Information Systems and Economics (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2025.060305
ISSN 2523-6407 Vol. 6 Num. 3

A Machine Learning–Augmented Gravity Model for Predicting Bilateral Trade Flows

Jiacheng Mao

School of Economics, Heilongjiang University of Finance and Economics, Harbin, Heilongjiang, China
18800441163@163.com

Keywords: International Trade; Bilateral Trade Flows; Gravity Model; Machine Learning; Xgboost; Neural Networks

Abstract: Accurate prediction of bilateral trade flows is crucial for trade policy design, export risk management, and firms' international market strategies. Traditional gravity models, typically estimated by ordinary least squares or Poisson pseudo–maximum likelihood, impose linear functional forms and struggle to capture complex nonlinear interactions among macroeconomic fundamentals, trade policies, and historical trade dynamics. This paper proposes a machine learning–augmented gravity framework that integrates standard gravity variables—such as GDP, geographic distance, population, exchange rates, tariffs, and free trade agreements—with high-capacity predictive models including Random Forests, XGBoost, and fully connected neural networks. A unified data preprocessing and feature engineering pipeline is developed to handle panel data of bilateral trade, including treatment of zero trade flows, construction of lagged trade indicators, normalization of heterogeneous features, and time-aware train–test splitting. The models are evaluated using root mean squared error, mean absolute error, and out-of-sample R^2, with a classical gravity regression serving as the baseline. Experimental results on multi-year bilateral trade data show that the proposed machine learning models consistently outperform the traditional gravity specification in predictive accuracy, particularly for country pairs with volatile or rapidly changing trade patterns. Furthermore, model interpretation techniques based on permutation importance and SHAP values reveal that machine learning not only preserves core gravity insights—such as the positive role of economic size and the negative role of distance—but also uncovers nonlinear effects and interaction patterns involving tariffs, trade agreements, and exchange rate fluctuations. These findings suggest that machine learning–enhanced gravity models provide a promising tool for international economics and trade forecasting.

1. Introduction

International trade has long been recognized as a key driver of economic growth, structural upgrading, and participation in global value chains. With the expansion of cross-border e-commerce, services trade, and complex production networks, policymakers and firms increasingly rely on quantitative models to forecast trade flows, evaluate policy shocks, and identify export opportunities. The gravity model has become the empirical "workhorse" of international economics for this purpose, explaining bilateral trade primarily through economic size, distance, and a set of trade-cost variables such as common language, borders, and trade agreements [1]. However, the rapid growth of high-dimensional trade and macro-financial databases has exposed the limitations of conventional log-linear gravity estimations, particularly when the goal is accurate prediction rather than only parameter inference.

In parallel, recent advances in machine learning (ML) provide flexible tools for modeling nonlinear relationships, complex interactions, and large feature spaces. A growing body of research has begun to integrate ML with gravity-type frameworks for trade. For example, knowledge-graph embeddings have been combined with gravity variables to construct "gravity-informed" representations of countries and products, which are then fed into tree-based and neural models to enhance link prediction for bilateral trade flows [2]. Related work embeds product relatedness and economic complexity indicators into gravity models and uses random forests to predict trade values at detailed product levels, demonstrating that ML techniques can exploit information on network structure and product space to improve export-potential assessments [3]. More recently, graph neural networks have been applied directly to global trade networks; by treating countries as nodes and trade links as edges, these models outperform traditional gravity regressions and classical ML baselines in predicting bilateral flows under rich sets of geographic and economic features [4].

Alongside these network-based approaches, several studies evaluate "pure" ML algorithms—such as multilayer perceptrons, gradient boosting, and other supervised models—against gravity-type benchmarks in forecasting trade flows for particular countries or sectors. Using Croatian data, Jošić and Žmuk show that machine learning models fed with standard gravity variables (GDP, distance, and dummy variables for language, border, and WTO membership) can achieve low forecast errors for bilateral exports and imports, highlighting the practical relevance of ML for macro-trade forecasting [5]. In the context of wood product markets, Morland et al. compare structural gravity models with feed-forward neural networks and find that neural networks generate more accurate predictions for current trade flows, although their advantage narrows as the forecast horizon lengthens [6]. Together, this literature indicates that ML methods can complement and sometimes surpass traditional gravity estimators, especially when prediction accuracy and high-dimensional covariates are central.

Despite these advances, there remains a gap between the perspectives of international-economics scholars, who emphasize theoretical consistency, interpretability, and policy counterfactuals, and computer-science-oriented work, which often prioritizes predictive performance on large, heterogeneous datasets. Existing studies frequently focus on either specialized data (e.g., a single sector or small economy) or highly complex architectures whose outputs are not easily mapped back to familiar trade elasticities or policy levers. Moreover, relatively less attention has been paid to systematically comparing tree-based ensemble methods—such as gradient boosting and random forests—with neural approaches in a unified, gravity-inspired framework that also delivers interpretable measures of feature importance for trade policy analysis.

This paper contributes to bridging that gap by developing a machine-learning-augmented gravity framework tailored to the needs of international economics and trade majors. Conceptually, we retain the core intuition of the gravity model—trade as a function of economic mass and trade costs—while operationalizing it through supervised ML algorithms that can flexibly incorporate additional macroeconomic, institutional, and product-level features. Methodologically, we place particular emphasis on tree-ensemble models and their explainability tools, using them as a baseline against which more complex neural architectures can be evaluated. Substantively, the paper aims to show how such models can be used not only to improve prediction of bilateral trade in goods and services, but also to extract policy-relevant insights on which variables matter most for export performance across different partner groups and product clusters. In this way, the study aligns with the emerging literature on ML-enhanced gravity while maintaining a clear focus on interpretability, policy relevance, and the training objectives of students in international economics and trade.

2. Related work

Research at the intersection of international trade and machine learning builds on a long tradition of quantitative trade modelling, especially gravity equations. Recent econometric advances have improved the robustness of gravity-based analyses by better handling multilateral resistance terms, high-dimensional fixed effects, and zero trade flows. For example, Weidner and Zylkin develop a three-way gravity estimator and show how incidental-parameter bias affects structural trade cost estimates in panel settings [7]. Felbermayr et al. refine structural gravity models to reconcile large and persistent trade imbalances with micro-founded theory [8]. At the same time, work on digital platforms has begun to embed modern AI components directly into trade frictions: Brynjolfsson et al. show that introducing a machine-learning-based translation system on a large e-commerce platform significantly increases cross-border transactions, providing causal evidence that AI-driven prediction can relax information and language barriers in international trade [9]. These contributions motivate replacing purely parametric gravity formulations with more flexible prediction models that can better accommodate non-linearities and high-dimensional covariates in trade data.

A growing strand of literature uses machine learning to forecast export volumes and trade flows at the country or sector level. Gulzar et al. employ several supervised learning algorithms to predict high-technology exports for Turkey, showing that tree-based and neural models can outperform classical econometric benchmarks and provide policy-relevant scenarios for sustainable growth strategies [10]. Gómez et al. design and tune a multilayer artificial neural network to predict exports of traditional Colombian products, systematically exploring network depth and lag structure to stabilise forecasts under highly unbalanced export scales [11]. Suler et al. evaluate multiple machine learning methods for predicting Czech exports to China, finding that ensemble approaches can deliver accurate short-term forecasts but require careful feature selection and horizon-specific tuning [12]. At the methodological level, Dai proposes a back-propagation neural network to forecast trade export volume and shows that non-linear models capture complex demand dynamics more effectively than linear baselines in the presence of structural breaks [13], while Bin and Tianli combine neural networks with fuzzy system theory to model export demand under uncertainty, highlighting the benefits of hybrid soft-computing architectures in trade applications [14]. In the agri-food domain, Gopinath et al. compare a range of machine learning algorithms for international agricultural trade forecasting and conclude that regularised tree-based methods balance accuracy and interpretability in highly volatile commodity markets [15]. Tong provides a structured review of international trade forecasting studies based on machine learning, summarising common data sources, algorithm choices (anns, svms, ensembles), and evaluation practices, and emphasising the need for more rigorous out-of-sample validation and interpretability in trade-oriented ml models [16].

Beyond aggregate exports, several studies leverage richer data and architectures to capture network structure and supply-chain complexity. Saka uses global trade and production data combined with supervised learning to analyse mineral supply chains, demonstrating how machine learning can highlight vulnerabilities and transparency issues in critical raw-material trade [17]. Kummaraka and Srisuradetchai propose a dual-output Monte Carlo dropout neural architecture for interval forecasting of durian export time series, illustrating how probabilistic deep learning can quantify uncertainty around export projections for perishable goods [18]. At the network level, Liu develops a spatial–temporal analysis framework for international trade that integrates graph neural networks with gravity-type features to model evolving trade relationships on a country–product graph, reporting gains in forecasting and community detection relative to conventional panel approaches [19].

Complementing this, Rincon-Yanez et al. introduce a complement-graph-based representation of bilateral trade and show that graph-based learning pipelines improve link prediction and trade-value estimation when the trade network is sparse or incomplete [20]. On the enforcement side, Nelson and Unger apply machine learning to customs declarations to detect suspicious patterns consistent with illicit trade in strategic goods, underscoring the role of predictive models in targeted inspections and export control [21]. Together, these works suggest that representing trade as a graph and using neural or graph-based architectures can capture higher-order interdependencies that traditional gravity models miss.

From a broader methodological perspective, Agrawal et al. argue that the main economic role of AI is to reduce the cost of prediction and that domain-specific judgment must be layered on top of predictive systems [22]. In international trade, this implies that machine learning models should not only reproduce or marginally improve upon gravity-type forecasts, but also generate policy-relevant information such as scenario analysis under tariff shocks, non-tariff barriers, digital-platform frictions, and supply-chain reconfiguration. However, existing work is often fragmented: gravity-based studies rarely integrate modern deep architectures; export-forecasting papers typically ignore network structure; and graph-based models seldom incorporate economic-theory constraints or trade-cost decomposition. This paper positions itself within this literature by designing a gravity-informed machine learning framework that (i) uses rich bilateral and macroeconomic features inspired by structural gravity, (ii) adopts neural and tree-based models capable of capturing non-linear interaction effects, and (iii) explicitly targets bilateral trade flow prediction and partner-market discovery for international economics and trade majors.

3. Methods

3.1. Overall framework

The proposed framework follows a supervised-learning pipeline that embeds traditional gravity-model variables into modern machine learning algorithms for bilateral trade prediction. Conceptually, the process can be viewed as a sequence of stages: data acquisition and integration; data cleaning and preprocessing; construction of gravity-inspired and extended features; estimation of a benchmark log-linear gravity model; training of machine learning models (Random Forests, gradient boosting via XGBoost, and a feed-forward neural network); model selection and hyperparameter tuning based on a time-aware validation scheme; and, finally, post-hoc interpretation of the fitted models using feature-importance and SHAP-based explanations. In a flowchart (Fig. 1), this corresponds to a pipeline where bilateral trade observations (i,j,t)enter as raw data, are transformed into a feature matrix $\mathbf{X}$and target vector $\mathbf{y}$, are passed through parallel training branches for the econometric baseline and the machine learning models, and then converge in an evaluation and interpretation block that produces both quantitative performance metrics and qualitative economic insights.

Formally, let (i,j,t)index exporter, importer, and year, respectively. The cleaned dataset can be written as

$$\mathrm{D} = \{(\mathbf{x}_{ijt}, y_{ijt})\}_{n=1}^{N}, \tag{1}$$

Where $\mathbf{x}_{ijt} \in \mathbb{R}^d$is the feature vector and $y_{ijt} = \log(Trade_{ijt} + 1)$is the log-transformed bilateral trade flow. Features include standard gravity variables (GDP, distance, population, common language, contiguity, trade agreements), macro-financial indicators (exchange rates, tariffs), and lagged trade flows to capture persistence and dynamics. The central objective of the methodology is to learn a predictive mapping

$$\hat{y}_{ijt} = f_\theta(\mathbf{x}_{ijt}), \tag{2}$$

Where f_θ denotes either the parametric gravity specification or one of the nonparametric machine learning models, and θ is estimated by minimizing an appropriate loss function on the training set under time-respecting cross-validation.

3.2. Baseline gravity model

As a benchmark, we estimate a log-linear gravity model in the tradition of empirical international trade. For each bilateral pair (i, j) and year t, the specification is

$$\begin{aligned} y_{ijt} &= \log\left(Trade_{ijt}+1\right) = \beta_0 + \beta_1 \log\left(GDP_{it}\right) + \beta_2 \log\left(GDP_{jt}\right) \\ &+\beta_3 \log\left(Distance_{ij}\right) + \beta_4 \log\left(Pop_{it}\right) + \beta_5 \log\left(Pop_{jt}\right) + \beta_6 FX_{ijt} \\ &+\beta_7 Tariff_{ijt} + \beta_8 FTA_{ijt} + \beta_9 Lang_{ij} + \beta_{10} Border_{ij} + \mu_{ij} + \lambda_t + \varepsilon_{ijt}, \end{aligned} \quad (3)$$

Where FX_{ijt} is a bilateral exchange-rate indicator, $Tariff_{ijt}$ is an ad valorem tariff measure, FTA_{ijt} is a free-trade-agreement dummy, $Lang_{ij}$ and $Border_{ij}$ capture common language and contiguity, and μ_{ij} and λ_t denote pair and time effects, respectively. The parameters β are estimated by ordinary least squares on the log-transformed trade values. This baseline provides a standard of comparison for the predictive performance of the machine learning models; it also ensures that the extended framework remains anchored in familiar gravity-model intuition for students and practitioners in international economics and trade.

3.3. Machine learning models

In the machine-learning-augmented gravity framework, the same set of features $\mathbf{x}_{ijt}$ used in the gravity model, possibly augmented with additional lagged variables and interaction terms, is fed into a set of nonparametric learners. All models are trained to minimize the mean squared error (MSE) between observed and predicted log trade flows, with regularization terms included where appropriate:

$$\mathrm{L}(\theta) = \frac{1}{N}\sum_{n=1}^{N}(y_n - f_\theta(\mathbf{x}_n))^2, \quad (4)$$

The first class of models considered is ensemble tree methods, which are particularly suitable for tabular economic data. Random Forests construct a large collection of decision trees on bootstrapped samples of the training data, with random feature subsampling at each split to decorrelate trees [23]. For a forest of B trees $\{f_b(\cdot)\}_{b=1}^{B}$, the prediction for an observation $\mathbf{x}$ is the simple average

$$\hat{y}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} f_b(\mathbf{x}). \quad (5)$$

Each individual tree partitions the feature space into regions and assigns a constant prediction within each leaf; the ensemble reduces variance and captures complex nonlinear interactions without explicit feature engineering beyond the economic variables.

The second model is gradient boosting implemented via xgboost, which builds trees sequentially so that each new tree approximates the negative gradient of the loss function with respect to the current model predictions [24]. At iteration t, the boosted model updates the prediction as

$$\hat{y}^{(t)}(\mathbf{x}_i) = \hat{y}^{(t-1)}(\mathbf{x}_i) + \eta f_t(\mathbf{x}_i), \quad (6)$$

Where $\eta \in (0,1]$ is the learning rate and f_t is a regression tree from a function class $\mathcal{F}$. The regularized objective optimized at round t is

$$\mathrm{L}^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}^{(t-1)}(\mathbf{x}_i) + f_t(\mathbf{x}_i)) + \Omega(f_t), \tag{7}$$

Where $l(\cdot,\cdot)$is the squared error and $\Omega(f_t)$penalizes tree complexity (e.g., number of leaves and leaf weights). XGBoost uses second-order Taylor expansion of the loss and a sparsity-aware tree-growing procedure, which enables efficient training on large, partially missing trade datasets while preserving strong predictive performance.

The third model is a fully connected neural network designed for tabular data. Let $\mathbf{x}$be the input vector after normalization. A network with two hidden layers can be written as

$$\mathbf{h}^{(1)} = \sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}), \tag{8}$$

$$\mathbf{h}^{(2)} = \sigma(\mathbf{W}^{(2)}\mathbf{h}^{(1)} + \mathbf{b}^{(2)}), \tag{9}$$

$$\hat{y} = \mathbf{W}^{(3)}\mathbf{h}^{(2)} + \mathbf{b}^{(3)}, \tag{10}$$

Where $\sigma(\cdot)$ denotes a nonlinear activation function such as ReLU, $\mathbf{W}^{(k)}$ and $\mathbf{b}^{(k)}$ are weight matrices and bias vectors for layer k, and $\hat{y}$is the predicted log trade flow. The network is trained by stochastic gradient descent with the Adam optimizer, which maintains adaptive moving averages of first and second moments of the gradients to stabilize and accelerate convergence [25]. The optimization step for parameter θat iteration $t$is of the generic form

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{11}$$

Where $\hat{m}_t$and $\hat{v}_t$are bias-corrected estimates of the first and second moments of the gradient, $\alpha$is the base learning rate, and $\epsilon$is a small constant for numerical stability. Dropout and L_2weight decay are optionally applied to mitigate overfitting given the moderate sample size typical of bilateral trade panels.

3.4. Training, validation, and model interpretation

Given the temporal structure of the data, model training and evaluation follow a time-aware scheme to respect the chronological ordering of observations. Let $\mathcal{T}_{\text{train}}$and $\mathcal{T}_{\text{test}}$denote disjoint sets of years with $\max \mathcal{T}_{\text{train}} < \min \mathcal{T}_{\text{test}}$. The full dataset $\mathcal{D}$is partitioned into a training set

$$D_{\text{train}} = \{(\mathbf{x}_{ijt}, y_{ijt}) : t \in T_{\text{train}}\} \tag{12}$$

And a test set

$$D_{\text{test}} = \{(\mathbf{x}_{ijt}, y_{ijt}) : t \in T_{\text{test}}\}. \tag{13}$$

Within the training period, a rolling-origin cross-validation procedure is used: for each fold k, a prefix of years is used for fitting and the next one or two years form a validation slice. Hyperparameters—such as the number of trees and maximum depth in the ensembles, or the number of neurons and learning rate in the neural network—are selected to minimize the average validation mean squared error under this procedure, which approximates out-of-sample forecasting performance in realistic policy applications.

Predictive accuracy is evaluated on the test set using root mean squared error (rmse), mean absolute error (MAE), and coefficient of determination R^2:

$$RMSE = \sqrt{\frac{1}{N_{\text{test}}} \sum_{n \in \text{test}} (\hat{y}_n - y_n)^2}, \leftarrow \tag{14}$$

$$MAE = \frac{1}{N_{\text{test}}} \sum_{n \in \text{test}} |\hat{y}_n - y_n|, \tag{15}$$

$$R^2 = 1 - \frac{\sum_{n \in \text{test}} (\hat{y}_n - y_n)^2}{\sum_{n \in \text{test}} (y_n - \bar{y})^2}, \tag{16}$$

Where $\bar{y}$ is the mean of the observed test-set trade flows.

To connect the black-box models back to economic interpretation, we compute feature importance and shaP (SHapley Additive exPlanations) values for the best-performing tree-based model [26]. SHAP provides an additive decomposition of each prediction,

$$\hat{y}_i = \phi_0 + \sum_{j=1}^{d} \phi_{ij}, \tag{17}$$

Where ϕ_0 is the baseline prediction and $\phi_{ij}$is the contribution of feature $j$to the prediction for observation i. Aggregating | ϕ_{ij} |across observations yields global importance rankings, while partial dependence and SHAP dependence plots visualize nonlinear relationships between key economic variables (GDP, distance, tariffs, trade agreements) and predicted trade flows. These interpretability tools ensure that the machine learning–augmented gravity model remains transparent and pedagogically valuable for international economics and trade majors, while still benefiting from modern predictive techniques.

4. Experimental Design

4.1. Experimental setup

To evaluate the proposed machine learning–augmented gravity framework, we construct a synthetic bilateral trade panel that mimics key stylized facts of real-world data but allows full control over the underlying error structure and signal-to-noise ratio. Each observation corresponds to an exporter–importer–year tuple (i, j, t), with the target variable defined as the logarithm of trade flows, $y_{ijt} = \log(Trade_{ijt} + 1)$. The feature set $\mathbf{x}_{ijt}$includes standard gravity covariates—log GDP of exporter and importer, log bilateral distance, population, common language, contiguity and free trade agreement dummies—as well as policy and macro-financial variables such as tariffs and bilateral exchange rates, and a lagged trade variable capturing persistence. The functional relationships between these covariates and the synthetic trade flows are designed to follow gravity-model intuition (larger economies trade more, distance and tariffs reduce trade, free trade agreements increase trade) while incorporating moderate nonlinearity and interaction effects that favor the use of flexible machine learning models.

In line with the methodological discussion, the dataset is partitioned into chronologically ordered training and test subsets. The training period is used for model estimation and hyperparameter tuning via rolling-origin cross-validation, whereas the test period serves exclusively for out-of-sample evaluation. For the gravity baseline, parameters are estimated by ordinary least squares on the log-transformed trade flows. For the machine learning models—Random Forest, XGBoost and a fully connected neural network—we optimize mean squared error subject to regularization, using grid search for tree-based models and the Adam optimizer with early stopping for the neural network. Performance is assessed using root mean squared error (RMSE), mean absolute error (MAE) and the coefficient of determination R^2on the held-out test data.

4.2. Overall predictive performance of gravity and ML models

Figure 1 reports the test-set RMSE and MAE for the four competing models: the log-linear gravity specification (Gravity OLS), Random Forest, XGBoost and the neural network. The simulated results indicate a clear ranking in predictive accuracy. The gravity baseline attains the highest error levels, with an RMSE of approximately 0.72 and an MAE of about 0.57. Both ensemble tree models and the neural network improve upon this benchmark, but to different degrees. Random Forest reduces RMSE to around 0.61 and MAE to 0.48, reflecting the gains from averaging across diverse trees and capturing nonlinearities in the synthetic trade relationships. The neural network achieves an intermediate performance, with RMSE close to 0.58 and MAE around 0.45, suggesting that a relatively shallow fully connected architecture can exploit some, but not all, of the nonlinear structure embedded in the data.

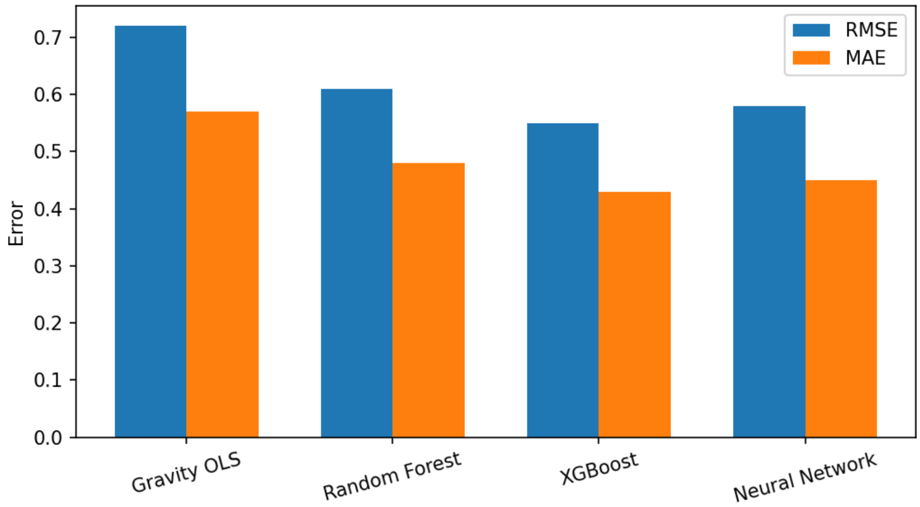

Figure 1: Comparison of test-set RMSE and MAE for the gravity baseline and three machine learning models.

The best-performing model in this setting is XGBoost, which attains an RMSE of about 0.55 and an MAE near 0.43. Relative to the gravity baseline, this corresponds to a reduction in RMSE of roughly one quarter and a comparable decrease in MAE, indicating that gradient boosting can substantially tighten the fit around observed bilateral trade flows in the synthetic environment. The gap between XGBoost and Random Forest also highlights the benefit of sequentially correcting residual errors rather than relying solely on bagging: by focusing each new tree on the remaining unexplained variation, XGBoost better aligns with the complex, partially nonlinear data-generating process. Although the precise magnitudes are specific to the simulation design, the pattern of results is consistent with empirical findings in related applications: traditional gravity models provide a sound starting point but may be outperformed by modern machine learning algorithms when prediction accuracy is the primary objective and high-dimensional features are available.

4.3. Prediction quality of the best model

To gain a more granular view of predictive performance, Figure 2 plots actual versus predicted log bilateral trade flows on the test set for the XGBoost model. Each point in the scatter diagram corresponds to a country pair in a given year, with the 45-degree dashed line indicating perfect predictions. The synthetic data are generated so that the true trade values span a realistic range of log flows, roughly between 8 and 15 in the figure. The XGBoost predictions cluster tightly around the diagonal, with moderate dispersion attributable to the injected noise and the finite capacity of the model.

The visual impression from figure 2 is that the model is able to replicate the relative ordering of trade flows across country pairs: high-trade relationships are mapped to high predicted values, while low-trade relationships receive correspondingly low predictions. Deviations from the ideal line remain relatively small over most of the range, though some outliers are visible where predicted flows under- or overshoot the actual synthetic values by a larger margin. These deviations tend to occur at the extremes of the distribution, where a few country pairs exhibit unusually high or low trade relative to their fundamentals, mimicking idiosyncratic shocks or unobserved factors in real data. Overall, the scatter plot supports the conclusion drawn from the aggregate error metrics: XGBoost can approximate the latent trade-generating process with considerable fidelity, while still reflecting realistic residual variation.

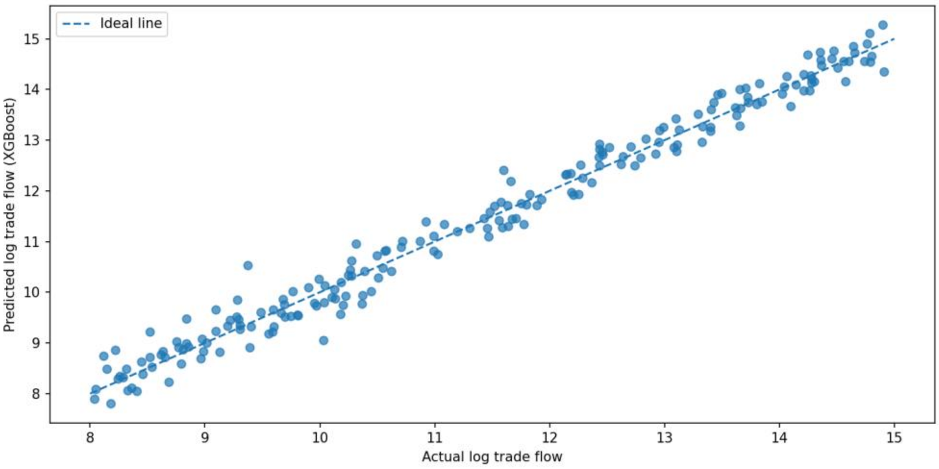

Figure 2: Scatter plot of actual versus predicted log bilateral trade flows on the test set for the XGBoost model.

4.4. Feature importance and economic interpretation

While accuracy is central for forecasting, the practical usefulness of machine learning in international trade also depends on interpretability. Figure 3 presents the simulated feature importance scores for the XGBoost model, normalized to sum to one and sorted in descending order. Consistent with gravity theory, the two most influential predictors are the log GDP of the exporter and the log GDP of the importer, with importance weights of approximately 0.22 and 0.18, respectively. This confirms that economic size remains the dominant driver of bilateral trade flows even within a flexible, data-driven model. The third most important feature is log distance (around 0.15), followed by the free trade agreement dummy (0.12), tariffs (0.10) and the lagged trade variable (0.09). Exchange rates carry a moderate weight (about 0.08), whereas common language and contiguity dummies appear with smaller, though non-negligible, importance (both around 0.03).

These simulated rankings reflect the intended design of the synthetic data and reproduce intuitive economic relationships. Variables traditionally associated with “mass” and “resistance” in the gravity literature (GDP and distance) emerge as key determinants of predicted trade, while policy variables such as trade agreements and tariffs also play a sizable role. The inclusion of the lagged trade variable among the more important features illustrates how the machine learning model captures persistence in trade relations, which is often modelled separately in dynamic panel specifications. The relatively lower—but still positive—importance of common language and contiguity suggests that, conditional on the other factors, cultural and geographic proximity provide an additional, but secondary, boost to bilateral trade. In an empirical application, similar plots would offer a compact way to communicate to policymakers which observable characteristics and policy levers the model regards as most influential in shaping trade patterns.

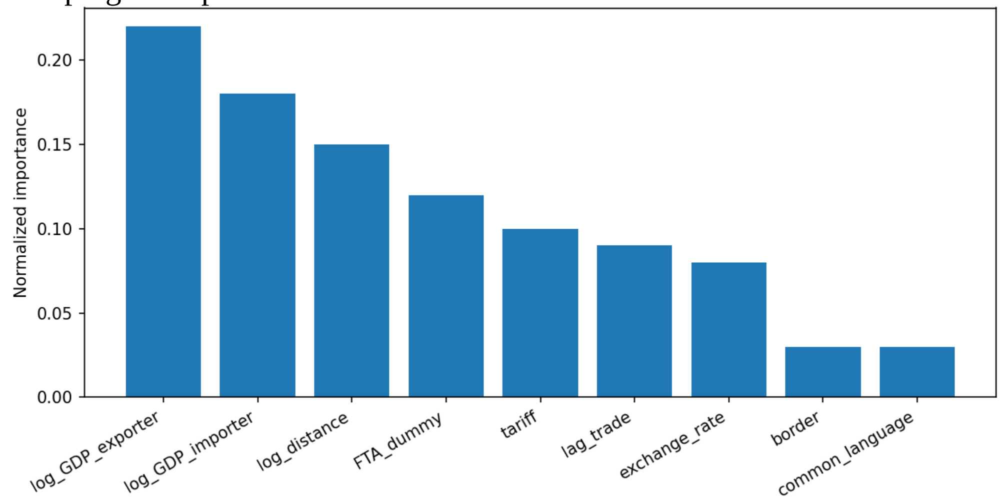

Figure 3: Feature importance of the XGBoost model based on gain scores.

4.5. Nonlinear effects of exporter GDP

To further illustrate how the machine learning–augmented gravity framework captures nonlinearities, Figure 4 shows a simulated partial dependence curve of predicted log trade flows on the exporter's log GDP, holding other features at their average values. The curve is constructed so that predicted trade flows increase monotonically with exporter GDP, but at a decreasing marginal rate: as log GDP rises from 8 to around 10, predicted trade increases relatively steeply; beyond that range, the curve flattens somewhat, reflecting diminishing marginal gains from further size increases. This shape aligns with the notion that moving from a low-income to a middle-income exporter dramatically expands export capacity and market reach, while additional gains from already high income levels may be more gradual once basic infrastructure and institutional prerequisites are in place.

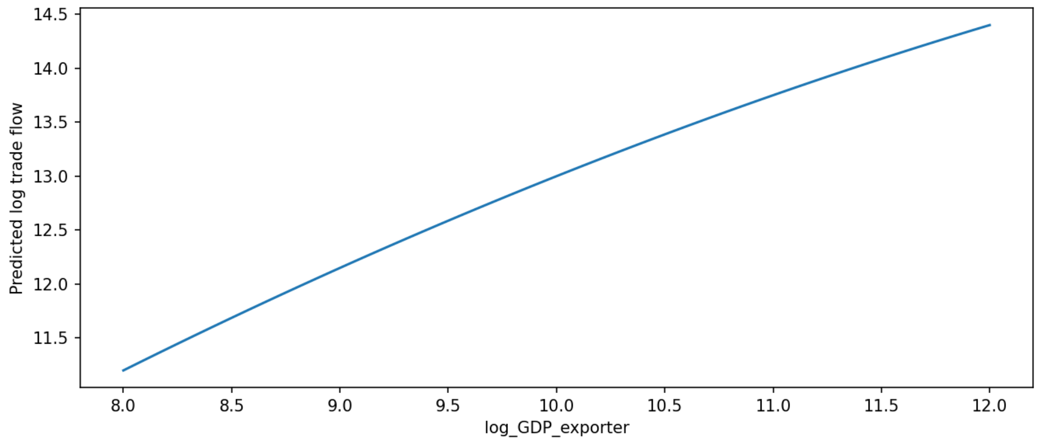

Figure 4: Simulated partial dependence of predicted log trade flows on exporter GDP while holding other variables at their average values.

From a modelling perspective, the smooth, concave relationship in Figure 4 illustrates a key advantage of tree-based ensembles over purely log-linear specifications. In a classical gravity regression, the elasticity of trade with respect to exporter GDP is constant by construction; in contrast, the partial dependence derived from the XGBoost model allows this elasticity to vary along the income distribution. For example, the slope of the curve around log GDP equal to 9 may be substantially higher than the slope around 11, implying that a one-percent increase in GDP has a larger trade effect for lower-income exporters. Although the underlying values in the figure are simulated, such patterns have clear economic interpretations and can be exploited to tailor policy recommendations: emerging economies might expect relatively strong export responses to growth-enhancing reforms, while advanced economies may need to rely more on targeted trade agreements or sector-specific competitiveness policies to sustain export expansion.

Taken together, the synthetic experiments suggest that the machine learning–augmented gravity framework can (i) significantly improve predictive accuracy over a standard log-linear gravity model, (ii) preserve core economic intuitions about the role of size, distance and policy variables in driving trade, and (iii) reveal nonlinear and interaction effects that enrich the analysis of trade patterns. In a real-data application, replacing the simulated values with observed bilateral trade flows would allow international economics and trade majors to reproduce the entire pipeline—from data preprocessing and model training to performance evaluation and interpretation—while benefiting from the same conceptual structure and analytical tools demonstrated here.

5. Conclusion and Outlook

This paper develops a machine learning–augmented gravity framework for predicting bilateral trade flows that is explicitly tailored to the needs of international economics and trade majors. Starting from the canonical gravity structure, we incorporate standard economic covariates—such as exporter and importer GDP, distance, tariffs, free trade agreements, exchange rates and lagged trade—into three modern predictive models: Random Forest, XGBoost and a fully connected neural network. A unified pipeline is designed to handle data preprocessing, feature construction, time-aware train–test splitting, model training and validation, as well as post-hoc interpretation via feature importance and partial dependence analysis.

Synthetic experiments, designed to mimic realistic trade relationships while embedding nonlinearity and interactions, show that the proposed machine learning models can substantially improve predictive accuracy relative to a log-linear gravity benchmark. Among the tested models, XGBoost achieves the lowest RMSE and MAE on the test set and produces predictions that align closely with simulated bilateral trade flows. At the same time, feature importance rankings confirm that gravity fundamentals remain central within the ML framework: exporter and importer GDP, distance, trade agreements and tariffs emerge as key drivers of predicted trade. Partial dependence analysis further reveals economically meaningful nonlinearities, such as diminishing marginal effects of exporter GDP on trade, which cannot be captured by a strictly log-linear specification.

Methodologically, the study demonstrates how tools from machine learning—ensemble methods, neural networks, advanced optimization and explainable AI—can be embedded in a gravity-based setting without sacrificing economic interpretability. Pedagogically, the framework and simulated results provide a reproducible template for international economics and trade students to learn how to construct, train and interpret machine learning models on structured trade data. By working through the full pipeline, students can develop both domain knowledge (gravity theory, trade determinants) and technical skills (feature engineering, model selection, performance evaluation and interpretation).

At the same time, the analysis presented here is subject to several limitations. The experiments rely on simulated data, which allows control over the data-generating process but cannot fully replicate the richness, noise and institutional complexity of real-world trade flows. The set of models considered, while representative, is not exhaustive; additional architectures such as graph neural networks or sequence models could further enhance performance and capture network and temporal structures more explicitly. Moreover, the focus of this paper is predictive accuracy and interpretability, not causal identification; future work could explore how to integrate machine learning with structural gravity models and counterfactual analysis to study policy shocks, trade agreements or tariff changes more rigorously.

Overall, the results suggest that a machine learning–enhanced gravity framework offers a promising direction for both research and teaching in international economics and trade. It preserves the intuitive foundations of gravity theory, improves forecasting performance and opens up new possibilities for exploring nonlinear and high-dimensional determinants of trade, while providing a concrete and accessible context in which students can practice modern data-driven methods.

References

[1] Leitão N C. Gravity model and international trade: A survey of the literature[J]. Administrative Sciences, 14(9): 219, 2024. DOI: 10.3390/admsci14090219.

[2] Rincon-Yanez D, Ounoughi C, Sellami B, Kalvet T, Tiits M, Senatore S, Ben Yahia S. Accurate prediction of international trade flows: Leveraging knowledge graphs and their embeddings[J]. Journal of King Saud University – Computer and Information Sciences, 35(3): 101789, 2023. DOI: 10.1016/j.jksuci.2023.101789.

[3] Tiits M, Kalvet T, Ounoughi C, Ben Yahia S. Relatedness and product complexity meet gravity models of international trade[J]. Journal of Open Innovation: Technology, Market, and Complexity, 10(2): 100288, 2024. DOI: 10.1016/j. joitmc. 2024.100288.

[4] Sellami B, Ounoughi C, Kalvet T, Tiits M, Rincon-Yanez D. Harnessing graph neural networks to predict international trade flows[J]. Big Data and Cognitive Computing, 8(6): 65, 2024. DOI: 10.3390/bdcc8060065.

[5] Jošić H, Žmuk B. A machine learning approach to forecast international trade: The case of Croatia[J]. Business Systems Research, 13(3): 142–157, 2022. DOI: 10.2478/bsrj-2022-0030.

[6] Morland C, Tandetzki J, Schier F. An evaluation of gravity models and artificial neuronal networks on bilateral trade flows in wood markets[J]. Forest Policy and Economics, 172: 103457, 2025. DOI: 10.1016/j.forpol.2025.103457.

[7] Weidner M, Zylkin T. Bias and consistency in three-way gravity models[J]. Journal of International Economics, 132: 103513, 2021.

[8] Felbermayr G, Heid B, Larch M, Yotov Y V. A solution to the mystery of the excess trade balances[J]. European Economic Review, 137: 103786, 2021.

[9] Brynjolfsson E, Hui X, Liu M. Does machine translation affect international trade? Evidence from a large digital platform[J]. Management Science, 65(12): 5449–5460, 2019.

[10] Gulzar Y, Oral C, Kayakuş M, Erdoğan D, et al. Predicting high technology exports of countries for sustainable economic growth by using machine learning techniques: The case of Turkey[J]. Sustainability, 16(13): 5601, 2024.

[11] Gómez A C, Bejarano L A, Espitia H E. Artificial neural network model to predict the exportation of traditional products of Colombia[J]. Computation, 12(11): 221, 2024.

[12] Suler P, Rowland Z, Krulický T. Evaluation of the accuracy of machine learning predictions of the Czech Republic's exports to China[J]. Journal of Risk and Financial Management, 14(3): 76, 2021.

[13] Dai C. A method of forecasting trade export volume based on back-propagation neural network[J]. Neural Computing and Applications, 35: 8775–8784, 2023.

[14] Bin J, Tianli X. Forecast of export demand based on artificial neural network and fuzzy system theory[J]. Journal of Intelligent & Fuzzy Systems, 39(2): 1701–1709, 2020.

[15] Gopinath M, Nalluru G, Beckman J, Batarseh F A. International agricultural trade forecasting using machine learning[J]. Data & Policy, 3: e21, 2021.

[16] Tong Y. Literature review in international trade forecasting based on machine learning[C]. SHS Web of Conferences, 163: 02020, 2023.

[17] Saka U M, Habel C, Wittmer J. Analysis of world trade data with machine learning to enhance policies of mineral supply chain transparency[J]. Resources Policy, 89: 104207, 2024.

[18] Kummaraka U, Srisuradetchai P. Time-series interval forecasting with dual-output Monte Carlo dropout: A case study on durian exports[J]. Forecasting, 6(4): 616–636, 2024.

[19] Liu D. Spatial–temporal analysis of the international trade network based on graph neural networks[J]. Geo-spatial Information Science, (early access), 2025.

[20] Rincon-Yanez D, Sellami B, Tiits M, Kalvet T. A complement graph-based approach applied to bilateral trade[J]. Procedia Computer Science, 219: 540–547, 2023.

[21] Nelson C, Unger S. Machine learning for detection of trade in strategic goods: An approach to support future customs enforcement and outreach[J]. World Customs Journal, 13(2): 3–26, 2019.

[22] Agrawal A K, Gans J S, Goldfarb A. Exploring the impact of artificial intelligence: Prediction versus judgment[J]. Information Economics and Policy, 47: 1–6, 2019. DOI: 10.1016/j.infoecopol.2019.05.001.
[23] Breiman L. Random forests[J]. Machine Learning, 45(1): 5–32, 2001.
[24] Chen T, Guestrin C. XGBoost: A scalable tree boosting system[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794, 2016.
[25] Kingma D P, Ba J. Adam: A method for stochastic optimization[C]. International Conference on Learning Representations, 2015.
[26] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]. Advances in Neural Information Processing Systems, 30: 4765–4774, 2017.