moreover, our method is generally immune to the incidental parameter problem when incorporating two-way fixed effects.[6] In contrast, Jochmans (2022) finds the instrumental-variable estimators are in general inconsistent and necessitate bias correction. This property is valuable in international economics applications for two primary reasons: First, Anderson and van Wincoop (2003) show that omitting multilateral resistance (MR) terms—which capture how a region's average trade barriers with all partners affect bilateral trade flows—leads to biased estimates. Second, these MR terms are crucial for subsequent economic analyses, including welfare calculations and trade policies evaluations in general equilibrium models. Following Feenstra (2004), Redding and Venables (2004), and Fally (2015), researchers are able to recover the MR terms of the structural gravity model by estimating the gravity equation using a PPML estimator with two-way fixed effects. While the IV estimator proposed by Jochmans and Verardi (2022) addresses incidental parameter problems by differencing out the two-way fixed effects, requiring other methods to recover them for further analysis. Additionally, their method's inability to produce predicted values for trade flows limits its applicability in empirical work.

The empirical challenge of conducting the proposed 2SPPML method is its computational load when the model consists of hundreds of parameters, even though we successfully derive the corresponding pseudo-log-likelihood function of the estimator.[7] The computational burdens lie in the demand for approximating the gradients and Hessians of the pseudo-log-likelihood function using numerical methods. In international trade applications, the number of observations grows exponentially, and the number of parameters increases linearly with the number of countries, creating substantial computational challenges. For instance, when using data from 100 countries in a Poisson model, there are 9,900 ($= 100 \times 99$) observations (i.e., bilateral country pairs) and more than 200 ($= 2 \times 100$) parameters.

To resolve the daunting tasks of estimating the proposed model with huge observations and a massive number of parameters, we work out the exact analytical gradients and Hessian matrices of the pseudo-log-likelihood function as outlined in the mathematical appendices of this paper. Accordingly, the computational performance tremendously improves by a thousand-fold of acceleration, as shown in the Monte Carlo simulations of this paper. Furthermore, the use of exact analytical formulas ensures that the calculation is free from approximation errors. In fact, our method is not only fast and easily implemented, but it also exhibits promising finite sample performance, as displayed in the simulation results under a Poisson model with a two-way fixed effect specification consisting of several hundreds of parameters.

We then employ the newly developed Poisson model on the dataset of Egger et al. (2011). This dataset provides a cross-section of trade flows and PTA relationships among 126 countries for the year 2005. To augment the data, we have included BIT relationships. The rationale behind this extension is that, apart from PTAs, BITs can influence the volume of foreign direct investment (FDI), which subsequently affects bilateral trade values via numerous channels such as vertical integration and horizontal expansion (Berger et al., 2013).

Vertical integration allows for the specialization of production stages across different countries, capitalizing on their comparative advantages in factors of production and boosting trade in intermediate goods. Conversely, horizontal expansion involves the replication of production facilities in various countries, improving market access, but potentially reducing bilateral trade values. The inclusion of such significant PEIA allows for the discovery of more intricate relationships and possible interactions among these factors, thereby deepening our comprehension of the underlying mechanisms.

Our results show that the decisions of signing PTAs and BITs are positively correlated, echoing the findings of Bergstrand and Egger (2013). While signing a PTA has a strong positive impact on bilateral trade flows (i.e., 172.92%), a BIT has a negative effect (i.e., −46.79%). The partial effects of PTAs and BITs become much smaller (74.16% and −25.46%, respectively) if endogeneity is ignored.

Given that, BIT lowers the cross-border investment costs and encourages FDI, and our "negative BIT effect" finding highlights the so-called proximity-concentration trade-off (Brainard, 1997; Neary, 2009), which emphasizes the strong role of horizontal FDI. Since horizontal FDI aims at replicating production facilities abroad to improve access to foreign markets, a BIT will cause a negative effect on trade if horizontal FDI is encouraged by the BIT. Brainard (1997) finds that FDI is high in industry-country pairs with high transport costs and low plant scale economies, while international differences in relative factor abundance have little effect on FDI. In the same vein, Markusen (2002) finds evidence that bilateral flows of FDI at the industry level are encouraged by similarities in market size and in relative endowments of skilled and unskilled labor between countries. Such empirical evidence is consistent with the view that FDI is primarily horizontal rather than vertical.[8]

We also investigate the interaction effect of PTAs and BITs in another model specification that adds the interaction term, $PTA \times BIT$, into the econometric model. The coefficients of PTA, BIT, and $PTA \times BIT$ are 0.8342, -0.6893, and 0.2762, respectively. Relative to the case where there is neither PTA nor BIT between a specific country pair, negotiating a PTA increases the expectation of the bilateral trade flow by 130.30%, while signing a BIT decreases the expectation of the bilateral trade flow by 49.81%. Having both PTAs and BITs will raise the expectation of the bilateral trade flow by 52.36% relative to the case without any PTA and BIT. Through the positive $PTA \times BIT$ term, the negative effect of BITs on trade is mitigated. This implies that the trade effect of a BIT depends on whether this country pair possesses a PTA and belongs to the same trade bloc. It reminds us that multinational corporations pursue what Yeaple (2003), following UNCTAD (1998), calls "complex integration strategies". Adopting a strategy of horizontal FDI, vertical FDI, or both, definitely depends on the trade barriers between the home country and the host country.[9]

Our work empirically highlights the distinct motives of FDI between cases with or without PTAs. Horizontal FDI dominates when the home country and the host country are not in the same trade bloc. Thus, a BIT will encourage this FDI strategy and in turn, decrease the trade flow from the home to the host. In contrast, if the two countries already have a PTA and participate in the same trade bloc, then a BIT will benefit from the motive of vertical FDI, which generates another channel to increase the bilateral trade flow by rising trade in intra-firm intermediate goods.

The rest of this paper runs as follows. Section 2 presents our econometric model, the two-stage Poisson pseudo-maximum likelihood (2SPPML) estimation procedure, and two sets of Monte Carlo simulations. Section 3 describes the data and the estimation results of the trade effects of endogenous PTAs and BITs. Section 4 concludes.

II. Econometric model and estimation method

This section presents the econometric model and estimation procedures for exploring the effects of dual BEEVs on count data outcomes. As shown in the introduction section, we follow the suggestions of Gourieroux, Monfort and Trognon (1984) and Santos Silva and Tenreyro (2006) to employ a pseudo-maximum likelihood (PML) estimator to our Poisson model, mainly because the PPML estimator delivers consistent parameter estimates provided that the conditional expectation of the response variable is correctly specified. Moreover, our 2SPPML estimator is consistent even when the dependent variable is not count data (i.e., non-negative integer). It can be any non-negative real number (Gourieroux, Monfort and Trognon, 1984). Finally, our 2SPPML estimator is immune to the incidental parameter problem when incorporating two-way fixed effects (Fernández-Val and Weidner, 2016; Weidner and Zylkin, 2021).

Econometric model

When the response variable, y, is count data, it is popular to model y with the exponential function:

$$\mathbf{E}[y|\boldsymbol{x}] = \exp(\boldsymbol{x}\boldsymbol{\beta}), \tag{1}$$

where $\boldsymbol{x}$ is a vector of explanatory variables, and $\boldsymbol{\beta}$ is the vector of parameters to be estimated. For example, in the gravity model, y is the nominal exports of a country pair, and $\boldsymbol{x}$ is a vector containing trade cost and trade facilitating variables.

To capture potential unobserved idiosyncratic characteristics of the country pair, we add an error term r_1 and assume that the conditional expectation of the response variable has the following form:

$$\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3, r_1] = \exp(\boldsymbol{x}_1\boldsymbol{\beta}_1 + r_1), \tag{2}$$

where $\boldsymbol{x}_1$ contains a constant term, a vector of exogenous variables $\boldsymbol{x}$, and the two binary variables, y_2 and y_3 (like the statuses of PTA and BIT), respectively. Note that $\boldsymbol{z}$ denotes a vector of exogenous variables, including a constant term, $\boldsymbol{x}$, and instrumental variables that are excluded from $\boldsymbol{x}$. The binary variables, y_2 and y_3, are endogenous. For example, the decisions of signing PTAs and BITs might relate to expected bilateral trade values. In fact, Bergstrand and Egger (2013) model the PTA and BIT statuses of a country pair with the following bivariate probit models:

$$y_2 = 1[\boldsymbol{z\delta}_2 + v_2 \geq 0], \quad v_2|\boldsymbol{z} \sim \text{Normal}(0, 1), \tag{3}$$

$$y_3 = 1[\boldsymbol{z\delta}_3 + v_3 \geq 0], \quad v_3|\boldsymbol{z} \sim \text{Normal}(0, 1), \tag{4}$$

where $(\boldsymbol{\delta}_2, \boldsymbol{\delta}_3)$ and (v_2, v_3) are the vectors of true parameters and error terms corresponding to Eq. (3) and Eq. (4), respectively. One can interpret these two equations as decisions based on the following differences in utility levels from having a PTA or a BIT, denoted respectively by:

$$y_2^* = \boldsymbol{z\delta}_2 + v_2, \text{ and } y_3^* = \boldsymbol{z\delta}_3 + v_3.$$

Thus, a pair of countries sign a PTA or a BIT if they have a positive difference in these utility levels, y_2^* or y_3^*. However, Bergstrand and Egger (2013) do not consider the impacts of PTA and BIT on trade flows.

We characterize the endogeneity of y_2 and y_3 by assuming the error terms (r_1, v_2, v_3) as multivariate normal variables (conditional on the exogenous variables) with mean vector zero and variance-covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \alpha^2 & \mu_2 & \mu_3 \\ \mu_2 & 1 & \gamma \\ \mu_3 & \gamma & 1 \end{bmatrix}. \tag{5}$$

Resorting to Theorem 3.10 and Corollary 5.5 of White (1994), the parameters of interest, $\boldsymbol{\beta}_1$, can be consistently estimated by the 2SPPML method under suitable regularity conditions with the correct functional form of $\mathbf{E}(y_1|\boldsymbol{z}, y_2, y_3)$

and the associated log-likelihood function displayed in the following Section 2.2. Accordingly, the first stage for estimating $\boldsymbol{\beta}_1$ is to derive this correct functional form of $\mathbf{E}(y_1|\boldsymbol{z}, y_2, y_3)$.

We thus begin by applying the law of iterative expectation to Eq. (2):

$$\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3] = \mathbf{E}\big[\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3, r_1]\big|\boldsymbol{z}, y_2, y_3\big].$$

Via Eq. (2) again, we replace $\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3, r_1]$ on the right-hand side of the above equation and arrive:

$$\mathbf{E}\big[\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3, r_1]\big|\boldsymbol{z}, y_2, y_3\big] = \exp(\boldsymbol{x}_1\boldsymbol{\beta}_1)\mathbf{E}[\exp(r_1)|\boldsymbol{z}, y_2, y_3].$$

Given the normal distribution assumption of the error terms, (r_1, v_2, v_3), we derive an analytic form of the above conditional expectation. We leave the details to the mathematical appendices of the paper and present the formula in the following:

$$\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3] = \exp(\boldsymbol{x}_1\boldsymbol{\beta}_1^*)\Psi(\mu_2, \mu_3, \gamma, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3), \tag{6}$$

where

$$\Psi(\mu_2, \mu_3, \gamma, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3) = \frac{\Phi_2(w_2 + q_2\mu_2, w_3 + q_3\mu_3, \gamma_\star)}{\Phi_2(w_2, w_3, \gamma_\star)}, \tag{7}$$

and $\Phi_2(a, b, \rho)$ denotes the bivariate standard normal cumulative distribution function (CDF) such that $a, b \in \mathbb{R}$ and $\rho \in [-1, 1]$. Moreover, $q_2 = 2y_2 - 1$, $q_3 = 2y_3 - 1$, $w_2 = q_2\boldsymbol{z}\boldsymbol{\delta}_2$, $w_3 = q_3\boldsymbol{z}\boldsymbol{\delta}_3$, and $\gamma_\star = q_2q_3\gamma$. In Eq. (6), since we cannot identify α, the term $\alpha^2/2$ is "merged" with the first element of $\boldsymbol{\beta}_1$, and we denote $\boldsymbol{\beta}_1^*$ as $\boldsymbol{\beta}_1$ with its first element plus an extra $\alpha^2/2$.

It is clear now that one might face omitted variable bias if the endogenous binary variables, y_2 and y_3, are treated as exogenous variables by using the following conditional expectation from conventional PPML approach to estimate the parameters of interest:

$$\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3] = \exp(\boldsymbol{x}_1\boldsymbol{\beta}_1) \equiv \lambda(\boldsymbol{\beta}_1). \tag{8}$$

The term Ψ in Eq. (6) can be interpreted as the nonzero conditional error expectation in Eq. (7) of Heckman (1978) (the inverse Mill's ratio correction term). Analogous to Heckman's more conventional endogenous dummy variable model, neglecting the "correction term" in the conditional mean function results in a form of omitted variable bias.

To distinguish from the notation used for the exogenous y_2 and y_3 scenario, we define the correctly specified conditional expectation, Eq. (6), as $\lambda^*(\boldsymbol{\beta}_1^*, \mu_2, \mu_3, \gamma, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3)$:

$$\mathbf{E}[y_1|\boldsymbol{z}, y_2, y_3] = \exp(\boldsymbol{x}_1\boldsymbol{\beta}_1^*)\Psi(\mu_2, \mu_3, \gamma, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3) \equiv \lambda^*(\boldsymbol{\beta}_1^*, \mu_2, \mu_3, \gamma, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3). \tag{9}$$

Pseudo-maximum likelihood estimation

We estimate the parameters of interest, $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}_1^{*\top}, \mu_2, \mu_3)^\top$ and $\boldsymbol{\xi} \equiv (\gamma, \boldsymbol{\delta}_2^\top, \boldsymbol{\delta}_3^\top)^\top$, based on Eq. (6) with a 2SPPML method. Different from the maximum likelihood (ML) method that specifies the conditional distribution of y_1, the PPML method **approximates** it by a Poisson density function (White, 1994; Wooldridge, 2010):

$$f(y_1|y_2, y_3, \boldsymbol{x}_1, \boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{\xi}) = \frac{\exp(-\lambda^*) \cdot (\lambda^*)^{y_1}}{y_1!},$$

where λ^* is a function of $(\boldsymbol{\theta}, \boldsymbol{\xi})$ defined in Eq. (9).

We now describe the estimation procedures. The first stage is to use a bivariate probit model to estimate Eq. (3) and Eq. (4). The estimates thus obtained are denoted as $\widehat{\boldsymbol{\xi}} \equiv (\widehat{\boldsymbol{\delta}}_2, \widehat{\boldsymbol{\delta}}_3, \widehat{\gamma})$.[10] We then plug these estimates in the second stage of the Poisson pseudo-maximum log-likelihood function for observation i:

$$\begin{aligned} l_i(\boldsymbol{\theta}, \widehat{\boldsymbol{\xi}}) &\equiv l(y_{1i}, y_{2i}, y_{3i}, \boldsymbol{x}_{1i}, \boldsymbol{z}_i; \boldsymbol{\theta}, \widehat{\boldsymbol{\xi}}) = \ln f(y_{1i}|y_{2i}, y_{3i}, \boldsymbol{x}_{1i}, \boldsymbol{z}_i; \boldsymbol{\theta}, \widehat{\boldsymbol{\xi}}) \\ &= -\widehat{\lambda}_i^* + y_{1i}\ln(\widehat{\lambda}_i^*) - \ln(y_{1i}!), \end{aligned} \tag{10}$$

where

$$\widehat{\lambda}_i^* \equiv \lambda_i(\boldsymbol{\beta}^*)\Psi_i(\mu_2, \mu_3, \widehat{\gamma}, \widehat{\boldsymbol{\delta}}_2, \widehat{\boldsymbol{\delta}}_3). \tag{11}$$

The PML estimator $\widehat{\boldsymbol{\theta}} \equiv (\widehat{\boldsymbol{\beta}}_1^*, \widehat{\mu}_2, \widehat{\mu}_3)$ is the solution to:

$$\max_{\boldsymbol{\vartheta}} \sum_{i=1}^{N} l_i(\boldsymbol{\vartheta}, \widehat{\boldsymbol{\xi}}).$$

Non-linear solvers in MATLAB's *fmincon* function are used to locate $\widehat{\boldsymbol{\theta}}$. Nevertheless, the task is time-consuming if we rely on the solvers to calculate the score and Hessian of the pseudo-log-likelihood function in Eq. (10) via numerical approaches. Another contribution of this paper is to resolve this numerical issue by providing the analytical formulae of the score and Hessian functions of the pseudo-log-likelihood function so as to speed up the estimation procedure by more than one thousand times. The details of these formulae are also presented in the mathematical appendices of the paper.

Numerical studies

This section reports the simulation results for two purposes:

(1) showing that the estimates from conventional PPML can be biased if endogeneity exists; and

(2) assessing the performance of our proposed model.

We first demonstrate that our algorithm can correctly deal with the Poisson model with dual BEEVs when the numbers of observations are relatively small. The algorithm also estimates the gravity models with tens of thousands of observations and hundreds of parameters and completes the computation tasks with 1,000 replications quickly.

We design the first Monte Carlo simulation for the scenario where the number of parameters is moderate. We choose different sample sizes, $N \in \{200, 800, 3200\}$, under the following data generating process (DGP):

$$y_{1,ij} = \exp(\beta_{11} + \beta_{12}x_{12,ij} + \beta_{13}y_{2,ij} + \beta_{14}y_{3,ij} + r_{1,ij}), \tag{12}$$

$$y_{2,ij} = 1[\delta_{21} + \delta_{22}z_{1,ij} + \delta_{23}z_{2,ij} + \delta_{24}z_{3,ij} + v_{2,ij} \geq 0], \tag{13}$$

$$y_{3,ij} = 1[\delta_{31} + \delta_{32}z_{1,ij} + \delta_{33}z_{2,ij} + \delta_{34}z_{3,ij} + v_{3,ij} \geq 0], \tag{14}$$

where $i, j = 1, 2, \ldots, N/2$, and $(r_{1,ij}, v_{2,ij}, v_{3,ij})$ are drawn from a multivariate normal distribution with mean zero and covariance, $\boldsymbol{\Sigma}$, as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \alpha^2 & \mu_2 & \mu_3 \\ \mu_2 & 1 & \gamma \\ \mu_3 & \gamma & 1 \end{bmatrix}. \tag{15}$$

We note that, in the spirit of "pseudo"-maximum likelihood, the dependent variable $y_{1,ij}$ is not count data (i.e., non-negative integer) nor follows a Poisson process.

The covariates $x_{12,ij}, z_{1,ij}, z_{2,ij}$, and $z_{3,ij}$ are all drawn independently from uniform distribution $U(-0.5, 0.5)$, and the parameters of interest are set by $(\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}) = (-1, 0, 1, 1)$, $(\delta_{21}, \delta_{22}, \delta_{23}, \delta_{24}) = (0, 0.5, -1, 1)$ as well as $(\delta_{31}, \delta_{32}, \delta_{33}, \delta_{34}) = (0, 0.5, 1, -1)$. Moreover, we use the estimates from the conventional PPML estimator of β_{12}, β_{13}, and β_{14} and add them with the extra random numbers generated from Uniform$(-0.5, 0.5)$ to serve as their initial values for implementing the proposed method and to create a realistic simulation scheme. As for the initial values for estimating μ_2 and μ_3, we simply use a random vector generated from Uniform$(-0.5, 0.5)$ as their starting values.

We evaluate the performance of the 2SPPML estimator under three different endogeneity levels: $(\mu_2, \mu_3, \gamma) = (0.3, 0.3, 0.3)$, $(0.6, 0.6, 0.6)$, and $(0.9, 0.9, 0.9)$. We compare the average biases, and root-mean-squared errors (RMSE) of the 2SPPML estimator with those of the conventional PPML method, which treats all covariates as exogenous variables. For ease of exposition, we refer to the conventional PPML as PPML for the rest of the paper. The simulations are all based on 1,000 replications, and the results are presented in Table 1.

Table 1 reports the finite sample performance of estimating β_{12}, β_{13}, and β_{14}. First, the average biases of the 2SPPML estimator are lower than those of the PPML approach across all configurations considered in the table. For example, the average biases of β_{12} for the 2SPPML estimator and the PPML estimator are

−0.0230 and −0.1565, respectively, when $N = 200$ and the level of endogeneity is low. Second, the average biases of the 2SPPML estimator decrease about half in magnitude when the sample size doubles, whereas those of the PPML estimator remain roughly intact. Accordingly, the PPML estimator cannot be used even if we have a large sample size. Finally, the RMSE of the 2SPPML estimator decreases with the sample size, indicating it has a well-behaved asymptotic property.

The second Monte Carlo simulation is designed to demonstrate the performance of the 2SPPML estimator for the widely known gravity model in international economics. We use N = 2,500, 10,000, and 22,500 to align with the regular sample size observed in international economics. For example, Egger et al. (2011) include 126 × 125 = 15,750 country-pairs in their dataset.

We also include two sets of dummies to imitate the scenario of a gravity model with two-way fixed effects along with the proceeding endogeneity setting in the error terms, $(r_{1,ij}, v_{2,ij}, v_{3,ij})$. The new DGP thus becomes:

$$y_{1,ij} = \exp(\beta_{11} + \beta_{12}x_{12,ij} + \beta_{13}y_{2,ij} + \beta_{14}y_{3,ij} + e_{1,i} + m_{1,j} + r_{1,ij}), \quad (16)$$

$$y_{2,ij} = 1[\delta_{21} + \delta_{22}z_{1,ij} + \delta_{23}z_{2,ij} + \delta_{24}z_{3,ij} + e_{2,i} + m_{2,j} + v_{2,ij} \geq 0], \quad (17)$$

$$y_{3,ij} = 1[\delta_{31} + \delta_{32}z_{1,ij} + \delta_{33}z_{2,ij} + \delta_{34}z_{3,ij} + e_{3,i} + m_{3,j} + v_{3,ij} \geq 0], \quad (18)$$

where the six terms, $e_{1,i}$, $e_{2,i}$, $e_{3,i}$, $m_{1,j}$, $m_{2,j}$, and $m_{3,j}$, represent the two-way fixed effects included in the above three equations. The letter 'e' denotes the meaning of exporter, while the letter 'm' conveys the idea of an importer. Moreover, these six terms are independently generated from the product of two random variables $O_1 \times O_2$, where $O_1 \sim \text{Uniform}(-0.5, 0.5)$, and $O_2 \sim \text{Normal}(0, 0.5^2)$, and Q_1 and Q_2 are independent of each other across all i and j considered in these six items.

Since the simulation design includes intra-national trade country pairs, the number of included importer and exporter dummies in Eqs. (16)–(18) is $2 \times \sqrt{N}$. For example, when $N = 2,500$, we have $\sqrt{N} = \sqrt{2,500} = 50$ countries. And the number of corresponding two-way fixed effects is thus $2 \times 50 = 100$. We summarize the results in Table 2 and Table 3 based on 1,000 replications again.

Table 2 reports the finite sample performance of estimating β_{12}, β_{13}, and β_{14}. We find similar patterns as those found in the first simulation experiment. The average biases of the 2SPPML estimator are always lower than those of the PPML method across all endogeneity levels and all sample sizes. The estimates of PPML are severely biased, as expected. When the average biases and RMSE of the 2SPPML estimator decrease with sample sizes, the corresponding values of the PPML estimator roughly remain intact even when facing the same larger sizes.

Table 3 presents the results from estimating μ_2, μ_3, and γ, which are the indicators of the endogeneity level of the binary regressors. Interestingly, we find similar patterns as we observe in Table 2; i.e., the average biases and RMSE of these estimators decrease with the sample size. Combining the results in Table 2 and Table 3, we find that the 2SPPML has promising performance, even though large numbers of two-way fixed effects are included in the model. This signals the potential of using our proposed method in many empirical studies.

III. Empirical results

In this section, we apply our proposed method to the dataset used in Egger et al. (2011) where they focus on the case with a single BEEV. In particular, they consider the effects of signing PTAs on the magnitude of bilateral trade flows. Building on their framework, we add an additional BEEV, BIT_{ij}, in the model, because global fragmentation motivates firms to influence trade policies, including BITs (Blanchard, 2010; Blanchard and Matschke, 2015; Blanchard, Bown and Johnson, 2016). This not only extends the coverage of the literature, but also helps us check the robustness of the findings in Egger et al. (2011).

Table 1: Monte Carlo Results

	β_{12}		β_{13}		β_{14}	
Estimator:	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\mu_2 = 0.3$, $\mu_3 = 0.3$, $\gamma = 0.3$						
$N = 200$						
2SPPML	-0.0230	0.3743	-0.1322	0.3347	0.2005	0.4015
PPML	-0.1565	0.3958	0.4155	0.4532	0.4154	0.4607
$N = 800$						
2SPPML	-0.0082	0.1867	-0.0637	0.1655	0.0888	0.1870
PPML	-0.1518	0.2353	0.4154	0.4254	0.4171	0.4293
$N = 3{,}200$						
2SPPML	-0.0044	0.0939	-0.0309	0.0807	0.0441	0.0959
PPML	-0.1523	0.1773	0.4158	0.4182	0.4171	0.4203
$\mu_2 = 0.6$, $\mu_3 = 0.6$, $\gamma = 0.6$						
$N = 200$						
2SPPML	-0.0297	0.3401	-0.0922	0.2609	0.2204	0.4105
PPML	-0.2557	0.3981	0.7151	0.7301	0.7475	0.7680
$N = 800$						
2SPPML	-0.0064	0.1684	-0.0549	0.1404	0.0914	0.1800
PPML	-0.2531	0.2959	0.7157	0.7195	0.7466	0.7516
$N = 3{,}200$						
2SPPML	-0.0053	0.0840	-0.0301	0.0748	0.0435	0.0895
PPML	-0.2572	0.2680	0.7172	0.7182	0.7448	0.7461
$\mu_2 = 0.9$, $\mu_3 = 0.9$, $\gamma = 0.9$						
$N = 200$						
2SPPML	-0.0236	0.3057	-0.0602	0.1679	0.1977	0.3440
PPML	-0.3155	0.4014	0.9239	0.9290	1.0252	1.0332
$N = 800$						
2SPPML	-0.0103	0.1546	-0.0406	0.0987	0.0846	0.1572
PPML	-0.3196	0.3433	0.9241	0.9254	1.0248	1.0268
$N = 3{,}200$						
2SPPML	-0.0030	0.0771	-0.0242	0.0546	0.0352	0.0707
PPML	-0.3181	0.3243	0.9260	0.9263	1.0249	1.0254

Notes: This table reports the simulation results of Eq. (12) based on 1,000 replications. β_{12} is the coefficient of an exogenous regressor, whereas β_{13} and β_{14} are the coefficients of BEEVs.

Table 2: Monte Carlo Results of the Gravity Model

	β_{12}		β_{13}		β_{14}	
Estimator:	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\mu_2 = 0.3$, $\mu_3 = 0.3$, $\gamma = 0.3$						
$N = 2{,}500$						
2SPPML	-0.0974	0.2537	0.2777	0.6733	0.2830	0.6701
PPML	-0.1452	0.1787	0.4182	0.4219	0.4200	0.4242
$N = 10{,}000$						
2SPPML	-0.0481	0.1440	0.1248	0.3738	0.1259	0.3679
PPML	-0.1520	0.1611	0.4167	0.4176	0.4205	0.4217
$N = 22{,}500$						
2SPPML	-0.0315	0.0939	0.0848	0.2513	0.0821	0.2441
PPML	-0.1505	0.1548	0.4164	0.4168	0.4202	0.4207
$\mu_2 = 0.6$, $\mu_3 = 0.6$, $\gamma = 0.6$						
$N = 2{,}500$						
2SPPML	-0.0977	0.1851	0.2787	0.4763	0.2815	0.4776
PPML	-0.2532	0.2683	0.7188	0.7202	0.7502	0.7520
$N = 10{,}000$						
2SPPML	-0.0451	0.0888	0.1427	0.2368	0.1350	0.2311
PPML	-0.2527	0.2563	0.7203	0.7206	0.7469	0.7474
$N = 22{,}500$						
2SPPML	-0.0279	0.0569	0.0850	0.1510	0.0828	0.1482
PPML	-0.2550	0.2567	0.7218	0.7220	0.7488	0.7490
$\mu_2 = 0.9$, $\mu_3 = 0.9$, $\gamma = 0.9$						
$N = 2{,}500$						
2SPPML	-0.0756	0.1300	0.2418	0.3073	0.2720	0.3401
PPML	-0.3170	0.3261	0.9308	0.9314	1.0351	1.0358
$N = 10{,}000$						
2SPPML	-0.0366	0.0622	0.1220	0.1522	0.1343	0.1673
PPML	-0.3171	0.3192	0.9330	0.9331	1.0300	1.0302
$N = 22{,}500$						
2SPPML	-0.0228	0.0419	0.0775	0.1002	0.0862	0.1097
PPML	-0.3175	0.3185	0.9353	0.9353	1.0272	1.0273

Notes: This table reports the simulation results of Eq. (16) based on 1,000 replications. β_{12} is the coefficient of an exogenous regressor, whereas β_{13} and β_{14} are the coefficients of BEEVs.

Table 3: Monte Carlo Results of the Gravity Model

	μ_2		μ_3		γ	
Num. Obs.	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$\mu_2 = 0.3$, $\mu_3 = 0.3$, $\gamma = 0.3$			
2,500	-0.1985	0.4707	-0.2055	0.4711	0.0094	0.0348
10,000	-0.0910	0.2653	-0.0915	0.2605	0.0031	0.0166
22,500	-0.0619	0.1788	-0.0597	0.1745	0.0032	0.0115
			$\mu_2 = 0.6$, $\mu_3 = 0.6$, $\gamma = 0.6$			
2,500	-0.2339	0.3872	-0.2291	0.3782	0.0177	0.0319
10,000	-0.1180	0.1913	-0.1110	0.1831	0.0081	0.0153
22,500	-0.0699	0.1218	-0.0663	0.1157	0.0054	0.0100
			$\mu_2 = 0.9$, $\mu_3 = 0.9$, $\gamma = 0.9$			
2,500	-0.2326	0.2875	-0.2281	0.2814	0.0250	0.0287
10,000	-0.1158	0.1409	-0.1122	0.1365	0.0107	0.0125
22,500	-0.0741	0.0925	-0.0719	0.0895	0.0072	0.0085

Notes: This table reports the simulation results of Eq. (15) based on 1,000 replications. The magnitudes of μ_2 and μ_3 stand for the level of endogeneity, and γ measures the correlation between the two BEEVs.

The dataset of Egger et al. (2011) includes the information concerning 126 countries for the year 2005. We thus have $126 \times 125 = 15{,}750$ bilateral trade country-pairs. We denote an indicator, BIT_{ij}, to be one, if the importer and exporter countries sign a BIT in force and zero otherwise. Table 4 reports the descriptive statistics of the data used for the estimation. To gain better computational precision, we rescale some variables to ensure their means fall between zero and one. When the data are subject to a rescaling procedure, we mark them with $*$. For example, the log distance between the importer and the exporter, $DIST_{ij}*$, is divided by 10.

We consider three different settings. The first one treats both PTA and BIT statuses as exogenous, and the conventional PPML method is used consequently. This setting is mostly employed in empirical studies and can be viewed as the benchmark of our investigation.

The second setting treats either PTA or BIT as an endogenous decision of the country pair involved. This is the model considered in Egger et al. (2011) where they employ the Poisson model with only one BEEV. For fairness of comparison, we use the programming code of Egger et al. (2011) to conduct the estimation task.

The last setting considers both PTA and BIT as endogenous variables and applies the proposed estimation procedure in Section 2. Furthermore, the importer and exporter fixed effects are both included with the other country-specific determinants such as a country's GDP, population, and capital-labor ratio in the model specifications.

In Table 5, columns (1)–(3) report the estimation results of the PPML estimates without considering the endogeneity of PTAs and BITs. The estimates found in these three columns are very similar to each other. Column (1) and column (2) report a positive $\left[\exp(0.5548) - 1\right] \times 100\% = 74.16\%$ PTA effect and a negative $\left[\exp(-0.2939) - 1\right] \times 100\% = -25.46\%$ BIT effect, respectively.

From column (3), we find the PTA effect is 69.93% and the BIT effect is -22.59% when both PTA and BIT statuses are included. The negative BIT effect is more in line with multinationals' horizontal FDI strategies, because the firms undertaking horizontal FDI can avoid trade costs by replicating production facilities abroad. Export values between the country pairs decrease consequently.

In columns (1) and (2) of Table 6, we replicate the empirical results of Egger et al. (2011) who only consider the endogeneity of PTAs. Our results are indeed identical to those of Egger et al. (2011). For the variables subject to the rescaling procedure, the magnitudes of the parameters require the same rescaling. For example, the estimated coefficient of $DIST_{ij}*$ is -10.7370. This is identical to that of Egger et al. (2011) after dividing by 10, which is the exact number we use to divide the original data of $DIST_{ij}$ in Egger et al. (2011).

We then examine μ_2 in Table 6. It is significantly negative, rejecting the null hypothesis of exogeneity of PTAs. A negative μ_2 reveals that the unobservables that we do not include in the model of signing PTA correlate negatively with the unobservables affecting bilateral trade values. In other words, the country pairs that have lower bilateral trade flows are more likely to sign PTAs after controlling the economic and political factors. This negative correlation causes a downward bias in the estimated PTA parameter when using the conventional PPML models. It is thus not surprising to find that the estimate in column (1) of Table 6 is about double the estimates of PTA_{ij} in column (1) of Table 5, revealing PTA has strong impacts (214.90%) on bilateral trade flows when we control its endogeneity.

We also apply the methodology of Egger et al. (2011) to explore the endogenous BIT effect and display the results in columns (3) and (4) of Table 6. Since μ_3 is significantly positive, we can reject the null hypothesis of exogeneity of BITs. In fact, a positive μ_3 indicates that the unobservables in the model of signing a BIT correlate positively with the unobservables affecting bilateral trade values.

It implies that the country pairs that have higher bilateral trade flows are more likely to sign BITs after controlling the economic and political factors. Such a positive correlation causes an upward bias in the BIT effects if the conventional exogenous PPML is used. As a result, the estimated BIT effect in column (2) of Table 5 is −25.46%, which is higher than −51.69% of the estimated BIT effect in column (4) of Table 6.

We finally discuss the estimation results based on our proposed 2SPPML method in Table 7. The robust standard errors used in this table are based on the approach of White (1994) and Wooldridge (2010), which considers the first-stage estimation effects on the second-stage Poisson-type regression. The computational details are also outlined in the mathematical appendices of this paper.

The first finding of Table 7 shows a positive γ, implying that the unobservable determinants of signing PTAs and BITs positively correlate, echoing the results of Bergstrand and Egger (2013). Second, the signs of μ_2 and μ_3 stay unchanged from the results in Table 6, indicating one would obtain an underestimated PTA effect and an overestimated BIT effect when the endogeneities of PTA and BIT are not taken into account at all. Moreover, when we consider the endogeneities of PTA and BIT simultaneously, their effects are $\big[\exp(1.0040) - 1\big] \times 100\% = 172.92\%$ and $\big[\exp(-0.6309) - 1\big] \times 100\% = -46.79\%$, respectively. These numbers are smaller in absolute value than those in Table 6 where the endogenous decisions of signing the PTA and BIT are not considered simultaneously.

Column (4) shows the results including an interaction term, $PTA_{ij} \times BIT_{ij}$. The coefficients of PTA_{ij}, BIT_{ij}, and $PTA_{ij} \times BIT_{ij}$ are 0.8342, −0.6893, and 0.2762, respectively. Comparing to the case where there is neither PTA nor BIT between a specific country pair, negotiating a PTA increases the expectation of the bilateral trade flow by $\big[\exp(0.8342) - 1\big] \times 100\% = 130.30\%$, while signing a BIT increases the expectation of the bilateral trade flow by

$\big[\exp(-0.6893) - 1\big] \times 100\% = -49.81\%$. Furthermore, having both PTAs and BITs raises the expectation of the bilateral trade flow by $\big[\exp(0.8342 - 0.6893 + 0.2762) - 1\big] \times 100\% = 52.36\%$ relative to the case without any PTA and BIT. The positive $PTA \times BIT$ term suggests that the partial effect of BITs on trade depends on whether this pair of countries signs a PTA in force and belongs to the same trade bloc. This finding supports the argument that whether multinational corporations adopt a strategy of horizontal FDI, or vertical FDI, or both, depends on the trade barriers between the home country and host country. Accordingly, horizontal FDI dominates when the two countries are not in the same trade bloc. A BIT will then encourage this FDI strategy and in turn decrease the trade flow from the home country to the host country. In contrast, if the two countries already have a PTA and join the same trade bloc, then signing a BIT will benefit the motive of vertical FDI and enhance the positive effect of BIT on trade through a rise in the trade of intra-firm intermediate goods trade.

We also summarize the estimated trade effects from different econometric models in Table 8 for ease of comparison. Apart from the findings already discussed previously, we emphasize here that the combined effects of having both PTA and BIT are shown in the rows of PPML and those of 2SPPML. The effects found for these two models are $\big[\exp(0.5302 - 0.2560) - 1\big] \times 100\% = 31.55\%$ and $\big[\exp(1.0040 - 0.6309) - 1\big] \times 100\% = 45.22\%$, respectively.

IV. Conclusion

This study contributes to the literature on modeling selectivity in a Poisson model via a full-information method that explicitly specifies both the selection process of the BEEVs and the Poisson regression equation. We develop a novel two-stage Poisson pseudo-maximum likelihood (2SPPML) estimator by extending the econometric model of Terza (1998) to estimate the effects of dual BEEVs

Table 4: Descriptive Statistics of the Egger et al. (2011) Dataset

Variable	Description	Mean	SD	Min.	Max.
X_{ij}	Nominal exports in million US dollars	305.9274	3,257.2670	0	213,763.06
I_{ij}	Indicator variable taking a value one if $X_{ij} > 0$	0.6280	0.4834	0	1.0000
PTA_{ij}	Indicator variable taking a value one if two countries belong to a common PTA since 2005 or earlier	0.2226	0.4160	0	1.0000
BIT_{ij}	Indicator variable taking a value one if two countries belong to a common BIT since 2005 or earlier.	0.1879	0.3907	0	1.0000
$DIST_{ij}*$	Log distance divided by 10	0.8200	0.0827	0.3247	0.9419
$BORD_{ij}$	Common border indicator variable	0.0210	0.1432	0	1.0000
$LANG_{ij}$	Common language/ethnicity indicator variable	0.1393	0.3463	0	1.0000
$COLONY_{ij}$	Colony indicator variable	0.0152	0.1225	0	1.0000
$COMCOL_{ij}$	Common colonizer indicator variable	0.0777	0.2677	0	1.0000
$CURCOL_{ij}$	Colony after 1945 indicator variable	0.0084	0.0912	0	1.0000
$SMCTRY_{ij}$	Same country indicator variable	0.0088	0.0935	0	1.0000
$CONT_{ij}$	Same continent indicator variable	0.2303	0.4211	0	1.0000
$RGDP_{sum_{ij}}*$	Log of sum of real GDPs divided by 100	0.2523	0.0181	0.1993	0.3018
$RGDP_{sim_{ij}}$	Similarity of real GDPs	-2.1131	1.4877	-9.7690	-0.6931
DKL_{ij}	Difference between log of capital-labor relative factor endowments between pair ij	1.8217	1.2944	0.0001	6.1001
$DROWKL_{ij}$	Difference between log of capital-labor relative factor endowment between pair ij and rest of the world	1.4852	0.6493	0.0659	3.7327
$DURAB_{ij}*$	Durability of an exporter's and an importer's political regime divided by 100	0.2940	0.2922	0	1.0000
$POLCOMP_{ij}*$	Political competition index divided by 10	0.8896	1.9944	0	9.8000
$AUTOC_{ij}*$	Autocracy index divided by 10	0.7987	1.8947	0	9.8000
Observations			15,750		

Note: The rescaled variables are marked with ∗.

Table 5: Estimation Results of the Gravity Model for Trade

		Exogenous	
	PTA	BIT	PTA and BIT
Regression	$\mathbf{E}(X_{ij}\|\cdot)$	$\mathbf{E}(X_{ij}\|\cdot)$	$\mathbf{E}(X_{ij}\|\cdot)$
Estimator	PPML	PPML	PPML
	(1)	(2)	(3)
PTA_{ij}	0.5548	–	0.5302
	(0.1256)		(0.1176)
BIT_{ij}	–	-0.2939	-0.2560
		(0.0945)	(0.0947)
$DIST_{ij}*$	-4.9979	-5.5027	-4.9964
	(0.4924)	(0.5158)	(0.4821)
$BORD_{ij}$	0.7263	0.7375	0.6924
	(0.0726)	(0.0733)	(0.0702)
$LANG_{ij}$	0.1553	0.1719	0.1781
	(0.0813)	(0.0829)	(0.0795)
$CONT_{ij}$	0.2736	0.5010	0.2912
	(0.1222)	(0.0905)	(0.1151)
$DURAB_{ij}*$	-0.3789	-0.4334	-0.3686
	(0.0899)	(0.0786)	(0.0841)
$POLCOMP_{ij}*$	0.7374	0.7290	0.8956
	(0.3266)	(0.3271)	(0.3100)
$AUTOC_{ij}*$	-1.0387	-1.0075	-1.2089
	(0.3254)	(0.3177)	(0.3155)
$CURCOL_{ij}$	0.7246	0.7012	0.7987
	(0.1695)	(0.1836)	(0.1652)
$COLONY_{ij}$	–	–	–
$COMCOL_{ij}$	–	–	–
$SMCTRY_{ij}$	–	–	–
γ_{ij}	–	–	–
μ_2	–	–	–
μ_3	–	–	–

Notes: All regressions include importer and exporter fixed effects. Standard errors are in the parentheses. The rescaled variables are marked with $*$ as described in Table 4. Since we model the endogeneity as Eq. (5), μ_2 and μ_3 measure the potential endogeneity for PTA_{ij} and BIT_{ij}, respectively, and γ estimates the level of correlation between these two BEEVs.

Table 6: Estimation Results of the Gravity Model for Trade

	Endogenous			
	PTA		BIT	
Regression Estimator	$\mathbf{Pr}(PTA_{ij}=1\vert\cdot)$ Probit ML	$\mathbf{E}(X_{ij}\vert\cdot)$ PPML	$\mathbf{Pr}(BIT_{ij}=1\vert\cdot)$ Probit ML	$\mathbf{E}(X_{ij}\vert\cdot)$ PPML
	(1)	(2)	(3)	(4)
PTA_{ij}	–	1.1471	–	–
		(0.3847)		
BIT_{ij}	–	–	–	-0.7276
				(0.2052)
$DIST_{ij}*$	-10.7370	-3.9706	-4.8596	-6.2448
	(0.5006)	(0.6780)	(0.3679)	(0.6384)
$BORD_{ij}$	-0.4687	0.7405	-0.8360	0.6214
	(0.1652)	(0.0757)	(0.1368)	(0.0842)
$LANG_{ij}$	-0.1193	0.2079	0.3356	0.2184
	(0.0676)	(0.0701)	(0.0637)	(0.0782)
$CONT_{ij}$	0.7650	0.1506	-0.0198	0.4911
	(0.0499)	(0.1582)	(0.0525)	(0.0907)
$DURAB_{ij}*$	-0.7190	-0.4054	1.9551	-0.2628
	(0.0905)	(0.0862)	(0.1292)	(0.1074)
$POLCOMP_{ij}*$	-0.4829	1.0437	0.3355	0.7295
	(0.1029)	(0.2712)	(0.0932)	(0.3222)
$AUTOC_{ij}*$	0.4800	-1.3906	-0.4303	-1.0219
	(0.1109)	(0.2909)	(0.1005)	(0.3155)
$CURCOL_{ij}$	0.5189	0.6179	-0.5549	0.7570
	(0.2741)	(0.1855)	(0.2296)	(0.1840)
$COLONY_{ij}$	0.1356	–	1.0337	–
	(0.2107)		(0.1830)	
$COMCOL_{ij}$	0.5519	–	0.2571	–
	(0.0798)		(0.0809)	
$SMCTRY_{ij}$	1.2275	–	-0.0703	–
	(0.3279)		(0.1817)	
γ_{ij}	–	–		–
μ_2	–	-0.3708	–	–
		(0.1818)		
μ_3	–	–	–	0.2412
				(0.0869)

Notes: All regressions include importer and exporter fixed effects. Standard errors are in the parentheses. The rescaled variables are marked with $*$ as descibed in Table 4. Since we model the endogeneity as Eq. (5), μ_2 and μ_3 measure the potential endogeneity for PTA_{ij} and BIT_{ij}, respectively, and γ estimates the level of correlation between these two BEEVs.

Table 7: Estimation Results of the Gravity Model for Trade

Regression Estimator	$\mathbf{Pr}(PTA_{ij}=1\|\cdot)$ Bivariate Probit ML (1)	$\mathbf{Pr}(BIT_{ij}=1\|\cdot)$ Bivariate Probit ML (2)	Endogenous PTA and BIT: $\mathbf{E}(X_{ij}\|\cdot)$ PPML (3)	$\mathbf{E}(X_{ij}\|\cdot)$ PPML (4)
PTA_{ij}	–	–	1.0040	0.8342
			(0.4219)	(0.4437)
BIT_{ij}	–	–	-0.6309	-0.6893
			(0.2158)	(0.2176)
$PTA_{ij} \times BIT_{ij}$				0.2762
				(0.1353)
$DIST_{ij}*$	-10.7430	-4.8873	-4.8505	-4.9253
	(0.4995)	(0.3678)	(0.7332)	(0.7171)
$BORD_{ij}$	-0.4727	-0.8371	0.6028	0.6437
	(0.1646)	(0.1361)	(0.0876)	(0.0840)
$LANG_{ij}$	-0.1201	0.3370	0.2584	0.2358
	(0.0676)	(0.0637)	(0.0680)	(0.0674)
$CONT_{ij}$	0.7649	-0.0195	0.1822	0.2244
	(0.0499)	(0.0526)	(0.1608)	(0.1657)
$DURAB_{ij}*$	-0.7156	1.9612	-0.2310	-0.2540
	(0.0898)	(0.1283)	(0.1046)	(0.1046)
$POLCOMP_{ij}*$	-0.4781	0.3397	1.1150	1.1981
	(0.1029)	(0.0934)	(0.2520)	(0.2512)
$AUTOC_{ij}*$	0.4747	-0.4328	-1.4746	-1.5594
	(0.1109)	(0.1007)	(0.2789)	(0.2829)
$CURCOL_{ij}$	0.5134	-0.5650	0.7555	0.7967
	(0.2752)	(0.2296)	(0.1771)	(0.1760)
$COLONY_{ij}$	0.1424	1.0399	–	
	(0.2115)	(0.1833)		
$COMCOL_{ij}$	0.5507	0.2514	–	
	(0.0797)	(0.0809)		
$SMCTRY_{ij}$	1.2344	-0.0757	–	
	(0.3296)	(0.1814)		
γ	0.0626			
	(0.0281)			
μ_2	–	–	-0.2915	-0.2459
			(0.2091)	(0.2158)
μ_3	–	–	0.2199	0.1895
			(0.0948)	(0.0891)

Notes: All regressions include importer and exporter fixed effects. Standard errors are in the parentheses. In column (3) & (4), robust standard errors are reported using the method in White (1994) and Wooldridge (2010). The rescaled variables are marked with $*$ as descibed in Table 4. Since we model the endogeneity as Eq. (5), μ_2 and μ_3 measure the potential endogeneity for PTA_{ij} and BIT_{ij}, respectively, and γ estimates the level of correlation between these two BEEVs.

Table 8: Summary of Estimated Trade Effects

Econometric Method	Effects on Trade Values		
	PTA	BIT	Combined
PPML	69.93%	−22.59%	31.55%
PPML of Egger et al. (2011)	214.90%	–	–
	–	−51.69%	–
2SPPML	172.92%	−46.79%	45.22%
	130.30%	−49.81%	52.36%

Notes: PPML summarizes the estimates in column (3) of Table 5. PPML of Egger et al. (2011) reports the estimates in columns (2) and (4) of Table 6. 2SPPML shows the estimates in columns (3) and (4) of Table 7.

in the count regression model. Our methodology can deal with the structural gravity model based on the Poisson model consisting of both two-way fixed effects and dual BEEVs. Since we derive the analytical form of the gradient and Hessian matrices of the log-likelihood function of the proposed model, we are able to accelerate the computational speed by thousands of times as compared to the one relying on software to approximate these gradient and Hessian matrices. In so doing, we conduct the Monte Carlo simulations with a large number of data observations in an econometric model and estimate hundreds of parameters with ease. Doing so shows its promising potential for application in many fields such as international economics, health economics, management, and industrial organization.

For the purpose of illustration and comparison, we apply our method to the dataset of Egger et al. (2011) covering 126 countries for the year 2005 in order to evaluate the impacts of PTAs and BITs on bilateral trade flows of country pairs. We find that the decisions of signing PTAs and BITs positively correlate. While PTAs have a strong positive impact on trade, which is consistent with the literature, BITs impose a negative effect on bilateral trade flow, supporting the proximity-concentration motive of horizontal FDI. Most notably, we also find an interaction effect exists among the effects of these policy instruments. Specifically speaking, the trade effect of BITs depends on whether this country pair possesses a PTA and belongs to the same trade bloc.

Relative to the case where neither a PTA nor a BIT exists between a specific country pair, establishing a PTA alone, signing a BIT alone, and having both a PTA and a BIT increase trade by 130.30%, decrease trade by 49.81%, and raise trade by 52.36%, respectively. The positive coefficient of the $PTA \times BIT$ term indicates that the negative effect of BITs on trade has been mitigated when a PTA exists between the country pair. This finding implies the distinct motives of FDIs between the situation with PTAs and the case without PTAs. For country pairs that are not in the same trade bloc, the horizontal FDIs dominate market access and BITs decrease the bilateral trade flows of country pairs. In contrast, if the two countries already have a PTA and are in the same trade bloc, then a BIT will encourage multinational firms to conduct a vertical FDI strategy and increase the bilateral trade flow by raising trade of intra-firm intermediate goods through global value chains.

Notes

1. PPML offers several benefits when estimating gravity equations that economists cannot easily overlook. Notably, Santos Silva and Tenreyro (2006) find that the PPML estimator remains consistent under heteroskedasticity, while the OLS estimator is biased. In addition, the PPML estimator is asymptotically unbiased with a single fixed effect or a two-way fixed effect (Fernández-Val and Weidner, 2016).

2. Country pairs may theoretically self-select into PEIAs to enhance trade and welfare. Influential research by Baier and Bergstrand (2004) and Bergstrand and Egger (2013) indicates that

the establishment of PEIAs is driven by various factors associated with trade levels, such as distance, remoteness, country size, and capital-labor ratio.

3. This method, originally developed by Terza (1998), has been widely applied in various scientific fields such as finance and health economics. For example, Fahlenbrach (2009) explore the comparative effects of founder-CEOs and successor-CEOs on a company's acquisition count. Kenkel and Terza (2001) investigate the effects of receiving physician advice on a drinker's alcohol consumption.

4. Terza (1998) develops a Poisson model that accommodates one BEEV and provides three different estimators: a full information maximum likelihood estimator, a two-stage method of moments estimator, and a non-linear weighted least-square estimator.

5. The choice of an econometric model should be contingent upon the specific economic question under consideration. For an extensive discussion comparing instrumental variables with structural models, see Hechman and Urzúa (2010). They discuss the limitation of IV estimation.

6. Fernández-Val and Weidner (2016) show the estimates of two-way fixed effect probit or Poisson models to be asymptotically unbiased and consistent. This implies that, under appropriate regularity conditions, the first-stage estimation of bivariate probit model described by Eqs. (3)–(4) is also asymptotically unbiased and consistent. Furthermore, integrating these first-stage estimates into the second-stage estimation of Poisson model (Eq. (2)) preserve asymptotically unbiasedness and consistency. In conclusion, our proposed two-stage estimator is asymptotically unbiased and consistent. Our Monte Carlo experiments further confirm such results. However, it is important to note that such properties do not hold in the three-way fixed effects case, as demonstrated by Weidner and Zylkin (2021), who elaborate on the distinctions from Fernández-Val and Weidner (2016)'s findings.

7. The terms "pseudo-likelihood" and "pseudo-likelihood" are used interchangeably when the distributional assumptions of their underlying econometric models are not satisfied (Gourieroux and Monfort, 1993). Following the conventions of the international economics literature, we use the term "pseudo-likelihood" throughout this paper.

8. The negative trade effect of BITs is in stark contrast with the results of Heid and Vozzo (2020) who show a strong positive effect of BITs, which supports vertical FDI. While they include 1990–2015 international and domestic trade data for 172 countries from the EORA26 database,

we use the dataset of Egger et al. (2011) covering 126 countries for the year 2005 for illustration and comparison.

9. With their calibrated quantitative model, Tintelnot (2017) explores how the trade and investment agreement between Canada and the European Union (Comprehensive Economic and Trade Agreement, CETA) affects the structure of multinationals' global production. He finds that EU multinationals would divert around 5% of their production from the United States to Canada if the agreement yields a 20% reduction of variable and fixed production costs between the signatories.

10. See Greene (2012) concerning the ML estimation for the bivariate probit models in greater detail.

References

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

Baier, Scott L., and Jeffrey H. Bergstrand. 2004. "Economic Determinants of Free Trade Agreements." *Journal of International Economics*, 64(1): 29–63.

Baier, Scott L., and Jeffrey H. Bergstrand. 2007. "Do Free Trade Agreements actually Increase Members' International Trade?" *Journal of International Economics*, 71(1): 72–95.

Berger, Axel, Matthias Busse, Peter Nunnenkamp, and Martin Roy. 2013. "Do Trade and Investment Agreements Lead to More FDI? Accounting for Key Provisions Inside the Black Box." *International Economics and Economic Policy*, 10(2): 247–275.

Bergstrand, Jeffrey H., and Peter Egger. 2013. "What Determines BITs?" *Journal of International Economics*, 90(1): 107–122.

Blanchard, Emily J. 2010. "Reevaluating the role of trade agreements: Does investment globalization make the WTO obsolete?" *Journal of International Economics*, 82(1): 63–72.

Blanchard, Emily J., and Xenia Matschke. 2015. "U.S. multinationals and preferential market access." *Review of Economics and Statistics*, 97(4): 839–854.

Blanchard, Emily J., Chad P. Bown, and Robert C. Johnson. 2016. "Global supply chains and trade policy." *NBER Working Papers No. 21883.*

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

Egger, Peter H., and Filip Tarlea. 2021. "Comparing Apples to Apples: Estimating Consistent Partial Effects of Preferential Economic Integration Agreements." *Economica*, 88: 456–473.

Egger, Peter, Mario Larch, Kevin E. Staub, and Rainer Winkelmann. 2011. "The Trade Effects of Endogenous Preferential Trade Agreement." *American Economic Journal: Economic Policy*, 3(3): 113–143.

Fahlenbrach, Rüdiger. 2009. "Founder-CEOs, Investment Decisions, and Stock Market Performance." *Journal of Financial and Quantitative Analysis*, 44(2): 439–466.

Fally, Thibault. 2015. "Structural Gravity and Fixed Effects." *Journal of International Economics*, 97(1): 76–85.

Feenstra, Robert C. 2004. *Advanced International Trade: Theory and Evidence.* Princeton, New Jersey:Princeton University Press.

Fernández-Val, Iván, and Martin Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large N, T." *Journal of Econometrics*, 192: 291–312.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52(3): 681–700.

Gourieroux, Christian, and Alain Monfort. 1993. "Pseudo-Likelihood Methods." In *Econometrics.* Vol. 11 of *Handbook of Statistics*, 335–362. Elsevier.

Greene, William H. 2012. *Econometric Analysis.* . 7th ed., Upper Saddle River, NJ:Prentice Hall.

Hechman, James J., and Sergio Urzúa. 2010. “Comparing IV with Structural Models: What Simple IV Can and Cannot Identify.” *Journal of Econometrics*, 156(1): 27–37.

Heckman, James J. 1978. “Dummy Endogenous Variables in A Simultaneous Equation System.” *Econometrica*, 46(6): 931–959.

Heid, Benedikt, and Isaac Vozzo. 2020. “The International Trade Effects of Bilateral Investment Treaties.” *Economics Letters*, 196: 109569.

Jochmans, Koen. 2022. “Bias in Instrumental-Variable Estimators of Fixed-Effect Models for Count Data.” *Economics Letters*, 212: 110318.

Jochmans, Koen, and Vincenzo Verardi. 2022. “Instrumental-Variable Estimation of Exponential Regression Models with Two-Way Fixed Effects with an Application to Gravity Equations.” *Journal of Applied Econometrics*, 37(6): 1121–1137.

Kenkel, Donald S., and Joseph V. Terza. 2001. “The Effect of Physician Advice on Alcohol Consumption: Count Regression with An Endogenous Treatment Effect.” *Journal of Applied Econometrics*, 16(2): 165–184.

Markusen, James R. 2002. *Multinational Firms and the Theory of International Trade.* Cambridge, Mass:MIT Press.

Mullahy, John. 1997. “Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior.” *The Review of Economics and Statistics*, 79(4): 586–593.

Neary, J. Peter. 2009. “Trade Costs and Foreign Direct Investment.” *International Review of of Economics and Finance*, 18(2): 207–218.

Redding, Stephen, and Anthony J. Venables. 2004. "Economic geography and international inequality." *Journal of International Economics*, 62(1): 53–82.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641–658.

Terza, Joseph V. 1998. "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects." *Journal of Economics*, 84(1): 129–154.

Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy.* New York:The Twentieth Century Fund.

Tintelnot, F. 2017. "Global Production with Export Platforms." *Quarterly Journal of Economics*, 132(1): 157–209.

UNCTAD. 1998. *World Investment Report: Trends and Determinants.* New York:United Nations.

Weidner, Martin, and Thomas Zylkin. 2021. "Bias and Consistency in Three-Way Gravity Models." *Journal of International Economics*, 132: 103513.

White, Halbert. 1994. *Estimation, Inference and Specification Analysis.* Cambridge University Press.

Windmeijer, Frank A. G., and J. M. C. Santos Silva. 1997. "Endogeneity in Count Data Models: An Application to Demand for Health Care." *Journal of Applied Econometrics*, 12(3): 281–294.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Wooldridge, Jeffrey M. 2014. "Quasi-maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables." *Journal of Econometrics*, 182(1): 226–234.

Yeaple, Stephen Ross. 2003. "The complex integration strategies of multinationals and cross country dependencies in the structure of foreign direct investment." *Journal of International Economics*, 60(2): 293–314.