CIRANO

Knowledge into action

BEYOND PPML: EXPLORING MACHINE LEARNING ALTERNATIVES FOR GRAVITY MODEL ESTIMATION IN INTERNATIONAL TRADE

LUCIEN CHAFFA
MARTIN TRÉPANIER
THIERRY WARIN

2025s-14
WORKING PAPER

ISSN 2292-0838 (online version)

Beyond PPML: Exploring Machine Learning Alternatives for Gravity Model Estimation in International Trade

Lucien Chaffa[*], *Martin Trépanier*[†], *Thierry Warin*[‡]

Abstract/Résumé

This study investigates the potential of machine learning (ML) methods to enhance the estimation of the gravity model, a cornerstone of international trade analysis that explains trade flows based on economic size and distance. Traditionally estimated using methods such as the Poisson Pseudo Maximum Likelihood (PPML) approach, gravity models often struggle to fully capture nonlinear relationships and intricate interactions among variables. Leveraging data from Canada and the US, one of the largest bilateral trading relationships in the world, this paper conducts a comparative analysis of traditional and ML approaches. The findings reveal that ML methods can significantly outperform traditional approaches in predicting trade flows, offering a robust alternative for capturing the complexities of global trade dynamics. These results underscore the value of integrating ML techniques into trade policy analysis, providing policymakers and economists with improved tools for decision-making.

Cette étude examine le potentiel des méthodes d'apprentissage automatique (ML) pour améliorer l'estimation du modèle de gravité, une méthode clé de l'analyse du commerce international qui explique les flux commerciaux en fonction de la taille de l'économie et de la distance. Traditionnellement estimés à l'aide de méthodes telles que l'approche du pseudo-maximum de vraisemblance de Poisson (PPML), les modèles de gravité ont souvent du mal à saisir pleinement les relations non linéaires et les interactions complexes entre les variables. En s'appuyant sur les données du Canada et des États-Unis, l'une des relations commerciales bilatérales les plus importantes au monde, cet article effectue une analyse comparative des approches traditionnelles et des approches par apprentissage automatique. Les résultats révèlent que les méthodes de ML peuvent être nettement plus performantes que les approches traditionnelles pour prédire les flux commerciaux, offrant ainsi une alternative robuste pour saisir les complexités de la dynamique du commerce mondial. Ces résultats soulignent la valeur de l'intégration des techniques de ML dans l'analyse de la politique commerciale, fournissant aux décideurs politiques et aux économistes des outils améliorés pour la prise de décision.

[*] CIRANO
[†] Polytechnique Montréal, CIRRELT
[‡] HEC Montréal, CIRANO

Keywords/Mots-clés: Gravity Model, PPML Machine Learning, International Trade, Trade Policy Analysis / Modèle de gravité, PPML, apprentissage automatique, commerce international, analyse de la politique commerciale

JEL Codes/Codes JEL: F10, F14, C13, C45

Pour citer ce document / To quote this document

Chaffa, L., Trépanier, M., & Warin, T. (2025). Beyond PPML: Exploring Machine Learning Alternatives for Gravity Model Estimation in International Trade (2025s-14, Cahiers scientifiques, CIRANO.) https://doi.org/10.54932/BFKY4995

Introduction

The gravity model has long occupied a central position in the analysis of international trade, offering a systematic framework that links trade flows between two regions to their respective economic sizes and the geographic distance separating them. Building on analogies to Newtonian physics, this conceptual framework posits that larger economies trade more intensively, while distance and other trade barriers exert dampening effects on the intensity of bilateral exchanges. Over the decades, this model has been refined to account for a host of additional factors, including border effects (McCallum, 1995), multilateral resistance terms (Anderson & Van Wincoop, 2003), and more complex specifications that incorporate unobserved heterogeneity (Silva & Tenreyro, 2006). Despite these advances, considerable methodological challenges remain. Traditional estimation techniques such as Ordinary Least Squares (OLS) and Poisson Pseudo Maximum Likelihood (PPML) often struggle with nonlinearities, zero trade flows, and the growing dimensionality of modern datasets.

These challenges have prompted researchers to explore alternative methods that may capture the intricate dynamics of trade relationships more effectively. Recent literature has begun to investigate whether machine learning (ML) approaches, renowned for their predictive power in large and complex datasets, can enrich gravity model estimations by uncovering patterns that elude standard econometric procedures. The proliferation of high-frequency and granular trade data—coupled with ongoing enhancements in computational capacity—presents a clear opportunity to integrate ML-based techniques into trade modeling. Yet the adoption of such methods remains comparatively limited in mainstream trade analysis, in part because conventional estimators maintain an advantage in terms of coefficient interpretability (Santos Silva & Tenreyro, 2022).

This study addresses the following overarching research question: to what extent can machine learning methods improve the predictive accuracy of gravity models in explaining bilateral trade flows, including those involving zero observations, when compared to more traditional estimation approaches such as OLS, PPML, Gamma Pseudo Maximum Likelihood (GPML), and Negative Binomial Pseudo Maximum Likelihood (NBPML)? To investigate this question, we employ a comprehensive dataset covering trade between Canadian provinces and U.S. states, capturing one of the world's largest bilateral trading relationships. By examining both scenarios in which zero trade flows are included and excluded from the estimation process, we aim to clarify how distinct methodologies manage sparse or zero-inflated data. In doing so, we evaluate the trade-offs among predictive accuracy, robustness to zero flows, and interpretability across a range of estimation procedures.

Our inquiry is motivated by dual objectives. First, we seek to determine whether modern ML techniques—specifically Random Forest, XGBoost, and Neural Networks—deliver meaningful improvements over econometric estimators that have become standard in the literature following the work of (Silva & Tenreyro, 2006). Second, by comparing predictive performance against interpretative clarity, we aim to offer guidance on the conditions under which each class of methods may be optimally employed. This is of particular relevance for policymakers and economists who rely on trade flow projections to inform decisions relating to tariff policies, regional integration agreements, and other macroeconomic interventions. By examining a rich empirical setting and systematically comparing methods, this study contributes to the evolving discourse on quantitative trade modeling and illuminates the potential value of machine learning in addressing the methodological shortcomings that have historically complicated the gravity model's application.

1. Literature Review

1.1. Historical Foundations

The gravity model has established itself as a pivotal framework in the analysis of trade and spatial flows, drawing its conceptual foundation from Newtonian physics, where trade interactions are treated analogously to gravitational forces. The seminal work of Anderson & Van Wincoop (2003) significantly advanced the theoretical framework of the gravity model by introducing multilateral resistance terms, which account for the relative trade costs faced by countries in a network of trade relationships.

This theoretical enhancement has been crucial in refining the understanding of how trade flows are influenced not only by the economic size of trading partners but also by the distance between them and the trade barriers they face (Anderson, 2011). In the empirical realm, the gravity model has undergone substantial refinement, particularly with the introduction of the Poisson pseudo-maximum likelihood (PPML) estimator by Silva & Tenreyro (2006).

This methodological advancement addresses critical issues such as zero trade flows and heteroskedasticity, which have historically plagued trade data analysis. The PPML estimator has been widely adopted in various studies, demonstrating its effectiveness in providing robust estimates of trade flows (Akhvlediani & Śledziewska, 2017). For instance, Karemera et al. (2009) Karemera et al. (2009) highlighted the empirical success of gravity models in capturing the complexities of bilateral trade flows, emphasizing their log-linear relationship with economic size and transaction costs (Karemera et al., 2009).

Over the last two decades, the gravity model has become the dominant analytical tool for assessing the impacts of trade agreements, tariffs, and other policy interventions. Its robustness and flexibility have facilitated extensive research across various domains, including trade, migration, and investment. Studies such as those by Masudur Rahman & Kim (2012) and Akhtar & Ghani (2010) illustrate the model's adaptability in different contexts, demonstrating its utility in evaluating trade potential and the effects of regional integration (Masudur Rahman & Kim, 2012; Akhtar & Ghani, 2010).

Furthermore, the gravity model's application has extended beyond traditional trade analysis to encompass agricultural trade, as evidenced by Muganyi & Chen (2016) exploration of China's agricultural trade flows, which underscores the model's versatility (Muganyi & Chen, 2016). The gravity model's empirical robustness is further supported by its historical application, tracing back to the pioneering works of Tinbergen (1962) and Poyhonen (1963), who laid the groundwork for estimating trade volumes based on economic size and distance (Esmaeili & Pourebrahim, 2011).

This foundational work has been built upon by numerous scholars, leading to a rich body of literature that continues to evolve. Recent studies, such as those by Beenstock et al. (2015) and Elshehawy et al. (2014), have employed the gravity model to analyze immigration flows and export determinants, respectively, showcasing its broad applicability across various economic phenomena (Beenstock et al., 2015; Elshehawy et al., 2014).

1.2. Key Methodological Developments

The methodological sophistication of gravity models has evolved significantly, driven by advancements in both econometrics and computational tools. This evolution has expanded the scope and applicability of the model in several key areas:

- Heterogeneity and Weighting: Traditional gravity models often assume homogeneity in coefficients across observations, which can lead to biased estimates when this assumption does not hold. Recent studies, such as those by Breinlich et al. (2024), emphasize the importance of explicitly modeling heterogeneity. These works demonstrate that estimators like Poisson pseudo-maximum likelihood (PPML) assign greater weight to larger trade flows, which can skew results if heterogeneity is unmodeled. By incorporating varying coefficients or using techniques such as quantile regression, researchers can better capture the diversity in trade relationships across different contexts and countries (Anderson, 2011).

- Dynamic and Spatial Dimensions: The incorporation of dynamic panel data models and spatial econometric techniques has significantly enhanced the gravity model's ability to capture temporal and spatial dependencies. Dynamic models allow for the analysis of trade persistence, reflecting how past trade flows influence current trade patterns. Spatial econometrics, on the other hand, accounts for the interconnectedness of global markets by considering spatial autocorrelation and spillover effects. This methodological advancement is particularly relevant for understanding how trade relationships evolve over time and how regional dynamics can influence bilateral trade flows (Akhvlediani & Śledziewska, 2017).

- Dealing with Zeroes and Small Flows: While PPML effectively addresses zero trade flows, recent advancements have sought to refine these methods further. Techniques such as Gamma PML and multinomial PML have been proposed as alternatives, offering nuanced approaches to handle small or zero flows. These methods allow researchers to model the distribution of trade flows more accurately, accommodating the prevalence of zero and small trade values in many datasets. This is crucial for ensuring that estimates reflect the underlying economic realities without being distorted by the peculiarities of the data (Karemera et al., 2009).

- Econometric Tools for Causal Inference: The integration of instrumental variables, synthetic control methods, and causal machine learning into gravity model estimation has become increasingly prevalent. These tools enable researchers to disentangle causal relationships in complex settings, addressing issues of endogeneity that can arise in observational data. For instance, using instrumental variables can help identify the causal impact of trade policies on trade flows, while synthetic control methods allow for the comparison of treated and untreated units in a quasi-experimental framework (Masudur Rahman & Kim, 2012).

2. Study Design

This study employs a multi-stage methodological framework that combines both traditional econometric methods and advanced machine learning techniques to estimate gravity models of trade. The approach entails specifying two alternative versions of the gravity equation, drawing respectively on McCallum (1995) and Anderson & Van Wincoop (2003), and then applying a series of estimation procedures to gauge each method's predictive performance. The data selection and processing strategies are designed to systematically compare the capabilities of these different estimators under conditions that commonly arise in trade analysis, including the presence of zero trade flows and the need for robust yet interpretable coefficient estimates.

The core dataset consists of trade flows between Canada's ten provinces and between Canadian provinces and fifty U.S. states, resulting in a cross-sectional sample for a single reference year. These data are retrieved from Statistics Canada's publicly available trade databases, ensuring consistency in measurement and reporting standards. The dependent variable is bilateral trade value, while the explanatory variables include GDP of the exporting and importing regions, geographical distance measured by centroid-to-centroid calculations, and a border indicator distinguishing intra-Canadian provincial trade from cross-border provincial-state trade. Supplementary data checks are conducted by comparing centroid distances with alternative distance measures such as capital-to-capital distances, though the centroid-based approach is maintained in the main results for consistency and ease of replication.

Two sets of estimations are carried out. In the first set, all trade flows are included, including zero observations, to assess the performance of each estimator in managing sparse or zero-inflated data. In the second set, zero flows are excluded, providing a comparative benchmark that highlights the sensitivity of results to the presence or absence of these often-problematic observations. For both sets, the sample is split into training and test partitions. The models are fitted using the training partition, and their performance is subsequently evaluated using observations in the test partition. Root Mean Square Error (RMSE) and R-square constitute the principal metrics for measuring predictive accuracy. In addition, Mean Absolute Error (MAE) is reported in order to capture the average magnitude of the errors in a more direct manner.

Within each estimation exercise, four econometric techniques—Ordinary Least Squares (OLS), Poisson Pseudo Maximum Likelihood (PPML), Gamma Pseudo Maximum Likelihood (GPML), and Negative Binomial Pseudo Maximum Likelihood (NBPML)—are systematically compared. These traditional methods serve as a benchmark for model interpretability and coefficient stability, particularly in the context of zero trade flows. Alongside these methods, three machine learning algorithms—Random Forest, XGBoost, and Neural Networks—are employed to capture potential nonlinearities and interactions in the determinants of bilateral trade. The machine learning approaches are tested both on level-transformed and log-transformed versions of the data, in order to ascertain whether log-transformations enhance predictive accuracy by mitigating the influence of large outliers in trade flows.

This experimental design makes it possible to assess each estimator's strengths and limitations. Traditional approaches, known for their well-established theoretical underpinnings and interpretability, can be benchmarked against machine learning methods, which often excel in predictive tasks but may be less transparent in their coefficient estimates. By applying these methods to two canonical versions of the gravity equation (Anderson & Van Wincoop, 2003; McCallum, 1995) and systematically including or excluding zero trade flows, the study generates insights into how methodological choices affect both predictive accuracy and interpretative clarity. The following sections detail the estimation results and highlight the context in which each class of models may be most suitable for empirical trade analysis.

2.1. Selection of gravity equations estimated

The traditional form of gravity model is inspired by Newton law of gravitation states as follows: $F_{ij} = G\frac{M_i M_j}{D_{ij}^2}$, where the force F between two bodies i and j with $i \neq j$ is proportional to the masses M of theses bodies and inversely proportional to the square of the distance between their centers. G is the gravitational constant.

In international trade, this model is adapted as follows:

$$X_{ij} = G\frac{Y_i^{\beta_1} Y_j^{\beta_2}}{D_{ij}^{\beta_3}}$$

Where X, the trade flow is explained by Y_i and Y_j that represents the masses of the exporting and importing countries/ regions. Generally, we approximate these masses by GDP of each country/region. D_{ij} is the distance between the countries/regions.

To incorporate additional variables that influence trade flows, D_{ij} is replaced with t_{ij}, representing transaction costs, including distance. This yields the following equation:

$$X_{ij} = G\frac{Y_i^{\beta_1}Y_j^{\beta_2}}{t_{ij}^{\beta_3}} \qquad (1)$$

The straightforward way to estimate (1) is by applying log transformation and use standard estimation methods such as OLS.

In McCallum (1995), investigating the effects of national borders on trades between the US and Canada, he define trade cost as $t_{ij} = D_{ij}exp(\delta ij)$ with δ_{ij} takes 1 if the exporting and importing region are Canadian provinces and 0 if there state-provinces regions.

Anderson & Van Wincoop (2003) suggest that the estimates from McCallum (1995) are biased, and one needs to account for multilateral resistance to identify the national border effect. Thus, trade cost expression in that case could be approximated as follows: $t_{ij} = D_{ij}exp\left(\delta_{ij} + \eta_i + \eta_j\right)$, with η_i and η_j are exporter and importer fixed effects.

In this project we are interested in comparing the predictions power of gravity models' estimation methods. We compare traditional methods of estimation and machine learning methods of McCallum (1995) equations (2) and Anderson & Van Wincoop (2003) (3) state as follows:

$$X_{ij} = G\frac{Y_i^{\beta_1}Y_j^{\beta_2}}{D_{ij}^{\alpha}\left(exp(\delta_{ij})\right)^{\gamma}} \qquad (2)$$

$$X_{ij} = G\frac{Y_i^{\beta_1}Y_j^{\beta_2}}{D_{ij}^{\alpha}\left(exp(\delta_i j)\right)^{\alpha}\left(exp(\eta_i)\right)^{\gamma_1}\left(exp(\eta_j)\right)^{\gamma_2}} \qquad (3)$$

2.2. Selection of estimation methods

Traditional methods of estimation

Equations (2) and (3) was estimated by traditional methods as Ordinary Least Squares (OLS), respectively in McCallum (1995) and Anderson & Van Wincoop (2003). However, these traditional methods of estimating the gravity equation, particularly OLS have faced criticism for their biases and inefficiencies, particularly in the presence of zero trade flows and unobserved heterogeneity. Recent literature has increasingly favored the Poisson Pseudo Maximum Likelihood (PPML) estimator as a more robust alternative. Silva and Tenreyro argue that the PPML estimator is exceptionally well-suited for gravity equation estimation, as it effectively addresses the issues of heteroskedasticity and zero trade flows that plague OLS estimates (Santos Silva & Tenreyro, 2022). This sentiment is echoed by Breinlich et al., who demonstrate that PPML provides superior estimates even when disaggregated data is available, highlighting its robustness across different data scenarios (Breinlich et al., 2024). Other estimators, such as Gamma Pseudo Maximum Likelihood (GPML) and Negative Binomial Pseudo Maximum Likelihood (NBPML), have also gained attention for their robustness in handling trade data with a high incidence of zeros.

The GPML estimator is particularly noted for its ability to manage overdispersion in count data, which is often a characteristic of trade flow datasets. Silva & Tenreyro (2011) emphasize that GPML can provide consistent estimates even when the dependent variable contains a significant number of zeros, similar to the PPML. They argue that when both PPML and GPML yield similar coefficients, it indicates that the model is appropriately specified and that heteroskedasticity is effectively managed. This is further supported by findings from Gregori and Michela (Gregori & Nardo, 2021), who note that GPML performs well in contexts where the dependent variable is subject to a high number of zeros, reinforcing the robustness of this estimator in empirical applications.

On the other hand, the NBPML estimator is particularly useful when dealing with over-dispersed count data, as it allows for the modeling of variance that exceeds the mean. Ghazalian (2019) discusses the application of NBPML in estimating gravity models, highlighting its effectiveness alongside PPML and GPML in capturing the structural relationships inherent in trade data. This is crucial for ensuring that the estimates are not only statistically significant but also economically interpretable. The work of D'Ambrosio & Montresor (2022) further supports the preference for NBPML over OLS, particularly when analyzing datasets with significant zero trade flows, as it provides a more nuanced understanding of the underlying trade dynamics.

In summary, concerning point estimation of the equations (1) and (2) we compare OLS, PPML, GPML, and NBPML methods with some machine learning techniques of estimation.

Machine learning techniques of estimation

The application of machine learning techniques has gained traction in recent years. The estimation of traditional gravity models, which relate trade flows to economic size and distance between trading partners, could be enhanced through the integration of advanced machine learning methods such as Random Forests, XGBoost, and Neural Networks.

Random Forests (RF) is a powerful ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification or mean prediction for regression. This technique is particularly advantageous due to its ability to handle large datasets with numerous variables without overfitting, making it suitable for complex trade data (Mangalathu & Jeon, 2019). The flexibility and intuitive nature of RF allow it to capture nonlinear relationships and interactions between variables, which are often present in economic data (Mangalathu & Jeon, 2019). Moreover, RF has been successfully applied in various fields, demonstrating its robustness and effectiveness in predictive modeling.

XGBoost, or Extreme Gradient Boosting, is another machine learning technique that show promise in improving the predictive power of gravity models. XGBoost is an optimized implementation of gradient boosting that is designed for speed and performance. It excels in handling sparse data and can effectively manage missing values, which are common in trade datasets (Gopinath et al., 2021). The method's ability to incorporate regularization helps prevent overfitting, making it a strong candidate for estimating gravity models where the relationships between variables may be complex and nonlinear (Gopinath et al., 2021). Recent studies have highlighted the effectiveness of XGBoost in forecasting trade flows, showcasing its superior performance compared to traditional econometric methods (Park et al., 2024).

Neural Networks (NN), particularly deep learning architectures, have also been employed to enhance gravity model estimations. The Deep Gravity model proposed by Simini et al., (2022) illustrates how deep neural networks can be utilized to improve the predictive performance of traditional gravity models by incorporating non-linearity and additional geographical features. This model treats the gravity equation as a baseline and enhances it with hidden layers, allowing for more complex interactions among variables (Simini et al., 2022). The flexibility of neural networks enables them to learn intricate patterns in data, making them particularly suitable for high-dimensional datasets typical in trade analysis. Furthermore, the integration of geographic information into neural network models has been shown to significantly enhance the accuracy of predictions related to mobility and trade flows (Simini et al., 2022).

In this paper, we perform those three machines learning estimation methods (RF, XGBoost, and NN) and compare them with point estimation methods (OLS, PPML, GPML, and NBML). These methods are performed on equations (1) and (2) using Canadian interprovincial trade flow and province state trade flow between the US and Canada following McCallum (1995) and Anderson & Van Wincoop (2003).

2.3. Data Description

This study draws on trade flow data obtained from Statistics Canada to capture transactions within and across Canadian regions. First, interprovincial trade flows are derived from the Interprovincial and International Trade Flows datasets, covering commerce among Canada’s ten provinces and yielding 90 observations. Second, trade flows between Canada’s ten provinces and 50 US states are taken from the Canadian International Merchandise Trade dataset, resulting in 1,000 observations. Together, these sources yield a comprehensive dataset of 1,090 observations for the year 2021. A number of these observations feature zero trade values, primarily reflecting pairs of regions that do not engage in any recorded exchange; analyses are therefore conducted both with and without these zero observations to assess the robustness of the results under varying data conditions.

Distances between regions are computed using centroid-based measurements, although alternative methods, such as distances between provincial or state capitals, are also explored as a robustness check. In order to evaluate the performance of each estimation method for Equations (2) and (3), the dataset is partitioned into training and test subsets. Models are fitted using the training portion and subsequently evaluated based on their predictive accuracy in the test sample. The primary metrics for comparing predictive performance are the Root Mean Square Error (RMSE) and R-squared, with lower values of RMSE and higher values of R-squared indicative of superior explanatory power.

3. Results

3.1. Estimation of Equation (2)

Estimation of equation (2) without zero trade flows

Table 1 presents the estimated coefficients and statistical significance for equation (2) using four traditional econometric methods: OLS, PPML, GPML, and NBPML. The results show that the size of the origin and destination economies measured by their GDP has a positive and statistically significant effect on trade flows, confirming the fundamental intuition of the gravity model—larger economies trade more. Distance has a negative coefficient across all methods, indicating that trade decreases as geographic separation increases. The border variable is positive and significant, suggesting that interprovincial trade is higher than province-state trade. This result suggests that border significantly harms trade. Compared to the other methods, PPML produces lower coefficient estimates, likely due to its ability to handle zero trade flows and account for heteroskedasticity. The R-square values indicate that OLS and GPML fit the data better than PPML, while NBPML performs similarly to GPML.

Table 1: Traditional Models performance of equation (2)

Model Performance Comparison				
	OLS	PPML	GPML	NBPML
Origin size	1.09^{***}	0.89^{***}	0.90^{***}	1.00^{***}
	(0.05)	(0.05)	(0.05)	(0.001)
Destination size	1.64^{***}	0.97^{***}	0.98^{***}	0.92^{***}
	(0.06)	(0.05)	(0.05)	(0.001)
Distance	-1.89^{***}	-1.67^{***}	-1.72^{***}	-1.15^{***}

	(0.14)	(0.13)	(0.13)	(0.002)
Border	3.25^{***}	2.08^{***}	2.10^{***}	1.91^{***}
	(0.26)	(0.24)	(0.24)	(0.003)
Constant	-14.67^{***}	-4.49^{***}	-4.47^{***}	-9.24^{***}
	(1.43)	(1.32)	(1.32)	(0.03)
R-square	0.70	0.65	0.65	0.17
MAE	1.29	596.24	636.39	895.38
RMSE	1.87	3053.94	3340.50	4228.45
Observations	846	846	846	846

Note: This table presents estimated coefficients for equation (2) using OLS, PPML, GPML, and NBPML. Origine size and Destination size are approximate by GDP of regions; Distance is the distance between centroids of regions, and Border is dummy variable taking 1 if the trading regions are in the same country. R-square is computed on test dataset of 218 observations. Coefficients significance: *p<0.1; **p<0.05; ***p<0.01.

In Table 2 we report the performance metrics—Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-square —for machine learning models estimating equation (2). Among the models tested, XGBoost outperforms both Random Forest and Neural Networks, achieving the lowest RMSE and highest R-square, indicating superior predictive accuracy. The results also show that applying a logarithmic transformation to the variables improves model performance across all three machine learning methods. This suggests that log transformations help mitigate the impact of outliers and nonlinear relationships. Neural Networks perform the worst in both levels and logs, potentially due to data size constraints or hyperparameter tuning limitations.

Table 2: Machine Learning Models Performance of equation (2)

Model Performance Comparison						
	Variables in Level			Variables in Log		
	Random Forest	XGBoost	Neural Network	Random Forest	XGBoost	Neural Network
RMSE	2,914.70	2,552.32	3,570.10	1.49	1.33	1.83
MAE	624.79	565.73	973.49	1.05	0.94	1.20
R-square	0.59	0.65	0.27	0.81	0.85	0.73

Note: This table presents performance metrics (RMSE, MAE, and R-square) for machine learning models estimating equation (2). The models evaluated include Random Forest, XGBoost, and Neural Networks, with results reported for both level and log-transformed variables. Lower RMSE and MAE indicate better predictive accuracy, while higher R-square suggests better model fit.

Table 3 compares the predictive accuracy of traditional econometric models with machine learning methods for equation (2). Traditional methods such as OLS, PPML, GPML, and NBPML produce lower R-square values compared to the best-performing machine learning models. XGBoost and Random Forest, particularly when applied to log-transformed variables, achieve significantly higher R-square values, indicating their superior ability to capture complex trade flow patterns. R-square, reinforcing its primary role as a robust estimator for handling zero trade flows rather than maximizing predictive accuracy.

Table 3: Traditional vs Machine Learning Models Performance of equation (2)

Model Performance Comparison				
		Performance Metrics		
Estimation Methods	Model Type	RMSE	MAE	R-square
Regression	OLS	1.87	1.29	0.70
	PPML	4228.45	895.38	0.17
	GPML	3053.94	596.24	0.65
	NBPML	3340.50	636.39	0.65
Machine Learning	Random Forest	1.49	1.05	0.81
	XGBoost	1.33	0.94	0.85
	Neural Network	1.83	1.20	0.73

Note: This table compares the predictive performance of traditional econometric estimators (OLS, PPML, GPML, and NBPML) with machine learning models (Random Forest, XGBoost, and Neural Networks) for equation (2). Performance is evaluated using RMSE, MAE, and R-square, with results reported for both level and log-transformed variables.

Estimation of equation (2) with zero trade flows

The estimates for equation (2) when zero trade flows are included in the dataset, using PPML and NBPML are presented in Table 4 . The results show that PPML and NBPML yield significantly different coefficient magnitudes, reflecting their distinct ways of handling zero trade flows. NBPML achieves a much higher R-square (0.802) compared to PPML (0.217), suggesting that NBPML provides a better fit when modeling sparse trade data. The signs of all predictors estimates are what expected and consistent with empiric results in literature of gravity models.

Table 4: Traditional Models Performance of equation (2) with zeros trade flows

Model Performance Comparison		
	PPML	NBPML
Distance	-1.22***	-1.73***
	(0.002)	(0.15)
Origin size	0.97***	0.99***
	(0.001)	(0.06)
Destination size	0.95***	1.07***
	(0.001)	(0.06)
Border	1.79***	2.05***
	(0.003)	(0.28)
Constant	-8.71***	-6.65***
	(0.03)	(1.50)
R-square	0.22	0.82
MAE	930.28	852.79
RMSE	3974.63	4661.11
Observations	872	872

Note: This table presents estimated coefficients for equation (2) using PPML and NBPML when zero trade flows are included in the dataset. Origin size and Destination size are approximated by GDP, Distance represents centroid-based regional distances, and Border is a dummy variable taking 1 if the trading regions are in the same country. R-square is computed on test dataset of 218 observations. Coefficients significance: *p<0.1; **p<0.05; ***p<0.01.

Table 5 includes the performance of machine learning models when zero trade flows are included. Compared to Table 2, the overall model performance declines, as evidenced by higher RMSE values and lower R-square scores. Random Forest maintains the highest R-square among the three models, but XGBoost exhibits a lower RMSE, indicating a slightly better predictive fit. Neural Networks continue to perform worse than tree-based models, further suggesting that their application in gravity models may require additional fine-tuning or larger datasets.

The comparison of traditional and machine learning models when zero trade flows are also included in Table 5. NBPML achieves the highest R-square (0.80), surpassing both PPML and machine learning models, reinforcing its suitability for handling trade datasets with a significant number of zero observations. Machine learning models, particularly Random Forest and XGBoost, experience a decline in predictive accuracy, emphasizing the challenges of modeling sparse data with nonparametric approaches.

Table 5:Traditional vs Machine Learning Models Performance of equation (2) with zeros trade flows

Model Performance Comparison

		Performance Metrics		
Estimation Methods	Model Type	RMSE	MAE	R-square
Regression	PPML	3974.63	930.28	0.22
	NBPML	4661.11	852.79	0.82
Machine Learning	Random Forest	2678.07	689.71	0.53
	XGBoost	2190.64	748.74	0.68
	Neural Network	3120.31	864.29	0.40

Note: This table compares traditional econometric models (PPML and NBPML) with machine learning models (Random Forest, XGBoost, and Neural Networks) in estimating equation (2) with zero trade flows. Performance is evaluated using RMSE, MAE, and R-square, highlighting differences in model suitability for handling sparse trade data.

3.2. Estimation of Equation (3)

Estimation of equation (3) without zero trade flows

The results of equation (3) using traditional econometric methods are presented in Table 6. The findings are largely consistent with those from equation (2), with positive coefficients for origin and destination size, negative coefficients for distance, and a strong positive border effect. However, R-square values are slightly lower than in equation (2), suggesting that equation (3) introduces additional complexity in explaining trade flows.

Table 6: Traditional Models Performance of equation (3)

Model Performance Comparison				
	OLS	PPML	GPML	NBPML
Origin size	1.52^{***}	1.53^{***}	1.66^{***}	1.47^{***}
	(0.39)	(0.04)	(0.31)	(0.31)
Destination size	2.15^{***}	0.58^{***}	1.59^{***}	1.46^{***}
	(0.38)	(0.02)	(0.31)	(0.30)
Distance	-2.07^{***}	-1.01^{***}	-2.15^{***}	-2.07^{***}
	(0.13)	(0.003)	(0.10)	(0.10)
Border	4.28^{***}	1.88^{***}	4.22^{***}	3.65^{***}
	(0.47)	(0.02)	(0.38)	(0.36)
Constant	-23.52^{***}	-12.94^{***}	-17.59^{***}	-14.51^{***}
	(6.36)	(0.50)	(5.14)	(4.99)
R-square	0.83	0.20	0.53	0.53
MAE	1.05	1032.07	8377.11	4955.34
RMSE	1.47	4271.08	84051.74	49478.33
Observations	846	846	846	846

Note: This table presents estimated coefficients for equation (3) using OLS, PPML, GPML, and NBPML. Origin size and Destination size are approximated by GDP, Distance represents centroid-based distances between trading regions, and Border is a dummy variable taking 1 if the trading regions are in the same country. R-square is computed on test dataset of 218 observations. Coefficients significance: *p<0.1; **p<0.05; ***p<0.01.

Table 7 applies machine learning techniques to equation (3) and compares the performances. As observed earlier (in Table 2), XGBoost performs the best, with the lowest RMSE and highest R-square. Random Forest follows closely behind, while Neural Networks continue to struggle in predictive accuracy. Again, log transformations improve model performance, highlighting their importance in reducing variability and improving fit in machine learning applications to trade models.

Table 7: Machine Learning Models Performance of equation (3)

Model Performance Comparison						
	Variables in Level			Variables in Log		
	Random Forest	XGBoost	Neural Network	Random Forest	XGBoost	Neural Network
RMSE	2,600.13	2,161.52	3,313.21	1.32	1.27	1.33
MAE	656.73	744.23	992.63	0.91	0.88	0.91
R-square	0.71	0.77	0.46	0.86	0.87	0.86

Note: This table presents performance metrics (RMSE, MAE, and R-square) for machine learning models estimating equation (3). Random Forest, XGBoost, and Neural Networks are evaluated using both level and log-transformed variables. Lower RMSE and MAE indicate better predictive accuracy, while higher R-square suggests better model fit.

Table 8 compares the predictive accuracy of traditional and machine learning models for equation (3). Traditional methods, particularly OLS and GPML, have lower predictive performance than ML methods. The best-performing models remain XGBoost and Random Forest with log-transformed variables, reinforcing the effectiveness of tree-based models in capturing trade flow determinants.

Table 8: Traditional vs Machine Learning Models Performance of equation (3)

Model Performance Comparison				
		Performance Metrics		
Estimation Methods	Model Type	RMSE	MAE	R-square
Regression	OLS	1.47	1.05	0.83
	PPML	4271.08	1032.07	0.20
	GPML	84051.74	8377.11	0.53
	NBPML	49478.33	4955.34	0.53
ML (Log transformation)	Random Forest	1.32	0.91	0.86
	XGBoost	1.27	0.88	0.87
	Neural Network	1.33	0.91	0.86

Note: This table compares traditional econometric estimators (OLS, PPML, GPML, and NBPML) with machine learning models (Random Forest, XGBoost, and Neural Networks) for equation (3). Performance is evaluated using RMSE, MAE, and R-square, with results reported for both level and log-transformed variables.

Estimation of equation (3) with zero trade flows

The estimation results for equation (3) when zero trade flows are included is presented in Table 9. PPML and NBPML again show different coefficient magnitudes, with NBPML producing slightly higher R-square values. The overall fit remains lower than in equation (2), suggesting that equation (3) is more sensitive to the presence of zero trade flows.

Table 9: Traditional Models Performance of equation (3) with zeros trade flows

Model Performance Comparison		
	PPML	NBPML
Distance	-0.91^{***}	-2.10^{***}
	(0.003)	(0.10)
Origin size	1.60^{***}	1.58^{***}
	(0.04)	(0.34)
Destination size	0.56^{***}	1.57^{***}
	(0.02)	(0.35)

Border	1.99^{***}	3.80^{***}
	(0.02)	(0.42)
Constant	-14.26^{***}	-16.79^{***}
	(0.52)	(5.72)
R-square	0.18	0.68
MAE	1072.20	5462.12
RMSE	3972.85	30867.10
Observations	872	872

Note: This table presents estimated coefficients for equation (3) using PPML and NBPML when zero trade flows are included. GDP approximates Origin size and Destination size, Distance represents centroid-based regional distances, and Border taking 1 if the trading regions are in the same country. R-square is computed on test dataset of 218 observations. Coefficients significance: *p<0.1; **p<0.05; ***p<0.01.

Table 10 assesses how well machine learning models handle equation (3) in the presence of zero trade flows. Compared to Table 8, all models show lower performance, with increased RMSE values and decreased scores. Random Forest achieves the R-square, though its predictive power is still reduced relative to scenarios without zero trade flows.

Table 10 provides also a final comparison of traditional and machine learning models under equation (3) with zero trade flows. NBPML marginally outperforms PPML, though both methods struggle to fully capture the effects of zero trade flows. Machine learning models again experience a decline in performance, further highlighting the challenges of applying ML methods to sparse trade data.

Table 10: Traditional vs Machine Learning Models Performance of equation (3) with zeros trade flows

Comparison of Estimation Methods				
		Performance Metrics		
Estimation Methods	Model Type	RMSE	MAE	R-square
Regression	PPML	3972.85	1072.20	0.18
	NBPML	30867.10	5462.12	0.68
Machine Learning	Random Forest	2624.04	730.89	0.54
	XGBoost	2169.56	714.91	0.70
	Neural Network	3033.20	869.84	0.43

Note: This table compares traditional econometric models (PPML and NBPML) with machine learning models (Random Forest, XGBoost, and Neural Networks) in estimating equation (3) with zero trade flows. Performance is evaluated using RMSE, MAE, and R-square, highlighting differences in model suitability for handling sparse trade data.

3.3. Synthesis

The comparative results presented in Tables 1 through 12 offer a clear lens through which to assess the central research question: namely, the extent to which machine learning methods can enhance predictive accuracy for gravity models of trade, and how they measure up against more traditional econometric approaches—especially when zero trade flows are taken into account. Across both equation specifications (McCallum's and Anderson and Van Wincoop's) and both data conditions (zero flows excluded vs. included), the findings consistently indicate that tree-based ML models (Random Forest and XGBoost) tend to yield higher R-squared values and lower RMSE under most settings without zero flows, thus providing stronger predictive performance than traditional estimators such as OLS, PPML, GPML, and NBPML. Neural Networks generally trail behind tree-based algorithms, highlighting that not all ML methods perform equally in a trade-flow context, especially in smaller or moderately sized datasets.

Yet, when zero trade flows are introduced into the analysis, PPML and NBPML often exhibit advantages in handling the resulting sparsity, sometimes matching or even surpassing ML algorithms in terms of R-squared. This pattern underscores that while ML can excel in capturing nonlinearity and complex interactions, traditional estimators retain an edge in zero-inflated environments, owing to their well-established theoretical properties and capacity to accommodate the discrete nature of trade data. In short, the tables demonstrate that ML can indeed improve gravity model predictions in many situations but that traditional estimators remain valuable for scenarios characterized by numerous zero trade values. These observations answer the research question by highlighting that the choice between ML methods and traditional approaches depends on the research objective—predictive accuracy versus robustness to zeros—and the empirical context in which the gravity model is applied.

4. Conclusion

In conclusion, the comparative analysis of traditional econometric methods and machine learning approaches for estimating gravity models demonstrates both the promise and limitations of each class of estimators. On the one hand, tree-based algorithms such as XGBoost and Random Forest consistently show superior predictive performance in terms of RMSE and R-squared, particularly when zero trade flows are excluded, and the data contain sufficient variation to capture nonlinearities and intricate interactions. These findings support the notion that advanced computational tools can reveal latent patterns in high-dimensional datasets that conventional linear or pseudo-maximum likelihood estimations may overlook. On the other hand, the robustness of PPML and NBPML in handling zero-inflated trade data remains a crucial advantage, underscoring the significance of model choice when zero flows constitute an appreciable portion of the dataset. In addition, traditional econometric estimators retain their appeal for policy-oriented studies that place a premium on coefficient interpretability and well-established theoretical underpinnings.

Taken together, the results suggest that the selection of an estimation strategy depends fundamentally on the research objective. Where predictive accuracy is paramount, and the data are rich enough to allow for flexible modeling of complex relationships, machine learning methods stand out as effective tools. Conversely, when zero trade values are prominent or when interpretative clarity is essential for policy formulation, PPML or NBPML may be more appropriate.

Future work might explore hybrid approaches that integrate the transparency of econometric modeling with the predictive strengths of machine learning, as well as investigate ways to adapt neural networks or other nonlinear frameworks to the zero-inflated nature of trade data. As global trade patterns become increasingly dynamic, the continued refinement of gravity model estimation methods—and the ability to synthesize advances in both econometrics and computational science—will remain a key priority for scholars and policymakers alike. From a methodological standpoint, new sources of trade-relevant data—ranging from satellite imagery to real-time shipping and supply chain information—continue to expand the dimensionality and complexity of economic datasets. Traditional econometric models, while rigorous and interpretable, may not fully exploit these emerging data streams, nor capture intricate patterns like nonlinearities and higher-order interactions that often characterize trade flows. Hence, integrating ML methods offers a pathway to harness these data more effectively, advancing both the predictive and explanatory power of trade models.

Moreover, the global trade landscape has grown increasingly multifaceted, influenced by shifting geopolitical alliances, evolving trade agreements, and unpredictable exogenous shocks. In this environment, forecasting precision becomes central to informing trade policy, such as tariff setting, multilateral negotiations, and the design of trade facilitation measures. Machine learning methods that excel in out-of-sample forecasting can substantially strengthen policy frameworks, provided that challenges associated with zero trade flows and interpretability are managed. By refining hybrid approaches that merge the clarity of established econometric structures with the adaptive capabilities of ML algorithms, future work stands to deliver both methodological innovations and practical tools for policymakers.

Such advancements hold promise for addressing longstanding debates in the trade literature. For instance, while extensive research has examined the role of distance, shared borders, and GDP in shaping bilateral trade, questions remain regarding how specific or transient factors—

like commodity-level specialization, supply chain disruptions, or trade policy shocks—manifest in disaggregated data. Machine learning can illuminate these nuanced effects by discovering patterns that elude conventional estimation. Consequently, continued investigation of ML-based gravity models can deepen our theoretical understanding of how trade determinants operate and provide more robust policy guidance in an era marked by heightened uncertainty and rapid market shifts.

References

Akhtar, N., & Ghani, E. (2010). Regional Integration in South Asia: An Analysis of Trade Flows Using the Gravity Model. *The Pakistan Development Review*, 105–118. https://doi.org/10.30541/v49i2pp.105-118

Akhvlediani, T., & Śledziewska, K. (2017). The Impacts of Common Commercial Policy on Export Performances of Visegrad Countries. *Prague Economic Papers*, *2017*(1), 3–18.

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, *3*(Volume 3, 2011), 133–160. https://doi.org/10.1146/annurev-economics-111809-125114

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, *93*(1), 170–192.

Beenstock, M., Felsenstein, D., & Rubin, Z. (2015). Immigration to the European Union from its neighborhoods: Testing welfare-chasing and related hypotheses by spatial gravity. *International Journal of Manpower*, *36*(4), 491–517. https://doi.org/10.1108/IJM-01-2014-0010

Breinlich, H., Novy, D., & Santos Silva, J. M. C. (2024). Trade, Gravity, and Aggregation. *Review of Economics and Statistics*, *106*(5), 1418–1426.

D'Ambrosio, A., & Montresor, S. (2022). The pro-export effect of subnational migration networks: New evidence from Spanish provinces. *Review of World Economics*, *158*(1), 53–107. https://doi.org/10.1007/s10290-021-00423-4

Elshehawy, M. A., Shen, H., & Ahmed, R. A. (2014). The Factors Affecting Egypt's Exports: Evidence from the Gravity Model Analysis. *Open Journal of Social Sciences*, *02*(11), Article 11. https://doi.org/10.4236/jss.2014.211020

Esmaeili, A., & Pourebrahim, F. (2011). Assessing Trade Potential in Agricultural Sector of Iran: Application of Gravity Model. *Journal of Food Products Marketing*, *17*(5), 459–469. https://doi.org/10.1080/10454446.2011.583534

Ghazalian, P. L. (2019). Canada's beef exports: Border effects and prospects for market access. *Canadian Journal of Agricultural Economics/Revue Canadienne d'agroeconomie*, *67*(1), 53–74. https://doi.org/10.1111/cjag.12173

Gopinath, M., Batarseh, F. A., Beckman, J., Kulkarni, A., & Jeong, S. (2021). International agricultural trade forecasting using machine learning. *Data & Policy*, *3*, e1.

Gregori, W. D., & Nardo, M. (2021). The effect of restrictive measures on cross-border investment in the European Union. *The World Economy*, *44*(7), 1914–1943. https://doi.org/10.1111/twec.13104

Karemera, D., Reinstra-Munnicha, P., & Onyeocha, J. (2009). Impacts of Free Trade Agreement on US State Vegetable and Fruit Trade Flows. *Journal of Economic Integration*, *24*(1), 116–134.

Mangalathu, S., & Jeon, J. (2019). Stripe-based fragility analysis of multispan concrete bridge classes using machine learning techniques. *Earthquake Engineering & Structural Dynamics*, *48*(11), 1238–1255. https://doi.org/10.1002/eqe.3183

Masudur Rahman, M., & Kim, C. (2012). Trade and investment potential among BCIM countries: Prospects for a dynamic growth quadrangle. *Journal of International Trade Law and Policy*, *11*(2), 163–190. https://doi.org/10.1108/14770021211239677

McCallum, J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *The American Economic Review*, *85*(3), 615–623. JSTOR.

Muganyi, T., & Chen, H. (2016). Strategic Economic Partnerships, Exchange Rate Policy and Agricultural Trade: A Gravity Model Analysis of China’s Agricultural Trade Flows. *Open Journal of Social Sciences*, *4*(5), Article 5. https://doi.org/10.4236/jss.2016.45008

Park, W., Kwon, O.-J., Chun, H.-W., & Kwon, L.-N. (2024). *Enhancing Trade Prediction of Environmental Goods: Integrating Traditional Gravity Models with Machine Learning*. https://www.researchsquare.com/article/rs-4742001/latest

Poyhonen, P. (1963). Toward a general theory of international trade. *Ekonomiska Samfundets Tidskrift*, *16*(2), 69–78.

Santos Silva, J. M. C., & Tenreyro, S. (2022). The Log of Gravity at 15. *Portuguese Economic Journal*, *21*(3), 423–437. https://doi.org/10.1007/s10258-021-00203-w

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

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

Simini, F., Barlacchi, G., Luca, M., & Pappalardo, L. (2022). *Deep Gravity: Enhancing mobility flows generation with deep neural networks and geographic information* (arXiv:2012.00489). arXiv. https://doi.org/10.48550/arXiv.2012.00489

Tinbergen, J. (1962). *Shaping the World Economy (New York: The Twentieth Century Fund)*.