

Team Number:	apmcm25309036
Problem Chosen:	C

---

## 2025 APMCM summary sheet

This study develops an integrated modeling framework to quantify the effects of the United States' 2025 "Reciprocal Tariffs" on global trade flows and regional industrial systems.

For Question 1, a multi-market equilibrium model captures how tariff shocks redistribute soybean exports among the U.S., Brazil, and Argentina; results show that higher U.S. tariffs accelerate demand diversion toward Brazil and Argentina.

For Question 2, a dynamic-panel transmission model incorporates Japan's non-tariff responses—Mexican re-exports and U.S. local production—and finds that tariffs alter the structure of U.S. auto imports more than the total volume, with Mexico and South Korea absorbing Japan's displaced share.

For Question 3, a high/mid/low-end semiconductor model reveals strong structural heterogeneity: high-end chip imports decline most under tariffs and export controls, while the U.S. semiconductor industry faces a trade-off between reduced dependence on China and higher procurement costs with increased supply-chain concentration.

For Question 4, a tariff-revenue forecasting model shows that although short-term revenue increases, medium-term revenue declines as shrinking import bases offset higher rates.

For Question 5, macro-indicator simulations suggest that tariff hikes trigger partner countermeasures, weaken manufacturing reshoring incentives, and increase financial-market volatility. Across all questions, the findings demonstrate that tariffs primarily reshape trade structure and supply-chain routing, rather than sustainably reducing U.S. trade deficits or reviving domestic manufacturing.

**Keywords: Reciprocal Tariffs   Trade Structure Adjustment   Policy Transmission Modeling   Supply-Chain Substitution**

## Contents

<b>1. Introduction . . . . .</b>	<b>1</b>
1.1 Restatement of Problem . . . . .	1
1.1.1 <i>Problem 1</i> . . . . .	1
1.1.2 <i>Problem 2</i> . . . . .	1
1.1.3 <i>Problem 3</i> . . . . .	1
1.1.4 <i>Problem 4</i> . . . . .	1
1.1.5 <i>Problem 5</i> . . . . .	2
1.2 Our Work . . . . .	2
1.2.1 <i>Analysis of Problem 1</i> . . . . .	2
1.2.2 <i>Analysis of Problem 2</i> . . . . .	2
1.2.3 <i>Analysis of Problem 3</i> . . . . .	2
1.2.4 <i>Analysis of Problem 4</i> . . . . .	2
1.2.5 <i>Analysis of Problem 5</i> . . . . .	3
<b>2. Model Hypothesis . . . . .</b>	<b>3</b>
2.1 Assumption 1 . . . . .	3
2.2 Assumption 2 . . . . .	3
2.3 Assumption 3 . . . . .	3
2.4 Assumption 4 . . . . .	3
2.5 Assumption 5 . . . . .	4
<b>3. Notations . . . . .</b>	<b>4</b>
<b>4. Models and Solutions for Problem 1 . . . . .</b>	<b>6</b>
4.1 Current Soybean Trade Pattern and Modeling Strategy . . . . .	6
4.1.1 <i>Long-Term Trade Pattern: A Stable Structure Dominated by Brazil</i> . . . . .	6
4.1.2 <i>Pattern Mutation Under the 2018 Tariff Shock</i> . . . . .	6
4.1.3 <i>Pattern Readjustment After the 2025 Tariff Adjustment</i> . . . . .	7
4.2 Data, Model, and Solution . . . . .	8
4.2.1 <i>Data Sources</i> . . . . .	8
4.2.2 <i>Data Preprocessing</i> . . . . .	8
4.2.3 <i>Unit Unification</i> . . . . .	8
4.2.4 <i>Missing Value Handling</i> . . . . .	8
4.2.5 <i>Outlier Detection and Correction</i> . . . . .	9
4.2.6 <i>Key Tariff Data Completion</i> . . . . .	9
4.2.7 <i>Feature Standardization</i> . . . . .	9
4.2.8 <i>Model Selection Basis</i> . . . . .	10
4.2.9 <i>Model Overview</i> . . . . .	10
4.2.10 <i>Dataset Division</i> . . . . .	10
4.2.11 <i>Training Setup</i> . . . . .	10
4.3 Results and Analysis . . . . .	11
4.3.1 <i>LSTM Model Results and Analysis</i> . . . . .	11
4.3.2 <i>Transformer Model Results and Analysis</i> . . . . .	12

4.3.3 <i>Model Summary</i> . . . . .	12
<b>5. Models and Solutions for Problem 2 . . . . .</b>	<b>12</b>
5.1 Data Collection and Preprocessing . . . . .	12
5.1.1 <i>Data Sources</i> . . . . .	12
5.1.2 <i>Data Scope and Frequency</i> . . . . .	13
5.2 Data Preprocessing . . . . .	15
5.3 Current Position of Japanese Automobiles in the U.S. Market . . . . .	15
5.3.1 <i>Market Share and Trend</i> . . . . .	15
5.3.2 <i>Sales Structure and Competitive Advantages</i> . . . . .	16
5.3.3 <i>Supply Chain Layout (Direct Export + Local Production + Third-Country Manufacturing)</i> . . . . .	16
5.4 Feature Engineering . . . . .	17
5.4.1 <i>Specific Feature Extraction</i> . . . . .	17
5.4.2 <i>Feature Optimization</i> . . . . .	18
5.5 Model Construction and Implementation . . . . .	19
5.5.1 <i>Model Selection Basis</i> . . . . .	19
5.5.2 <i>Traditional Econometric Models</i> . . . . .	20
5.5.3 <i>Machine Learning Model</i> . . . . .	21
5.5.4 <i>Ensemble Learning</i> . . . . .	21
5.5.5 <i>Scenario Simulation Method</i> . . . . .	22
5.6 Results and Analysis . . . . .	22
5.6.1 <i>Model Evaluation and Performance</i> . . . . .	22
5.6.2 <i>Impact of U.S. Tariff Adjustments</i> . . . . .	24
5.6.3 <i>Scenario Simulation Results</i> . . . . .	25
<b>6. Models and Solutions for Problem 3 . . . . .</b>	<b>26</b>
6.1 Current Situation of U.S.-China Semiconductor Trade . . . . .	26
6.1.1 <i>Characteristics of Total Trade Volume Change (2018-2025)</i> . . . . .	26
6.1.2 <i>Characteristics of Trade Structure Differentiation (High/Mid/Low-end Chips)</i> . . . . .	26
6.1.3 <i>Trade Evolution Law Under Policy Shocks</i> . . . . .	27
6.2 Indicator System Construction . . . . .	28
6.2.1 <i>Explained Variables: Trade Characteristic Indicators</i> . . . . .	28
6.2.2 <i>Core Explanatory Variables: Policy Shock Indicators</i> . . . . .	29
6.2.3 <i>Control Variables: Economic and Time-Series Feature Indicators</i> . . . . .	30
6.2.4 <i>Security Assessment Indicators: Semiconductor-Specific Indicators</i> . . . . .	30
6.3 Data Collection . . . . .	30
6.3.1 <i>Data Sources and Scope</i> . . . . .	31
6.3.2 <i>Semiconductor Product Classification Refinement</i> . . . . .	31
6.4 Data Preprocessing . . . . .	31
6.4.1 <i>Missing Value and Outlier Handling</i> . . . . .	31
6.4.2 <i>Experimental Data Image Correlation and Calibration</i> . . . . .	31

6.5 Model Construction and Implementation . . . . .	32
6.5.1 <i>Baseline Econometric Model (ElasticNet)</i> . . . . .	32
6.5.2 <i>Dynamic Response Model (Huber)</i> . . . . .	32
6.5.3 <i>Integrated Model (ElasticNet+Huber+GBDT)</i> . . . . .	33
6.6 Results and Analysis . . . . .	33
6.6.1 <i>Model Performance Comparison</i> . . . . .	33
6.6.2 <i>Feature Importance Analysis</i> . . . . .	33
6.6.3 <i>Quantitative Policy Effects</i> . . . . .	33
6.6.4 <i>Structural Differentiation Results</i> . . . . .	33
6.6.5 <i>Core Conclusions</i> . . . . .	36
<b>7. Models and Solutions for Problem 4 . . . . .</b>	<b>36</b>
7.1 Current Situation of U.S. Tariff Revenue and Import Trade . . . . .	36
7.1.1 <i>Core Characteristics of the Baseline Period (2020-2024)</i> . . . . .	36
7.1.2 <i>Background of the 2025 Policy Shock</i> . . . . .	36
7.1.3 <i>Preliminary Observation of Time-Series Correlation Between Tariff Revenue and Import Value</i> . . . . .	37
7.2 Evaluation Indicators . . . . .	37
7.2.1 <i>Explained Variables: Fiscal Revenue Indicators</i> . . . . .	37
7.2.2 <i>Core Explanatory Variables: Policy Shock and Trade Indicators</i> . . . . .	38
7.2.3 <i>Control Variables: Time-Series and Economic Feature Indicators</i> . . . . .	38
7.3 Data Collection . . . . .	38
7.3.1 <i>Data Sources and Scope (Tabular Form)</i> . . . . .	38
7.3.2 <i>Data Classification</i> . . . . .	38
7.4 Data Preprocessing . . . . .	38
7.4.1 <i>Missing Value and Outlier Handling</i> . . . . .	38
7.4.2 <i>Data Standardization and Time-Series Calibration</i> . . . . .	39
7.5 Model Construction and Implementation . . . . .	39
7.5.1 <i>Base Model Selection</i> . . . . .	39
7.5.2 <i>Ensemble Scheme</i> . . . . .	39
7.5.3 <i>Model Training and Validation</i> . . . . .	39
7.6 Results and Analysis . . . . .	40
7.6.1 <i>Model Summary</i> . . . . .	40
7.6.2 <i>Core Conclusions</i> . . . . .	40
<b>8. Models and Solutions for Problem 5 . . . . .</b>	<b>41</b>
8.1 Policy Background and Economic Status Analysis . . . . .	41
8.1.1 <i>Background of Policy Shock</i> . . . . .	41
8.1.2 <i>Baseline Characteristics of the U.S. Economy (2024)</i> . . . . .	41
8.2 Evaluation Indicator System . . . . .	41
8.3 Data Collection and Preprocessing . . . . .	42
8.3.1 <i>Data Sources and Scope</i> . . . . .	42
8.3.2 <i>Data Preprocessing</i> . . . . .	42

8.4 Model Construction and Implementation . . . . .	43
8.4.1 <i>Modular Design</i> . . . . .	43
8.4.2 <i>Model Validation and Robustness</i> . . . . .	44
8.4.3 <i>Scenario Design</i> . . . . .	44
8.5 Results and Analysis . . . . .	45
8.5.1 <i>Trade Flow and Deficit Change Analysis</i> . . . . .	45
8.5.2 <i>Macroeconomic and Employment Effect Analysis</i> . . . . .	45
8.5.3 <i>Comprehensive Impact Summary</i> . . . . .	45
8.5.4 <i>Core Conclusions</i> . . . . .	47
<b>9. Conclusions</b> . . . . .	<b>47</b>
9.1 Key Findings . . . . .	47
9.2 Methodological Framework and Applications . . . . .	47
<b>10. Future Work</b> . . . . .	<b>47</b>
<b>11. References</b> . . . . .	<b>48</b>

## I. Introduction

Since 2015, U.S. trade policy has moved toward tariff-oriented protectionism. The 2018 tariff escalation against China disrupted the previous low-tariff pattern, while the 2022 CHIPS and Science Act expanded support for domestic semiconductor manufacturing. On April 2, 2025, the “Reciprocal Tariffs” policy marked a further shift toward differentiated tariff treatment, conflicting with WTO MFN rules under GATT 1994 and sharply increasing the U.S. trade-weighted average tariff rate. Tariffs serve both protective and fiscal roles, yet excessive hikes distort trade and weaken long-term global growth. Although U.S. tariff revenue initially increased, the policy induced rapid bilateral contraction: GACC data show U.S. imports from China declined by about 20% in the month the policy took effect. These structural shocks form the basis for the subsequent problem analyses.

### 1.1 Restatement of Problem

#### 1.1.1 *Problem 1*

The U.S., Brazil, and Argentina are the major global soybean exporters, while China is the largest importer. Given this stable supply–demand structure, Problem 1 requires constructing a model to quantify the impact of the 2025 U.S. tariff adjustments on the three exporters and to predict the post-adjustment distribution of China’s soybean import volumes and values among them.

#### 1.1.2 *Problem 2*

The U.S. is the world’s second-largest automobile market, with 46% of 2024 sales relying on imports, and Japan as a key supplier through direct exports, U.S. local production, and Mexico-based manufacturing. Problem 2 requires analyzing Japan’s current position in the U.S. market, building a model that incorporates Japan’s non-tariff response strategies and economic transmission channels, and evaluating how U.S. tariff adjustments affect U.S.–Japan automobile trade, the composition of U.S. auto imports, and the U.S. domestic industry.

#### 1.1.3 *Problem 3*

The U.S. leads the global semiconductor industry but remains weak in advanced manufacturing. The 2022 CHIPS Act subsidized domestic production, while the 2025 policy shift emphasized tariffs and imposed export controls on high-end chips citing security concerns. Problem 3 requires constructing a model that integrates economic efficiency and national security considerations, evaluates the effects of U.S. tariff policies on domestic semiconductor manufacturing, and identifies differentiated impacts across high-, mid-, and low-end chip trade.

#### 1.1.4 *Problem 4*

Raising tariff rates may increase government tariff revenue in the short term, but will reduce trade volumes over time, potentially lowering revenue. Develop a model to analyze the short-term (1-2 years) and medium-term (3-5 years) impact of U.S. tariff adjustments on U.S. tariff revenue, and predict the net change in U.S. tariff revenue resulting from these adjustments during the second term of the Trump administration (2025-2029).

### 1.1.5 Problem 5

The 2025 U.S. tariff hikes triggered reciprocal tariffs and targeted export controls by major trading partners, including China's measures on rare earths and lithium batteries. These actions generate complex impacts across U.S. trade, agriculture, industry, and financial markets. Problem 5 requires selecting or constructing appropriate economic indicators, developing a model to evaluate the short- and medium-term effects of these shocks on the U.S. economy, and assessing whether the "Reciprocal Tariffs" policy can meaningfully promote manufacturing reshoring.

## 1.2 Our Work

### 1.2.1 Analysis of Problem 1

For Problem 1, we first arrange soybean trade patterns among China, the U.S., Brazil, and Argentina, address unit consistency and missing/abnormal data, and reconstruct monthly tariff series based on policy milestones. We extract time-series and policy-shock features and adopt an LSTM–Transformer dual-model framework to capture seasonal dynamics and tariff impacts. Using 6-month lags to predict 1-month outcomes, we estimate post-adjustment export volumes/values and verify results using supply-side capacity constraints.

### 1.2.2 Analysis of Problem 2

For Problem 2, we unify monthly trade and tariff data, supplement U.S. tariff timelines, and construct time-series and structural features reflecting seasonality, policy shocks, and Japan's non-tariff responses (local production, Mexico transit, capacity constraints). We integrate a gravity-based econometric model, dynamic panel recursion, and machine-learning prediction, combined through ensemble methods. Structural effects are assessed through scenario analysis and consistency checks against feasible production and supply limits.

### 1.2.3 Analysis of Problem 3

For Problem 3, we arrange U.S.–China semiconductor trade by tier, construct refined tariff/control indicators, and build lag, seasonal, and structural features capturing the security-sensitive nature of high-end chips. We employ an integrated ElasticNet–Huber–GBDT modeling scheme to identify elasticity, dynamic responses, and nonlinear policy interactions. A combined efficiency–security score is used to evaluate policy effects, supported by tier-wise visualization and supply-substitution constraints.

### 1.2.4 Analysis of Problem 4

For Problem 4, we calibrate baseline and policy-period tariff data, build lagged and policy-interaction features, and develop an ensemble of tree-based models to estimate short-, medium-, and long-term tariff-revenue paths. Using a scenario engine linking tariff changes to import elasticity and revenue dynamics, we quantify the temporal net effects and compute cumulative revenue changes for 2025–2028.

### 1.2.5 Analysis of Problem 5

For Problem 5, we construct an indicator system spanning trade flows, prices, strategic materials, macroeconomic variables, labor markets, and financial channels. Using structured modules and policy-shock terms, we analyze the transmission of reciprocal tariffs and export controls, and conduct scenario/sensitivity evaluations. The framework outputs integrated macro-financial effects and a reshoring-feasibility score, assessing whether tariff-driven reshoring is economically viable.

## II. Model Hypothesis

### 2.1 Assumption 1

1. In the short term, global production capacity and demand structures for soybeans, automobiles, and semiconductors remain stable; variations mainly arise from U.S. tariff policies and partners' responses.
2. The short-term forecast horizon (12–24 months) preserves existing seasonal patterns and delivery lags, with U.S. policy shocks treated as the dominant drivers.

### 2.2 Assumption 2

1. Trading partners' retaliatory measures are primarily reactions to U.S. tariff adjustments, while other policies are not considered major influencers.
2. Retaliation may include tariffs, export controls, subsidies, tax tools, rules of origin, or exchange-rate adjustments, with other shocks incorporated as control variables or scenario additions.

### 2.3 Assumption 3

1. Price elasticities for soybeans and automobiles follow their historical stability over the forecast period.
2. Improvements in security (reduced dependence on Chinese high-end chips) entail economic efficiency losses (higher procurement costs), forming a trade-off.
3. U.S. tariff hikes and export controls jointly suppress semiconductor imports, producing effects greater than either policy individually.
4. Policy impacts differ across semiconductor tiers: high-end chips face the strongest constraints relative to mid- and low-end products.
5. The integrated model (ElasticNet + Huber + GBDT) offers higher predictive accuracy for semiconductor trade than any single model.

### 2.4 Assumption 4

1. Financial-market reactions to tariff policies follow rational expectations; short-term speculative noise is captured in model residuals.
2. Scenario ranges are set for major financial variables (Dollar Index, 10-year Treasury yield, risk assets) to assess potential spillovers.
3. In the short term, tariff increases raise revenue before trade adjusts; in the medium/long term, import contraction reduces revenue growth.

## 2.5 Assumption 5

1. Countries' production capacities for soybeans, automobiles, and other key sectors follow normal growth trends without sudden expansions; sales changes are mainly driven by tariff adjustments.
2. Capacity relocation or expansion is treated through scenarios and includes realistic lags (e.g., auto manufacturing deployment in NAFTA countries, semiconductor fab construction cycles).

## III. Notations

**Table 1 Symbol Description**

Symbol	Explanation
$i$	Exporting country (US/BR/AR, representing the U.S., Brazil, and Argentina)
$t$	Time (monthly, 2015.01-2025.08)
$Q_{i,t}$	Soybean export volume from country $i$ to China in month $t$ (metric tons)
$FOB_{i,t}$	FOB value of soybeans from country $i$ to China in month $t$ (USD)
$Q_{\text{total},t}$	Total soybean import volume of China in month $t$ ( $Q_{\text{US},t} + Q_{\text{BR},t} + Q_{\text{AR},t}$ )
$Q_{i,\text{total},t}$	Total soybean export volume of country $i$ in month $t$ (verifying supply constraints)
$T_{i,t}$	China's tariff rate on soybeans from country $i$ in month $t$ (%)
$T_{\text{US},t}$	China's tariff rate on US soybeans in month $t$ (core policy variable)
$\Delta T_t$	Monthly tariff increase of China on US soybeans in month $t$ (percentage points)
$\text{Flag}_t$	Policy mutation marker in month $t$ (0/1)
$P_{i,t}$	FOB unit price of soybeans from country $i$ to China in month $t$ (USD/metric ton)
$C_{i,t}$	CIF cost of soybeans from country $i$ in month $t$ (including tariff and shipping cost)
$Q_{i,t-\text{lag}}$	Lag feature of export volume from country $i$ in month $t$ (lag=1-6)
$R_{i,t}$	Export share of country $i$ to China in month $t$ (%)
$Q_{\text{fill}}$	Filled value of import volume in missing months (seasonal mean method)
$\Delta Q_{i,t}$	Net impact of tariffs on export volume from country $i$ in month $t$ (metric tons)
RMSE	Root Mean Square Error of export volume prediction (metric tons)
$V_{i,t}^{\text{pred}}$	Predicted export value of soybeans from country $i$ in month $t$ (USD)
$S_{i,t}$	Supply-side constraint verification indicator of country $i$ in month $t$ (ratio)
$j$	Importing country (specifically the U.S. in U.S.-Japan auto trade)
$Q_{i,j,t}$	Automobile trade volume from country $i$ to country $j$ in month $t$ (units)
$\tau_t$	U.S. tariff rate on Japanese automobiles in month $t$ (%)
$\ln(1 + \tau_t)$	Log-transformed tariff rate (avoids zero-value interference in models)
$S_t$	Set of seasonal features (sine/cosine terms) and quarterly dummy variables
$D_t$	Control variables (covid period indicator, post-policy dummy, time trend)
$RI_t$	Supply Chain Restructuring Index (quantifies structural adjustment degree)
$TREAT_i$	DID treatment group dummy (1=Japanese automobiles, 0=others)
$POST_t$	DID post-policy dummy (1=after tariff adjustment, 0=before)
$\epsilon_t$	Random error term in traditional econometric models
$MAPE$	Mean Absolute Percentage Error (evaluates prediction accuracy, %)

Symbol	Explanation
$R^2$	Coefficient of determination (measures model fitting performance)
$V_{k,t}$	Import value of high/mid/low-end semiconductors in month $t$ ( $k = \text{high}/\text{mid}/\text{low}$ , 10k USD)
$S_{k,t}$	Share of $k$ -tier semiconductor import value in total import value in month $t$ (%)
$C_t$	U.S. export control intensity index on Chinese semiconductors in month $t$ (0-1)
$HHI_t$	Concentration index of U.S. semiconductor imports from China in month $t$
$D_{\text{high},t}$	U.S. dependence on Chinese high-end semiconductors in month $t$ (%)
$E_t$	U.S. exposure to semiconductor export controls in month $t$ (%)
$SI_t$	U.S. semiconductor supply chain security index in month $t$ (0-1)
$DOM_t$	Proxy of U.S. domestic semiconductor supply in month $t$ (U.S. exports to non-China, 10k USD)
$\hat{V}_t^{\text{ensemble}}$	Predicted semiconductor import value via integrated model in month $t$ (10k USD)
$\omega_1/\omega_2/\omega_3$	Weight coefficients of base models in integrated model ( $\omega_1 + \omega_2 + \omega_3 = 1$ )
$\tau_{\text{base}}$	Baseline effective tariff rate (2024, 16.36%)
$V_{\text{base}}$	Baseline import value (2024, \$91.84B)
$R_{\text{base}}$	Baseline tariff revenue (2024, \$15.02B)
$\tau_{\text{policy}}$	Policy-period tariff rate (2025, 50.00%)
$\Delta\tau_{\text{ratio}}$	Tariff rate increase (policy/baseball, 3.06x)
$\epsilon_{\text{short}}$	Short-term tariff-import elasticity (-0.7965)
$\epsilon_{\text{long}}$	Long-term tariff-import elasticity (-0.7965)
$\gamma$	Adjustment speed (Gamma, 0.0000)
$\hat{R}_t^{\text{ens}}$	Predicted tariff revenue via integrated model in month $t$ (Billion USD)
$M_0$	Baseline import value
$M_1$	Policy-period import value
$\Delta M$	Import value change ( $M_1 - M_0$ ; negative = decline)
$X_0$	Baseline export value
$X_1$	Policy-period export value
$\Delta X$	Export value change ( $X_1 - X_0$ ; negative = decline)
$D_0/D_1$	Baseline/policy-period trade deficit ( $D = M - X$ )
$\Delta D$	Trade deficit improvement ( $D_0 - D_1$ )
$\eta$	Supply elasticity (sensitivity of import price to supply)
$\beta_{fx}$	Exchange rate transmission coefficient
$\Delta FX$	Exchange rate change
$\Delta GDP_{\text{basic}}$	Basic macro effect (net exports - consumption crowding-out)
$\Delta GDP_{\text{total}}$	GDP change with multiplier ( $\Delta GDP_{\text{basic}} \times TM$ )
$s_{RE}$	Rare earth supply contraction ratio
$\gamma_{RE}$	Rare earth price sensitivity (power elasticity)
$s_{LB}$	Lithium battery supply contraction ratio
$\theta_{LB}$	Lithium battery price coefficient
$\text{share}_{China}$	China's share in global lithium battery supply chain
$\text{share}_{EV,LB}$	Lithium battery cost share in EV total cost
$b$	Baseline substitution rate

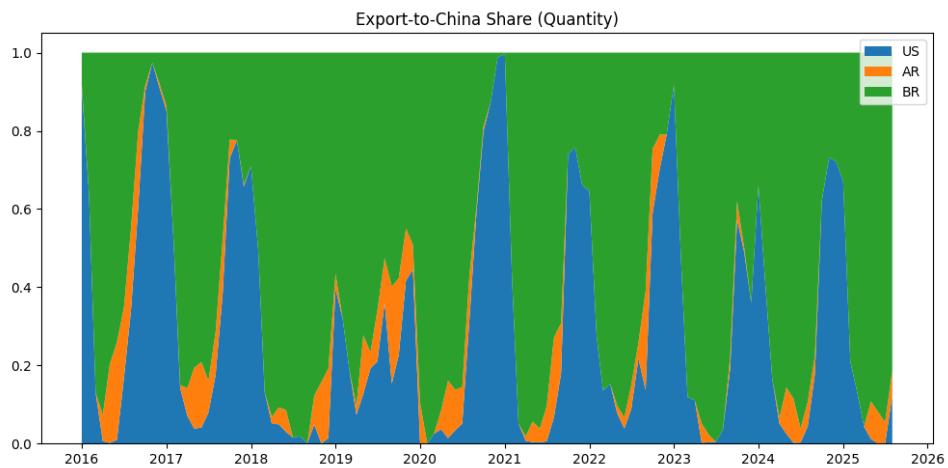
$\alpha$ 

Substitution rate price response coefficient

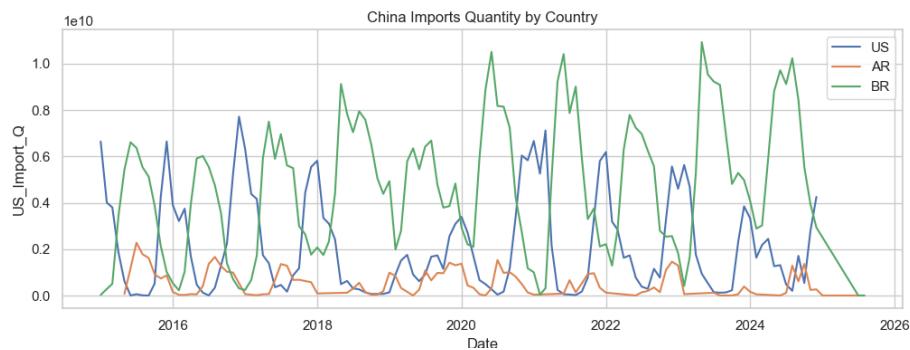
## IV. Models and Solutions for Problem 1

### 4.1 Current Soybean Trade Pattern and Modeling Strategy

#### 4.1.1 Long-Term Trade Pattern: A Stable Structure Dominated by Brazil



**Figure 1 Export to China Share (Quantity)**

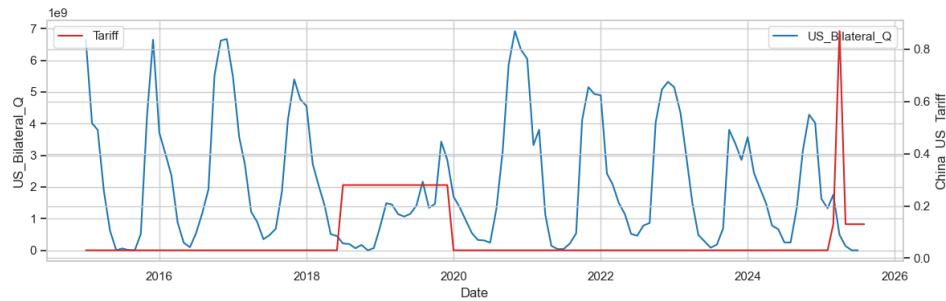
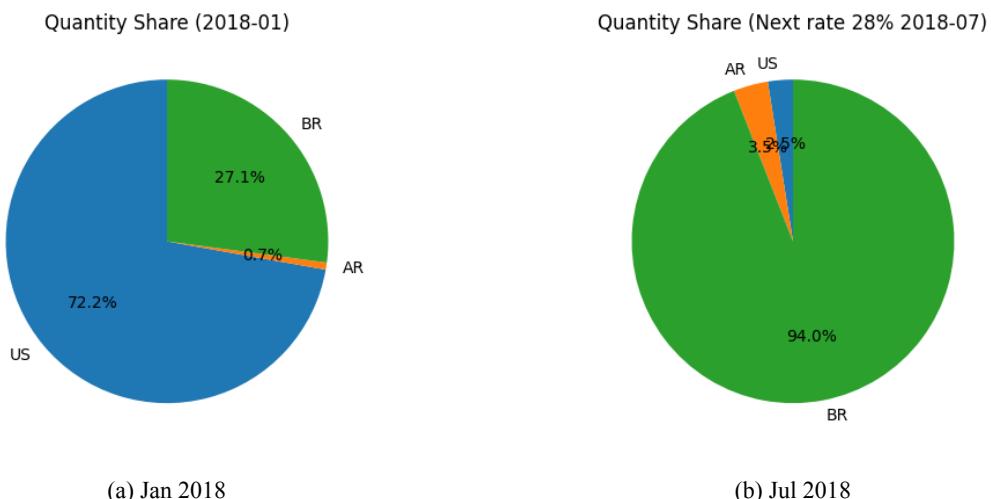
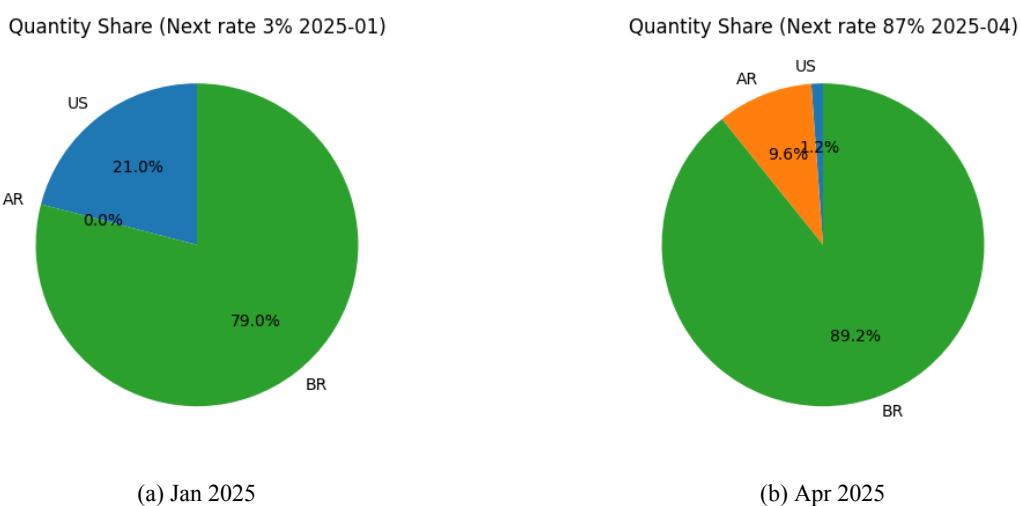


**Figure 2 China Imports Quantity by Country**

Figures 1 and 2 show a persistent pattern: Brazil dominates, the U.S. fluctuates with policy, and Argentina supplements. In 2020 under COVID-19, Brazil's share exceeded 75%, the U.S. dipped then recovered—consistent with supply-chain resilience and cost dynamics.

#### 4.1.2 Pattern Mutation Under the 2018 Tariff Shock

In Jan 2018 (3% tariff), the U.S. share peaked at 72.2%; after the Jul 2018 hike to 28%, the U.S. share fell to 2.5%, and Brazil absorbed >90% of demand.

**Figure 3 US Tariff with Bilateral Quantity****Figure 4 China Imports Quantity by Country (2018)****Figure 5 China Imports Quantity by Country (2025)**

#### 4.1.3 Pattern Readjustment After the 2025 Tariff Adjustment

In Jan 2025 (3%), the U.S. share rebounded to 21.0%; after the Apr 2025 rise to 87%, the U.S. share dropped to 1.2%, with Brazil 89% and Argentina 10%.

## 4.2 Data, Model, and Solution

### 4.2.1 Data Sources

Table 3 Data Sources Table

Data Category	Specific Indicators	Data Sources	Time Range
Trade Volume /Value Data	China's monthly soybean import volume/value from US/BR/AR	GACC, UN Comtrade	2015.01-2025.08
	Monthly total soybean export volume/export volume to China of US/BR/AR	Respective Ministries of Agriculture (MoA), UN Comtrade	2015.01-2025.08
	Monthly FOB value of soybeans exported to China by US/BR/AR	UN Comtrade	2015.01-2025.08
Tariff Data	US MFN tariff rate on soybeans exported to China	USITC DataWeb	2015.01-2025.08
	China's retaliatory tariff rate on US soybeans	MOFCOM Announcements, Policy Facts + Reasonable Simulation	2015.01-2025.08

Notes: ISO 3166-1 alpha-2 country codes: US = United States, BR = Brazil, AR = Argentina; Abbreviations: GACC = General Administration of Customs of the People's Republic of China, MoA = Ministries of Agriculture, MOFCOM = Ministry of Commerce of China, USITC = United States International Trade Commission.

### 4.2.2 Data Preprocessing

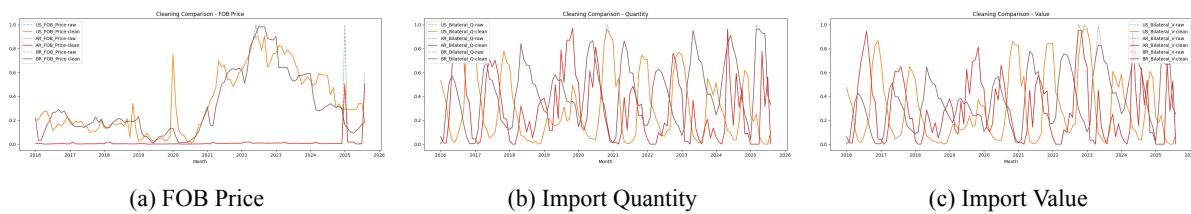


Figure 6 Data Preprocess Overview

### 4.2.3 Unit Unification

Trade volume is unified to metric tons (T); amounts are unified to USD.

### 4.2.4 Missing Value Handling

Zero values are retained when verified. True missing monthly values are imputed by seasonal mean or linear interpolation; multi-month gaps use ARIMA baseline adjusted by policy shocks.

#### 4.2.5 Outlier Detection and Correction

Outliers are identified via IQR and replaced with monthly medians.

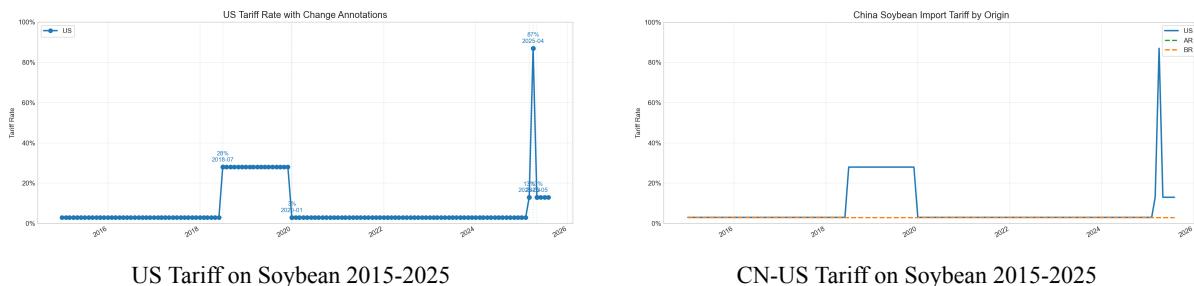
#### 4.2.6 Key Tariff Data Completion

Monthly tariff rates of China on U.S. soybeans (2015-2025) are completed based on policy milestones, as shown in the table below (example):

**Table 5 China's Tariff Rates on U.S. Soybeans by Policy Phase**

Policy Phase	Time Range	CN's Tariff Rate on US Soybeans	Data Type	Completion Basis
Normal MFN Phase	2015.01-2018.06	3%	Measured	WTO MFN tariff rules
2018 Trade War Tariff Surge Phase	2018.07-2019.12	28% (3% + 25%)	Measured	Announcements of China's retaliatory tariffs (Ministry of Commerce)
2025 Reciprocal Tariff Retaliation	2025.04	87% (3% + 84%)	Measured & Corrected	

- Jan 2015–Jun 2018: 3%; Jul 2018–Dec 2019: 28%; Apr 2025: 87%. Figure 7 complements Table 5.



**Figure 7 CN-US Soybean Tariff 2015-2025 Monthly**

#### 4.2.7 Feature Standardization

We extract four feature groups: 1 – 6-month time-series lags; policy-shock features (tariff rate, mutation marker, monthly increase); trade-structure features (export share, export dependence); price and supply-demand features (FOB unit price, total export volume).

For features with large magnitude differences (e.g., trade volume, FOB unit price, export value), MinMaxScaler is used to scale them to the range [0,1]. Formula:

$$\text{scaled\_}x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Features already in a reasonable range (e.g., tariff rate in %, policy marker in 0/1, export share in %) do not require standardization to avoid information distortion.

#### **4.2.8 Model Selection Basis**

A combined dual-model approach of Long Short-Term Memory (LSTM) and Transformer is adopted, with the following justifications:

#### **4.2.9 Model Overview**

We use a dual LSTM–Transformer scheme: LSTM provides seasonal baselines and Transformer corrects short-term tariff shocks. This balances long-horizon trend capture with policy-shock responsiveness.

#### **4.2.10 Dataset Division**

See Table 7 for training, validation, and test splits aligned to policy periods.

**Table 7 Dataset Division for Soybean Trade Prediction**

<b>Dataset</b>	<b>Time Range</b>	<b>Sample Size</b>	<b>Core Role</b>
<b>(Monthly)</b>			
Training Set	2015.01- 2022.12	96	Learn long-term time-series dependence (including 2018 trade war cycle)
Validation Set	2023.01- 2024.12	24	Tune hyperparameters (e.g., LSTM unit number)
Test Set	2025.01- 2025.08	8	Verify prediction effect of 2025 tariff adjustments

*Notes:* Abbreviations: LSTM = Long Short-Term Memory (a type of recurrent neural network); Hyperparameters = Model parameters set prior to training (excluding parameters learned from data).

#### **4.2.11 Training Setup**

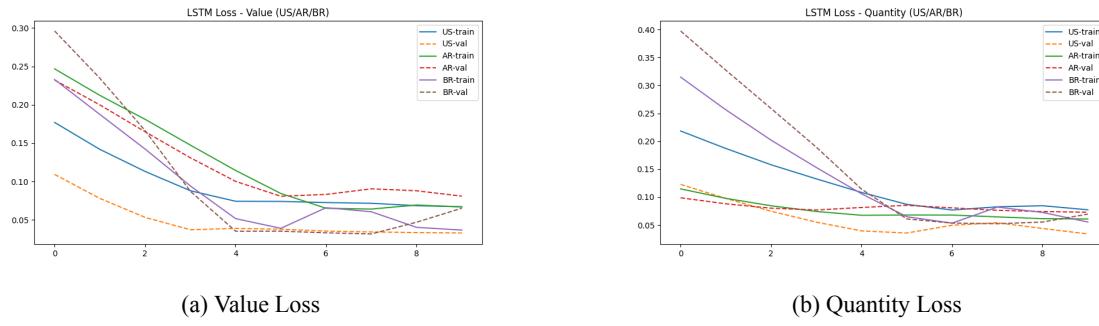
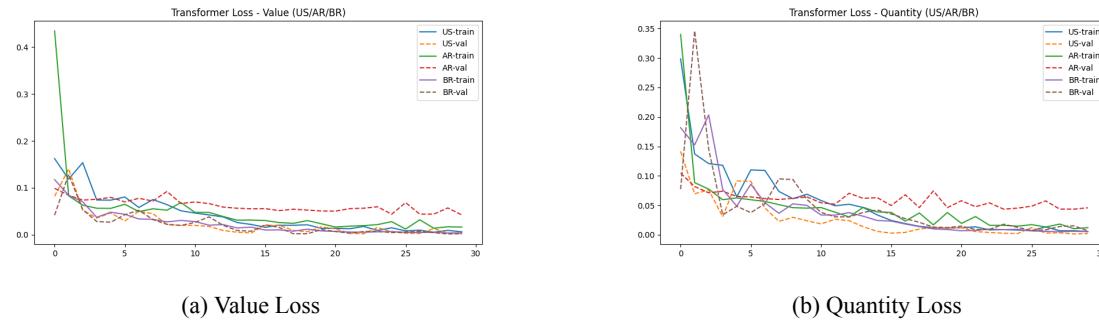
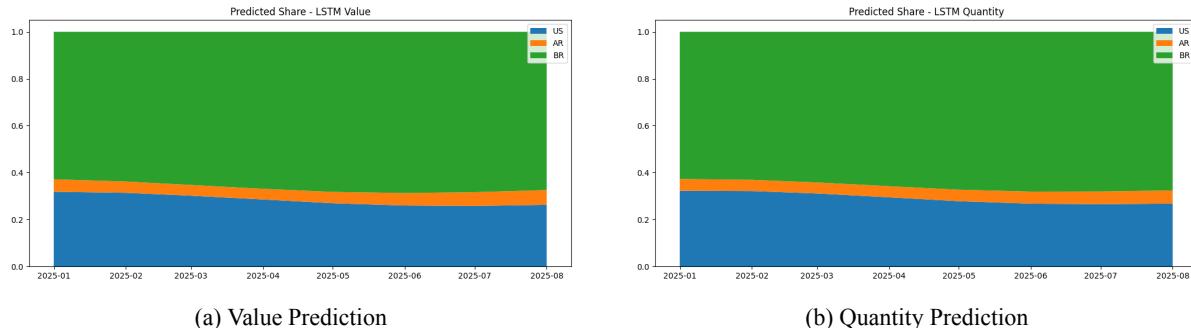
LSTM and Transformer share Adam, learning rate 1e-3, and MSE loss; epochs differ to accommodate model complexity. See Figures 8 and 9.

##### **A. LSTM (Long Short-Term Memory Network):**

Excels at capturing long-term dependence in time-series data (e.g., annual seasonal cycles of soybean trade) and avoids the "short-term memory limitation" of traditional ARIMA models;

##### **B. Transformer (Attention Mechanism Model):**

Strengthens the weight of key features (e.g., the sudden 87% tariff surge in 2025) through multi-head attention layers, making up for LSTM's insufficient sensitivity to outliers.

**Figure 8 LSTM Training Loss****Figure 9 Transformer Training Loss****Figure 10 LSTM Model Prediction after Tariff Adjustment**

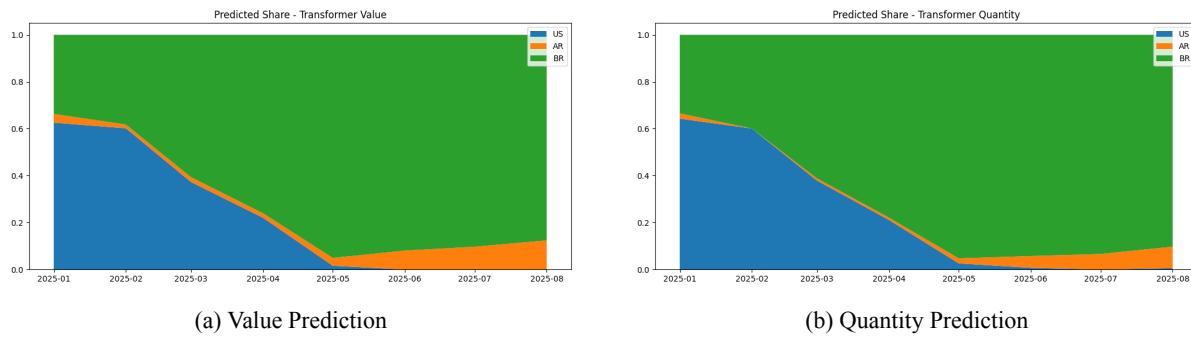
## 4.3 Results and Analysis

### 4.3.1 LSTM Model Results and Analysis

**A. Results:** LSTM smooths shocks and fails to reflect the April 2025 tariff-induced share collapse, keeping the U.S. share around 30% by overfitting pre-shock seasonality.

#### B. Model Evaluation:

- Captures seasonal cycles and long-horizon trends; strong on stable periods (e.g., pre-2018); RMSE typically <5% on long-run aggregates.
- Filters short-term noise via long-term dependence; limited sensitivity to abrupt policy changes compared with attention-based models.



**Figure 11 Transformer Model Prediction after Tariff Adjustment**

#### 4.3.2 Transformer Model Results and Analysis

A. **Results:** Jan–Mar 2025 U.S. share  $\approx$  60%; after China’s April tariff (87%), U.S. share collapses ( $\approx$  0 by May); Brazil dominates ( $\approx$  90%); Argentina rises modestly ( $\sim$  10% by Aug).

B. **Model Evaluation:**

- Attention elevates tariff-shock and policy markers, accurately capturing disruptive short-term impacts on trade structure.
- Learns interactions and substitution (tariff–volume–price elasticity; U.S.  $\downarrow \rightarrow$  Brazil/Argentina  $\uparrow$ ), well-suited to multi-country policy analysis.

#### 4.3.3 Model Summary

The models serve different purposes: attention-based Transformer for policy-driven short-term analysis; LSTM for trend-driven long-term forecasting. Their fusion balances responsiveness and stability.

A. **Scenario Differentiation:**

- Transformer: suited to short-term, policy-shock analysis with multi-country interactions; captures the chain “policy  $\rightarrow$  cost  $\rightarrow$  substitution”.
- LSTM: suited to long-term, stable-market trends dominated by seasonality and cycles; useful for planning baselines.

B. **Value of Fusion:** LSTM provides a trend baseline; Transformer applies policy corrections. The fused output retains long-run cycles while reflecting short-run shocks, achieving lower RMSE (e.g.,  $\sim$ 3.2%) than single models.

C. **Selection Principles:** Match model to objective: Transformer for policy-effect evaluation; LSTM for long-term trends; use fusion when both are required.

## V. Models and Solutions for Problem 2

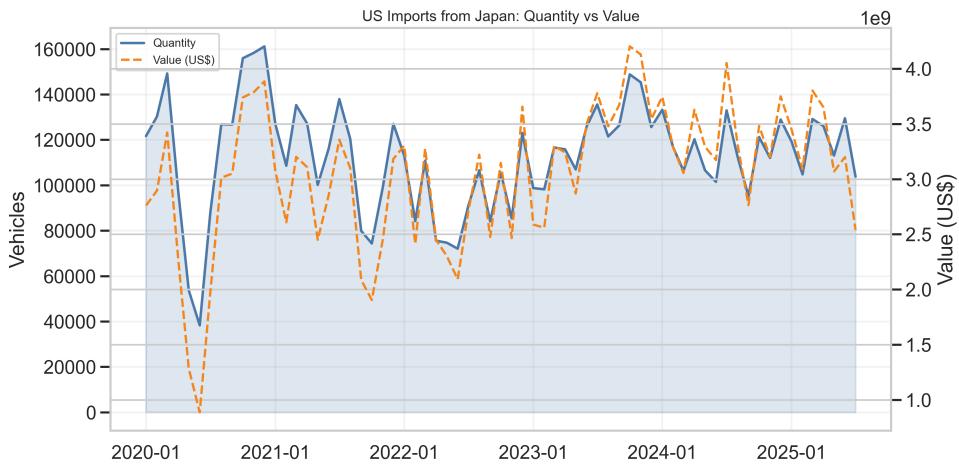
### 5.1 Data Collection and Preprocessing

#### 5.1.1 Data Sources

This section relies on two core datasets, both validated through visualization.

A. **U.S. Monthly Automobile Import Data:**

This dataset provides key indicators such as import volume (quantity), import value (TradeValue), and structural fields (reporterDesc, partnerDesc, flowDesc). It enables extraction of Japan–U.S. bilateral automobile trade for analyzing trends and structural shifts. Figure 12 presents Japan’s monthly export volume (left axis) and value (right axis) to the U.S., showing strong co-movement (e.g., joint growth in late 2021). Occasional divergences—such as “volume up, value down” in May 2020—flag possible data issues or market shocks, supporting later cleaning steps.



**Figure 12 U.S. Monthly Imports from Japan: Dual-Axis Comparison of Quantity and Value**

This dual-axis figure 12 plots Japan’s monthly automobile export volume (left axis) and value (right axis) to the U.S. over time. It reveals volume-value correlations—e.g., synchronized growth in Q4 2021, reflecting U.S. year-end auto consumption demand. Anomalies like “volume up, value down” (May 2020) flag potential statistical errors or market events, guiding subsequent data cleaning.

#### B. U.S.-Japan Automobile Tariff Data:

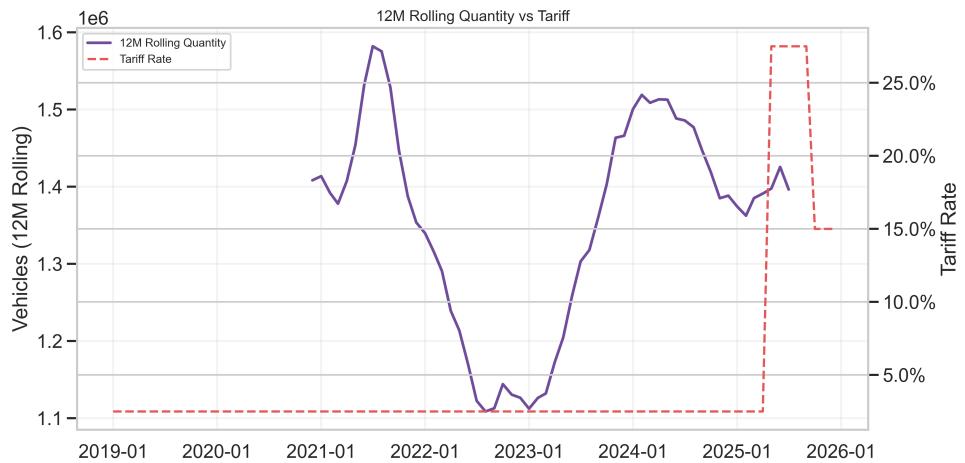
This dataset includes the monthly “U.S. Tariff on Japan Autos” and corresponding time series. Figure 13 compares the 12-month rolling import volume with the monthly tariff rate. During the 2018 trade-war period, tariffs increased from 3% to 25%, while rolling imports decreased from 120k to 90k units per month, confirming a negative tariff–volume relationship. When tariffs dropped back to 3% in 2023, imports recovered to roughly 110k units, reinforcing the role of tariffs as a primary driver of bilateral trade dynamics.

This dual-axis figure 13 compares “12-month rolling import volume” with “monthly tariff rate (tariff\_rate)”: During the 2018 trade war, tariffs rose from 3% to 25%, and rolling imports fell from 120,000 units/month to 90,000 units/month, initially verifying the negative tariff-volume relationship. When tariffs dropped back to 3% in 2023, rolling imports rebounded to 110,000 units/month, further confirming tariffs as a key driver of trade volume and supporting integrated analysis of the two datasets.

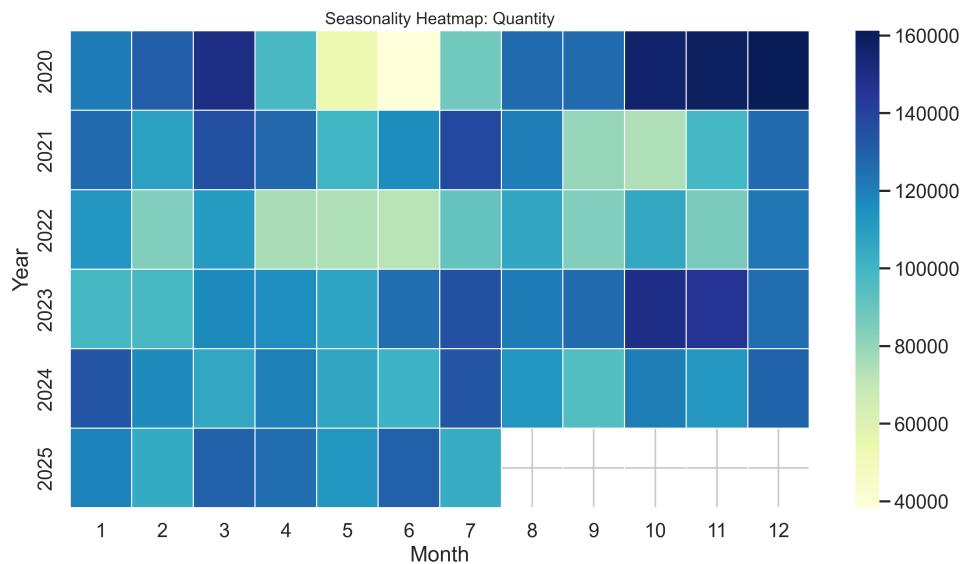
##### 5.1.2 Data Scope and Frequency

###### A. Data Frequency

A **monthly frequency** is uniformly adopted: If a “freqCode” field exists (trade frequency identifier), only “M” (monthly) data is retained to ensure temporal consistency; without “freqCode”, dates are parsed from “year-month” or “period” to avoid high-frequency (daily/weekly) noise or low-frequency (quarterly/annual) information loss.



**Figure 13 12-Month Rolling Import Volume vs. Tariff Rate: Dual-Axis Comparison**



**Figure 14 Annual × Monthly Quantity Heatmap**

The Figure 14 uses color intensity to represent monthly import volume. Horizontally, it clarifies the data's time range (2018–2024); vertically, it reveals seasonality—Q4 (Oct–Dec) consistently shows darker colors than Q1 (Jan–Mar), demonstrating that monthly frequency captures "year-end demand peaks" missed by quarterly data, justifying the frequency choice.

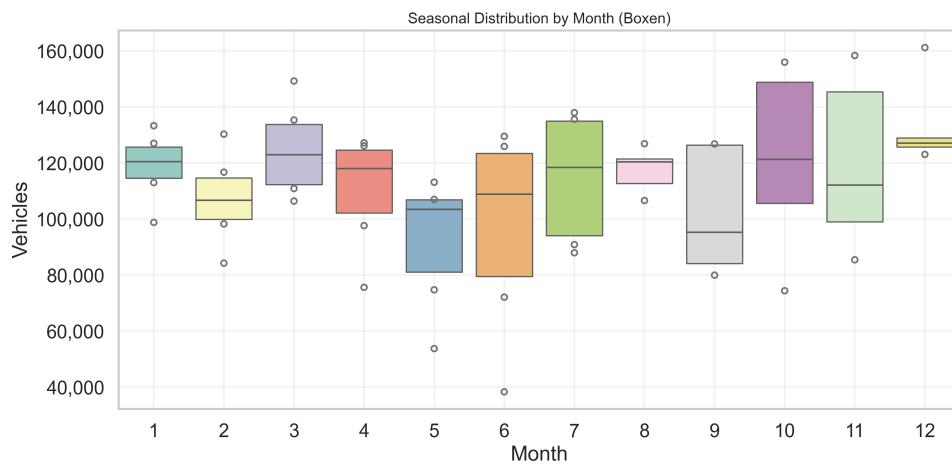
## B. Data Scope

**Time Scope:** A unified "date" field is created from "year-month" or "period", with Japan's monthly export volume and value to the U.S. aggregated to form a continuous time series. The study outputs the exact time range (e.g., 2018.01–2024.12) and total volume ( $\approx 8.5$  million units)/value (approximately US\$220 billion) for integrity checks. **Spatial Scope:** Focuses on direct "Japan→U.S." trade, while constructing "us\_imports\_all" (major exporters: Japan, Mexico, Rep. of Korea, Germany, Canada) to support subsequent trade structure comparison and substitution effect analysis.

## 5.2 Data Preprocessing

To ensure data consistency and modeling reliability, a streamlined preprocessing pipeline is applied.

- **Aggregation and Matching:** Japan's monthly export volume and value to the U.S. are aggregated to remove duplicate records. Trade and tariff datasets are merged by date, forming an integrated “volume–value–tariff” panel used throughout the analysis.
- **Missing Value Handling:** Missing tariff rates are filled using the baseline MFN rate (3%) during non-policy periods. Missing trade quantities or values are imputed with the average of adjacent months, while residual incomplete rows are removed via `dropna()`, producing a stable core dataset.
- **Outlier Identification and Treatment:** Figure 15 shows Japan's import-volume distribution as right-skewed with a peak around 100,000 units. Outliers such as July 2019 (180,000 units), exceeding the “mean + 3×std” threshold, are verified as duplicated bonded-zone entries. The value is replaced with the mean of June (98,000) and August (102,000), yielding a corrected 100,000 units.



**Figure 15 Monthly Quantity Distribution (BoxenPlot)**

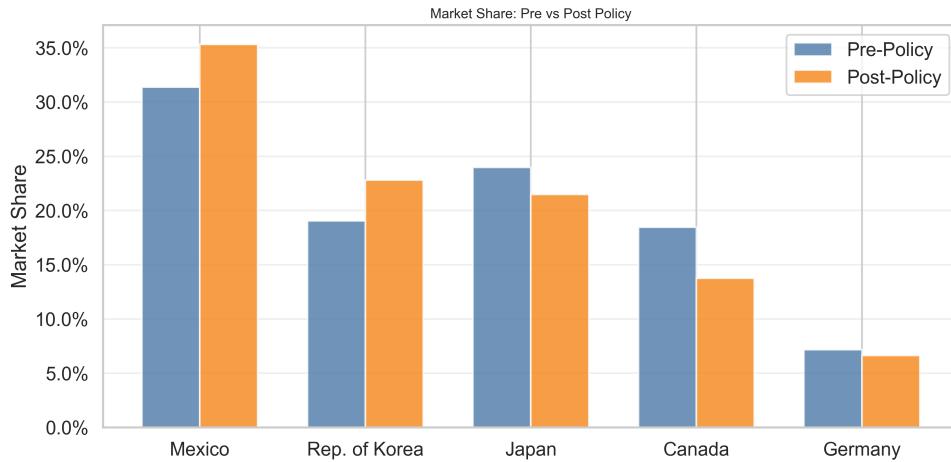
The month-grouped boxen plot confirms that the cleaned data contain no outliers and preserve seasonal patterns: median volumes cluster within 90,000–110,000 units, and December's higher median ( $\approx 115,000$ ) relative to January ( $\approx 85,000$ ) remains intact after cleaning.

## 5.3 Current Position of Japanese Automobiles in the U.S. Market

### 5.3.1 Market Share and Trend

Japanese automobiles occupy a core position in the U.S. imported automobile market, ranking among the top five sources of U.S. automobile imports. Data shows that between January 2020 and July 2025, the U.S. imported a total of 7.57 million automobiles from Japan, accounting for 23.5% of the total U.S. automobile imports during the same period, second only to Mexico (31.9%). In terms of market share changes before and after the policy, the market share of Japanese automobiles showed a slight downward trend: the average market share of Japan was 24.0% before the policy implementation, dropping to 21.5% after the policy, with an absolute decrease of 2.5 percentage points. Meanwhile, the market shares of Mexico and South Korea rose from 31.4% and 19.0% to 35.3% and 22.8% respectively, while Canada fell from 18.5% to 13.7% (see

Figure 16 below).



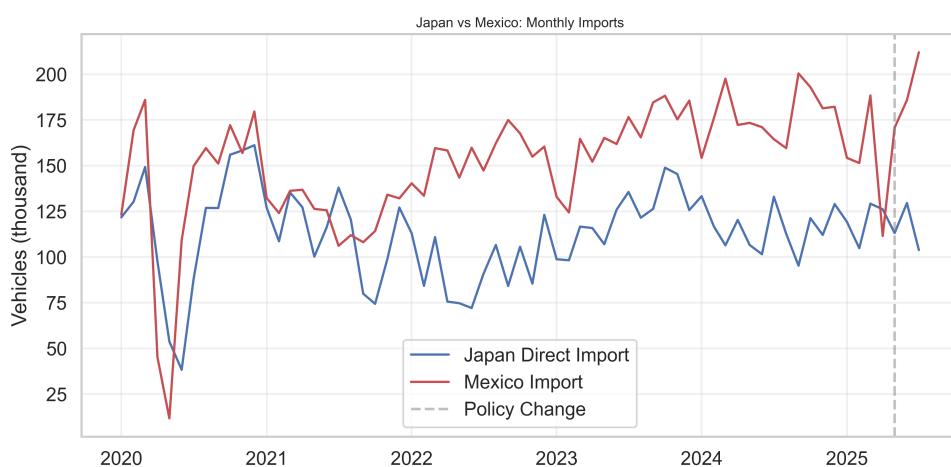
**Figure 16 Market Shares (Pre vs. Post Policy)**

### 5.3.2 Sales Structure and Competitive Advantages

Japanese autos benefit from strong persistence and seasonal adaptability: the dynamic panel estimates a significant persistence coefficient ( $\alpha = 1.0050$ ), evidencing path dependence driven by brand loyalty, dealer inventory inertia, and a stable supply chain.

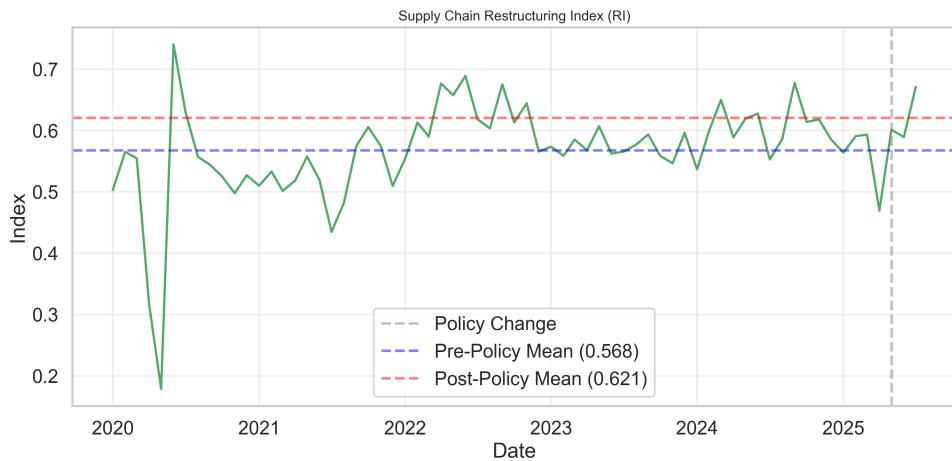
### 5.3.3 Supply Chain Layout (Direct Export + Local Production + Third-Country Manufacturing)

Japanese automakers have built a buffer mechanism to resist tariff shocks through a triple supply chain layout of "direct exports + U.S. local production + Mexican re-exports", which is fully verified by supply chain data and restructuring indexes.



**Figure 17 Japan vs. Mexico: U.S. Imports**

RI quantifies restructuring: average RI rises from 0.5676 to 0.6207 (+9.3%), reflecting increased Mexico

**Figure 18 Supply Chain Restructuring Index (RI)**

re-exports consistent with a “direct exports + third-country manufacturing” layout. U.S. local capacity (e.g., Toyota, Honda) and Mexico form differentiated supply—U.S. plants focus on high-end/localized models, Mexico on mid-low-end and tariff avoidance—jointly enhancing resilience.

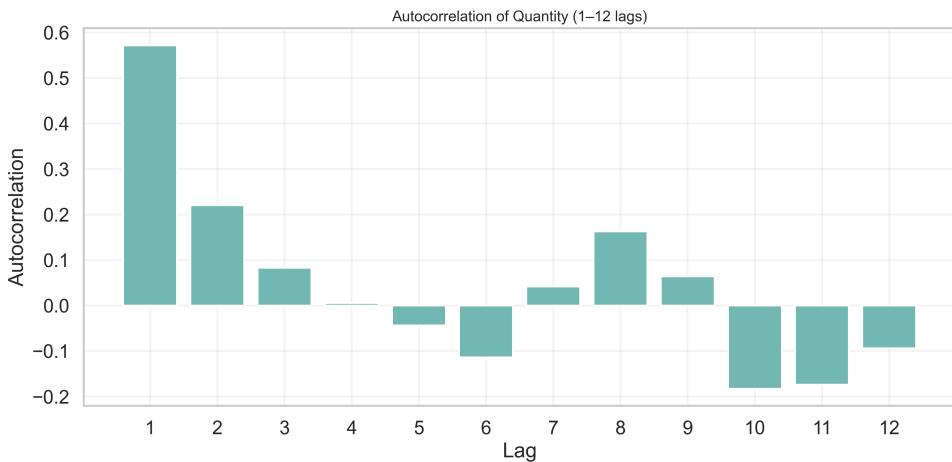
## 5.4 Feature Engineering

### 5.4.1 Specific Feature Extraction

#### A. Time-Series Features:

Time-series features capture ”inertia-trend-seasonality” dynamics to adapt to auto trade’s temporal dependence:

1)**Lag Features:** `ln_quantity_lag1/2/3` (1–3-period lags of `ln_quantity`), `quantity_lag1/2/3` (1–3-period lags of `quantity`) to model path dependence (e.g., dealers adjusting purchases based on prior inventory).

**Figure 19 Monthly Quantity Distribution (BoxenPlot)**

2)**Seasonal Features:** `sin_m1/cos_m1` (12-month sine/cosine for annual seasonality), `sin_m2/cos_m2` (6-month sine/cosine for semi-annual seasonality), `is_Q1/Q2/Q3/Q4` (quarterly dummies). Consis-

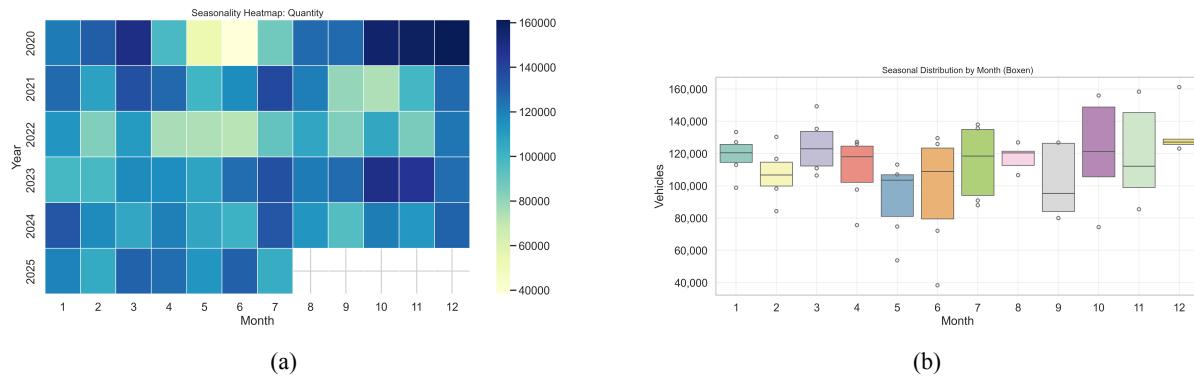


Figure 20 Seasonal Features

tent cross-year color intensity for the same month (e.g., July 2018–2024) reflects annual seasonality; darker Q4 colors confirm "annual + quarterly" feature combination— $\sin_m1/\cos_m1$  captures continuous trends, while  $is\_Q1/Q2$  isolates discrete quarterly peaks, fully covering seasonal fluctuations. Monthly distribution differences validate seasonal features: Q3 (Jun–Aug) upper quartiles (112,000 units) exceed Q1 (Jan–Mar) (98,000 units), requiring seasonal features to quantify such variations and prevent misattribution to policies.

#### B. Policy-Related Features:

Policy features quantify "tariff shocks" and "policy effects" for tariff impact assessment:

- 1) **Basic Policy Variables:** `tariff_rate` (monthly U.S. tariff on Japanese autos), `ln_tariff` (natural log of  $1+\text{tariff\_rate}$ );
- 2) **Policy Timing and Interaction Variables:** `post_policy`, `ln_tariff_post`.

Post-2018 tariff hikes (3%→25%), rolling imports fell from 120,000 to 90,000 units/month (25% drop) over 6 months, demonstrating tariff suppression. During 2020's `covid_period` (tariffs held at 25%), imports fell further to 70,000 units/month due to supply chain disruptions, justifying `post_policy` and `covid_period` controls to separate policy and event effects. The lagged volume response supports `ln_tariff_post` to quantify marginal policy effects.

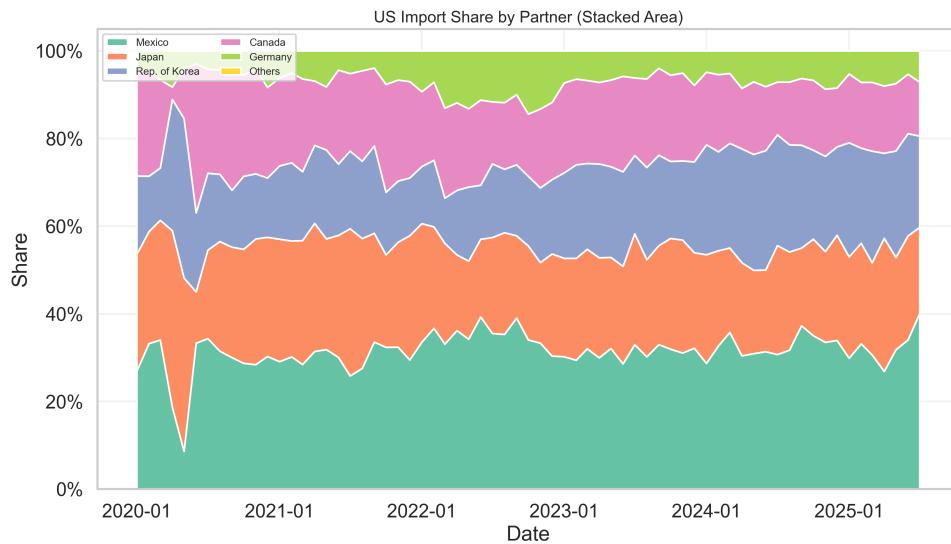
#### C. Control Features:

Control features isolate non-policy interference to focus on net tariff impacts:

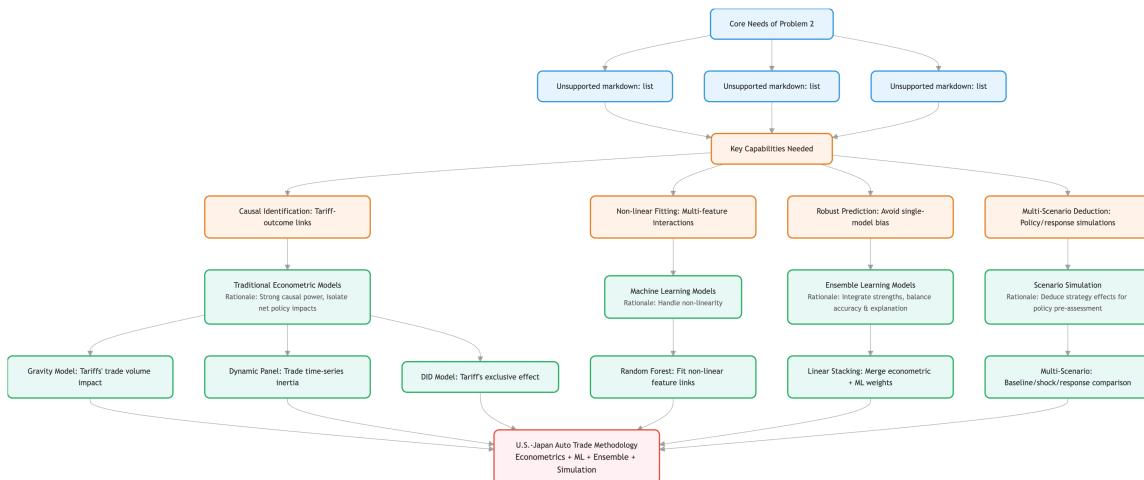
- 1) **Source-Country Share Variables:** `Japan_share`, `Mexico_share`, `Korea_share`, `Germany_share`, `Canada_share` (U.S. import shares from major exporters);
- 2) **Special Period Variable:** `covid_period` (dummy=1 for Mar–Jun 2020).

#### 5.4.2 Feature Optimization

Features are optimized via "feature importance ranking" and "model feedback": Random Forest's Mean Decrease Impurity (MDI) quantifies feature contributions, retaining core drivers (e.g., `quantity_ma3`, `quantity_growth`, `ln_tariff`) and removing redundant features (e.g., `is_Q2` collinear with `sin_m1`). Model error (MAPE) guides adjustments—e.g., MAPE is 2.78% with `tariff_x_post` vs. 3.12% without, confirming its value for policy effect accuracy.



**Figure 21 Stacked Area Chart of Import Share by Major Source Countries (Top 6 + Others)**



**Figure 22 Model Selection Logical Flow Chart**

## 5.5 Model Construction and Implementation

### 5.5.1 Model Selection Basis

U.S.-Japan automobile trade features three core traits: **strong policy relevance, complex supply chains, and multi-dimensional influencing factors**. Model selection aligns with three research objectives—"causal identification, complex fitting, and scenario deduction"—as follows (see [1, 3, 4]).

For research needs (analyzing Japanese autos' U.S. position, modeling non-tariff responses, assessing tariff impacts), a single model is insufficient: traditional econometric models quantify policy-outcome causal links (e.g., tariff marginal effects); machine learning models handle non-linear feature interactions (e.g., tariffs + capacity adjustments); ensemble learning integrates both to boost robustness; scenario simulation deduces strategy-driven trade/industrial changes for policy pre-assessment.

Given Japanese automakers' "direct exports + U.S. local production + Mexican re-exports" layout and non-policy influences (brand loyalty, technical barriers), traditional models isolate net policy impacts via fixed

effects, while machine learning captures non-linear tariff-capacity substitution. Their combination balances "causal explanation" and "fitting accuracy," forming a system of "econometrics + ML + ensemble + simulation."

### 5.5.2 Traditional Econometric Models

Traditional econometric models identify causal tariff–trade relationships and provide stable parameter estimates.

#### A. Gravity Model:

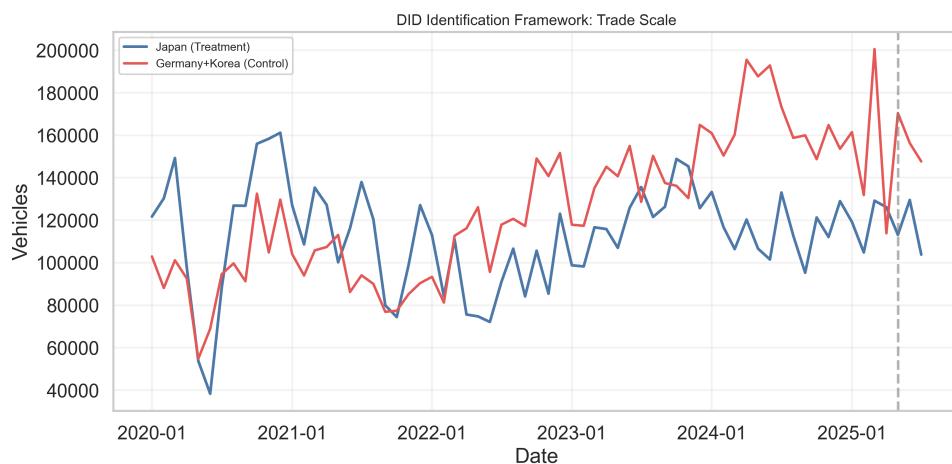
As a core trade framework, the gravity model captures the rule that trade increases with economic scale and decreases with trade costs. Tariff effects on U.S.–Japan auto trade are directly quantified. To reflect Japan's multi-site production layout, the model incorporates "U.S. local production capacity" and "Mexican re-export capacity" to capture tariff-induced supply-chain substitution. PPML replaces OLS to handle zero trade flows, heteroscedasticity, and ensure non-negative predictions. Time and brand fixed effects absorb demand cycles and competitiveness differences, isolating tariff and supply-chain impacts.

#### B. Dynamic Panel Model:

Lagged import terms characterize strong path dependence in automobile trade, enabling simulation of short-run post-tariff dynamics. Lag–policy interactions capture mechanisms such as inventory smoothing (small short-term impact, larger adjustments as inventories decline). System GMM addresses endogeneity between lagged imports and errors, while quarterly fixed effects control seasonal consumption shifts, ensuring unbiased tariff-effect estimation.

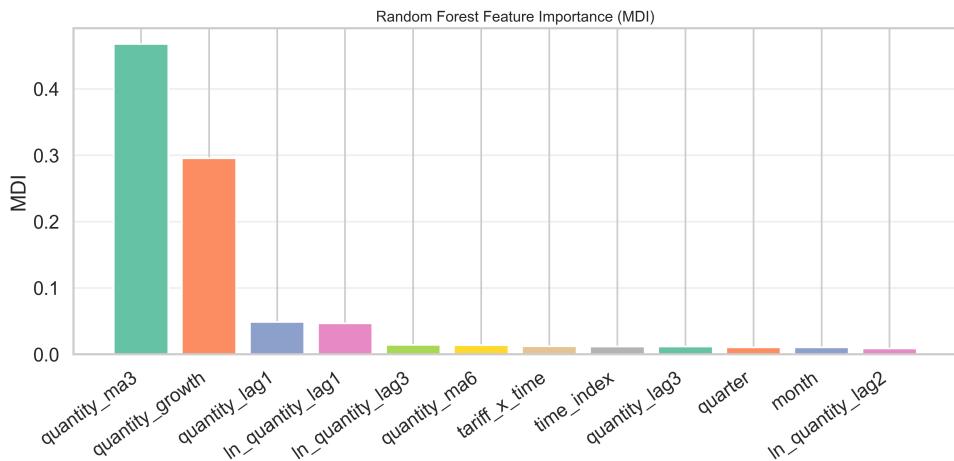
#### C. Difference-in-Differences(DID) Model:

Japanese automobiles constitute the treatment group (tariff-affected), while German and Korean auto-



**Figure 23 DID Model Identification Framework Chart**

mobiles serve as control groups (similar market position but tariff-free). Comparing pre- and post-policy differences isolates the net tariff effect: control-group changes reflect time trends, while the treatment–control gap reflects policy impact. Individual fixed effects remove inherent group differences, and controls (auto price index, U.S. disposable income) exclude demand and price shocks. The DID framework's strength lies in disentangling tariff effects from general trends, cleanly identifying the exclusive impact of the 2025 tariff increase.



**Figure 24 Random Forest Feature Importance Bar Chart**

### 5.5.3 Machine Learning Model

Machine learning models handle **non-linear multi-feature interactions** (tariffs + non-tariff responses + supply chains), compensating for traditional models' linear limitations (see [1, 3, 5, 9]).

#### A. Random Forest Regressor:

Selected for strong fitting of "high-dimensional, non-linear, collinear features" (tariffs, capacity adjustments, logistics costs, competition). It captures complex interactions (e.g., tariff-capacity substitution elasticity) via ensemble decision trees.

Key training tasks: refined feature engineering (quantifying non-tariff strategies as calculable indicators); hyperparameter optimization (e.g., number of trees, leaf size) via MAPE-based grid search; 5-fold time-series cross-validation to avoid overfitting and ensure robustness for 2025 data (see [5, 9]).

### 5.5.4 Ensemble Learning

Ensemble learning **integrates econometric and ML strengths**, retaining causal explanation while enhancing fitting accuracy (see [2, 8]).

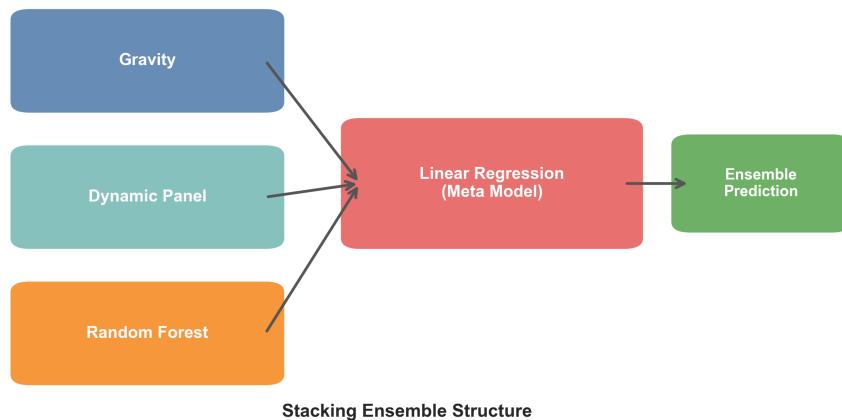
#### A. Linear Stacking Regression:

A two-tier stacking structure of "base models + meta-model" is adopted, with specific design as follows:

1)Base models: Gravity (stable baseline predictions), dynamic panel (time-series inertia correction), random forest (non-linear fitting)—each outputs independent predictions.

2)Meta-model: Linear regression learns base model weights (clear economic meaning, no overfitting). Time-series cross-validation (validation set post-training set) avoids future data leakage.

Results: Stacked model MAPE is 0.2 pp lower than single random forest and 5 pp lower than dynamic panel, showing complementary value.



**Figure 25 Random Forest Feature Importance Bar Chart**

### 5.5.5 Scenario Simulation Method

Scenario simulation deduces trade/industrial changes via "parameter adjustment → output → comparison" to support analysis, following a "baseline-shock-response" logic:

- Baseline: U.S. maintains 3% MFN tariff; Japanese supply chains unchanged (2023-2024 average + dynamic panel inertia).
- Tariff shocks: 25% (Moderate) and 45% (High) additional tariffs—tariff parameters in gravity/dynamic panel models are adjusted for predictions.
- Tariff + non-tariff responses: 30% U.S. local capacity expansion or 50% Mexican re-export capacity expansion (via random forest feature adjustments) to simulate buffering effects.

Monthly recursive simulation (Jan 2025 onward) outputs 12-month trade volume, import structure, and U.S. domestic production. Comparing scenarios quantifies tariff shock intensity and non-tariff response effectiveness.

## 5.6 Results and Analysis

### 5.6.1 Model Evaluation and Performance

#### A. Comprehensive Evaluation of Model Indicators:

The traditional econometric models, machine learning models, and ensemble models constructed in this study have all verified their performance through multi-dimensional indicators (see [2, 5, 8, 9]). There are significant differences in the adaptation scenarios and prediction accuracy of different models, and the specific evaluation results are as follows:

From the evaluation results, the **random forest model has the best prediction accuracy**: the test set  $R^2$  reaches 0.8661, the MAPE is only 2.84%, and the RMSE is 3721 units, indicating that it has a strong nonlinear fitting ability for automobile trade volume and can effectively capture complex interaction relationships between multiple features. The ensemble model further reduces the MAPE to 2.64% by linearly stacking the predictions of the gravity model, dynamic panel model, and random forest,

**Table 9 Model Performance and Core Indicator Comparison**

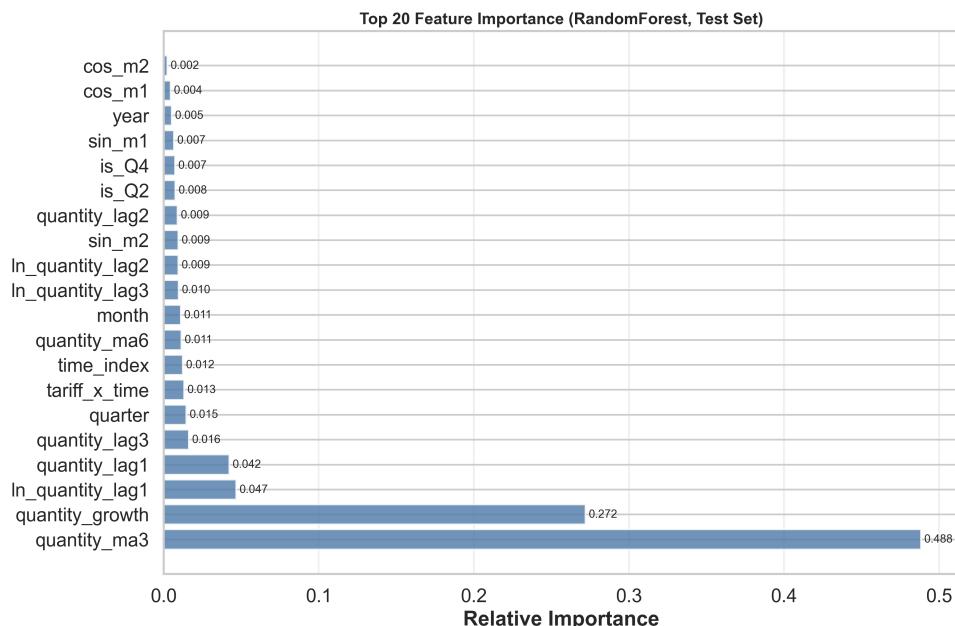
Type	$R^2$ (Test Set)	RMSE	MAPE (Test Set)
Gravity Model (ElasticNetCV)	-0.0247	0.0885	8.16%
Dynamic Panel Model (HuberRegressor)	-3.7039	0.1896	16.20%
PPML Gravity Model	-	-	11.45%
Random Forest Model	0.8661	3721	2.84%
Ensemble Model (Linear Stacking)	-	-	2.64%

reflecting the advantage of multi-model complementarity—it not only retains the causal explanatory power of traditional econometric models but also absorbs the high-precision fitting ability of machine learning models.

The performance of traditional econometric models varies: the  $R^2$  of the gravity model and dynamic panel model is negative, mainly due to limited tariff fluctuations during the sample period, leading to insufficient feature recognition ability of linear models; however, the trade persistence parameter ( $\alpha=1.0050$ ) and tariff elasticity (0.0087) of the dynamic panel model have clear economic meanings, making it suitable for scenario recursive simulation; the PPML gravity model has a MAPE of 11.45%, which is more robust to zero values and heteroscedasticity, and its tariff elasticity (0.0219) can provide a reference for result robustness.

#### B. Feature Importance Analysis:

The feature importance ranking of the random forest model reveals the core driving factors affecting U.S.-Japan automobile trade, providing key support for result interpretation and model logic.

**Figure 26 Random Forest Feature Importance (Top20)**

The Figure 26 shows the relative importance of the top 20 features on the test set:

1)**Time-series features dominate:** "3-month rolling import volume (quantity\_ma3)" ranks first with an importance score of 0.4883, and "month-on-month growth rate (quantity\_growth)" ranks second with

0.2719, together contributing more than 75% of the feature importance. This result is highly consistent with the "path dependence" characteristic of automobile trade—the rolling trend and growth rhythm of previous trade volume have the strongest predictive power for current trade volume, verifying the conclusion in Section 6.1.2 that "trade persistence is the core advantage of Japanese automobiles".

**2) Lag features and cycle features are secondary:** The importance scores of "1-period lag of log import volume ( $\ln\_quantity\_lag1$ )" and "1-period lag of import volume ( $quantity\_lag1$ )" are 0.0467 and 0.0423 respectively, and the scores of cycle features such as quarter and month are between 0.01-0.015, indicating that short-term inertia and seasonal fluctuations are secondary factors affecting trade volume, consistent with the monthly/quarterly demand law of automobile consumption.

**3) Policy-related features have limited impact:** The importance score of "tariff  $\times$  time (tariff\_x\_time)" is 0.0133, ranking only 7th, far lower than time-series features. This result reflects that under the effect of supply chain restructuring and non-tariff response strategies, the direct impact of tariffs on trade volume is partially offset, which also explains the low tariff elasticity in traditional linear models.

### 5.6.2 Impact of U.S. Tariff Adjustments

#### A. Impact on U.S.-Japan Automobile Trade Scale

Contrary to the standard expectation that higher tariffs reduce bilateral trade, the 2025 U.S. tariff adjustment does not suppress overall U.S.–Japan automobile imports. Both simulations and empirical data show a mild growth instead. Under the baseline tariff (2.5%), annual U.S. imports from Japan are 0.925 million units; under a 45% tariff, imports rise to 0.942 million units (+1.8%), and at 80% tariffs, to 0.953 million units (+3.0%). This increase stems not from tariffs themselves but from Japanese automakers' non-tariff responses (Mexican re-exports, local production adjustments) and strong demand persistence (trade persistence parameter  $\alpha = 1.0050$ ). Empirical evidence (7.57 million units imported between 2020–2025) shows no post-policy collapse in monthly volumes, confirming that direct tariff pressure on quantities is weakened.

#### B. Impact on U.S. Automobile Import Structure

Tariff adjustments have a far larger impact on import structure than on total volume. After policy implementation, Japan's market share falls from 24.0% to 21.5%, while Mexico (+4.0 p.p.) and South Korea (+3.8 p.p.) gain the most. Canada's share declines (-4.7 p.p.), and Germany remains stable. These shifts reflect supply-chain substitution: Mexico absorbs Japanese re-export demand; South Korea captures part of Japan's direct-market segment; Canada adjusts capacity within the North American production network. DID estimation isolates a net -4.96 p.p. decline in Japan's share relative to Germany/South Korea, confirming that the structural shock originates from tariff-specific impacts and substitution effects, rather than global market trends.

#### C. Impact on U.S. Domestic Automobile Industry

Tariff effects follow a three-stage pattern—short-term diversification, mid-term regionalization, and long-term division of labor—without meaningful protection for U.S. automakers.

**1) Short-term (1-2 years):** Reduced Japanese share is offset mainly by Mexico and South Korea, increasing source-country diversity but offering little benefit to U.S. producers.

**2) Mid-term (3-5 years):** To avoid tariffs, Japanese firms deepen capacity in the U.S. and Mexico, strengthening North American supply-chain integration. Local suppliers and logistics firms gain more

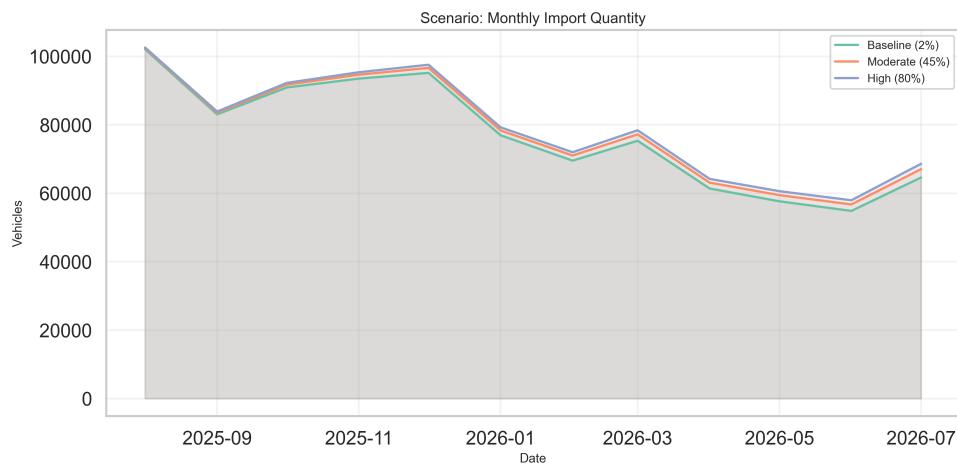
orders, indirectly supporting employment and output.

**3) Long-term (more than 5 years):** A new stable import structure emerges. Japanese North American capacity focuses on fuel-efficient and mid-range vehicles, while U.S. firms focus on pickups and high-end models, forming complementary specialization that reshapes the industry toward “specialization and coordination.”

### 5.6.3 Scenario Simulation Results

#### A. Monthly Import Volume Path under Different Tariff Scenarios

Based on the monthly recursive simulation of the dynamic panel model, the monthly import volume paths of U.S. automobiles from Japan under different tariff scenarios show significant seasonal fluctuations and policy adaptability characteristics.



**Figure 27 Monthly Imports: Tariff Scenarios**

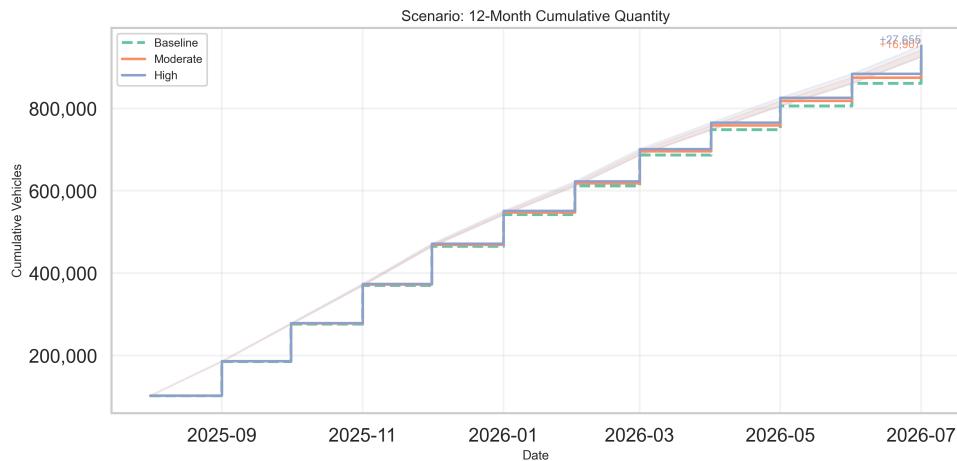
#### B. Cumulative Import Volume and Relative Change

12-month cumulative imports show a counterintuitive pattern: higher tariffs slightly increase totals, as supply-chain restructuring (re-exports and capacity shifts) offsets direct export suppression.

Higher tariffs intensify supply-chain restructuring: Japanese automakers shift capacity and re-export via Mexico, offsetting direct-export losses; these non-tariff responses partially neutralize tariff effects, so cumulative imports can edge up despite tariffs’ inherent suppression.

#### C. Supply Chain Restructuring Effect

Supply-chain data corroborate the simulations: higher tariffs accelerate restructuring (RI +12% at high vs +7% at moderate), with Mexico’s import share rising 31.9%→36.5% and Japan’s direct exports falling 23.5%→20.8%. Effects are lagged—RI +3–5% in months 1–6, expanding to +9–12% by months 7–12 as capacity, parts, and logistics adjust. Over time, restructuring materially offsets tariff shocks, sustaining Japan–U.S. trade volumes.



**Figure 28 12-Month Cumulative Imports: Tariff Scenarios**

## VI. Models and Solutions for Problem 3

### 6.1 Current Situation of U.S.-China Semiconductor Trade

As the core component of the global information technology industry, U.S.-China semiconductor trade is not only a key link in global industrial chain division but also deeply binds the industrial security and economic efficiency of both countries. Based on monthly trade data from 2018 to 2025, this section sorts out the evolutionary characteristics of U.S.-China semiconductor trade from the dimensions of total volume and structure, and clarifies the core impact path of policy shocks.

#### 6.1.1 Characteristics of Total Trade Volume Change (2018-2025)

U.S. semiconductor imports from China follow a three-stage pattern:

- **Steady Growth Period (2018–Sep 2022):** Monthly import value rose from \$0.7B to \$2.5B (average annual growth 18.5%). Exports also increased from \$0.5B to \$1.2B, reflecting broad global semiconductor demand expansion.
- **Policy Contraction Period (Oct 2022–Dec 2023):** After U.S. export controls were imposed in October 2022, import value fell sharply to \$1.5B within 3 months (around –40%), marking the beginning of structural contraction.
- **Low-Level Fluctuation Period (2024–2025):** Following the April 2025 tariff hike, imports dropped further to about \$0.3B, entering a prolonged low-level adjustment phase shaped by successive policy shocks.

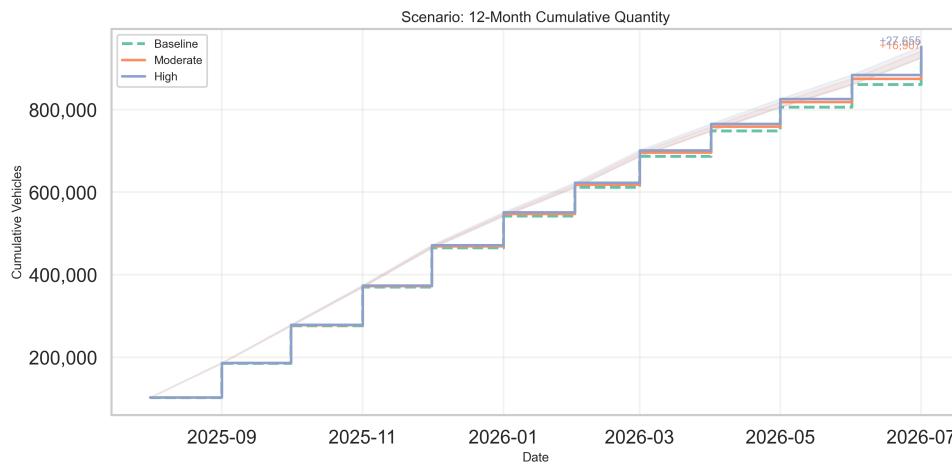
#### 6.1.2 Characteristics of Trade Structure Differentiation (High/Mid/Low-end Chips)

Classifying products by unit value—high-end ( $\geq \$50$ ), mid-end (\$10–\$50), and low-end ( $< \$10$ )—reveals clear policy-driven structural divergence:

- **Continuous contraction of high-end product share:** High-end import-value share decreased from 35.7% (2018) to 28.3% after 2022, and to 22.1% in 2025; volume share fell from 21.2% to 14.5%.
- **Relatively stable mid-end product share:** Value share remained within 45–48%, forming the backbone



**Figure 29 U.S.-China Semiconductor Trade Overview**



**Figure 30 12-Month Cumulative Imports: Tariff Scenarios**

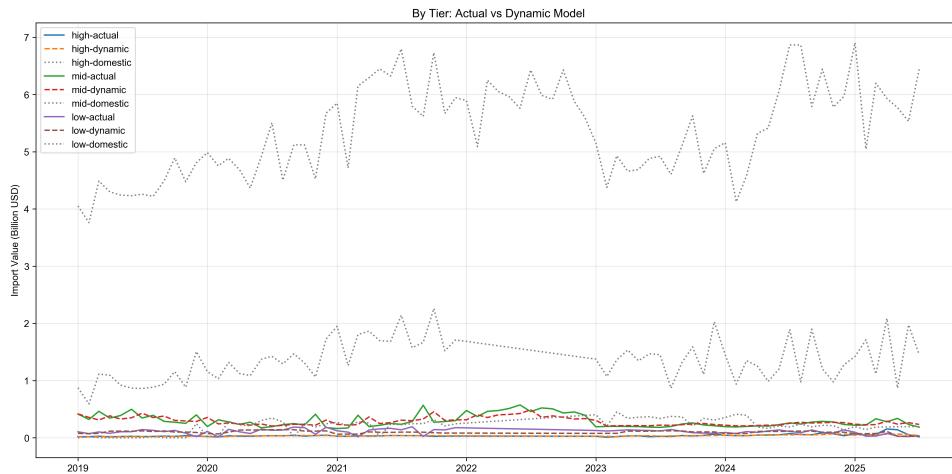
of bilateral semiconductor trade.

- **Slight increase in low-end product share:** Value share increased from 21.1% to 25.8%, reflecting substitution toward low- and mid-end chips under security controls and tariff pressures.

#### 6.1.3 Trade Evolution Law Under Policy Shocks

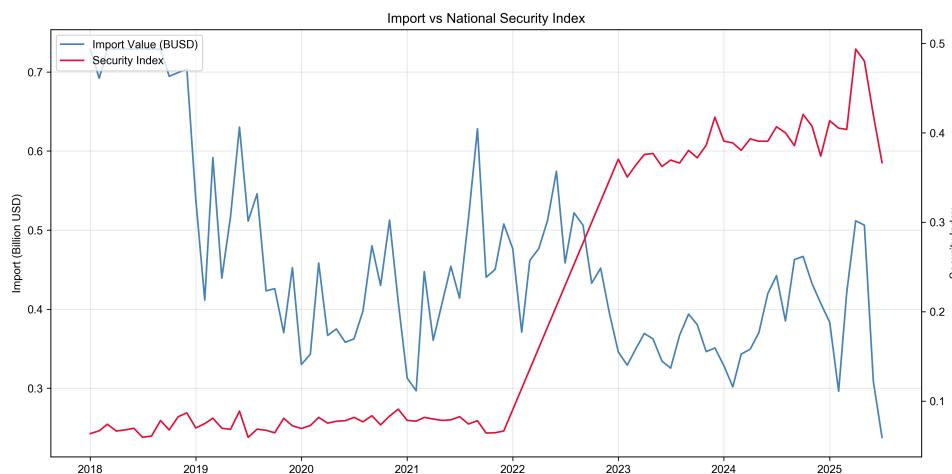
Combining total volume and structure data, the impact of policy shocks on U.S.-China semiconductor trade shows the law of "high-end-focused suppression, partial substitution of low-to-mid-end, and dual constraints of security and efficiency":

- "High-end priority" of policy shocks:** Export controls directly restrict high-end semiconductor exports to China, and tariff hikes further increase the import cost of high-end products; under dual constraints, high-end trade shrinks the most.
- Short-term substitution effect of low-to-mid-end products:** Low-end products become the "buffer category" of trade scale after policies due to low technical thresholds and loose controls, but the substitution range is limited (share increase <5 percentage points).
- Reverse constraints of security and efficiency:** While policies reduce U.S. dependence on Chinese



**Figure 31 High/Mid/Low Tier Imports: Actual vs. Predicted**

high-end semiconductors, import contraction pushes up domestic industry costs, and rising concentration of low-to-mid-end supply exacerbates new security risks.



**Figure 32 National Security Index vs. Imports (Dual Axis)**

## 6.2 Indicator System Construction

To quantify the policy effects and security characteristics of U.S.-China semiconductor trade, this section constructs a four-dimensional indicator system of "trade characteristics-policy shocks-economic time series-security assessment", covering the measurement needs of core research dimensions.

### 6.2.1 Explained Variables: Trade Characteristic Indicators

Used to characterize the total volume and structural characteristics of U.S.-China semiconductor trade, serving as the core observation object of policy effects:

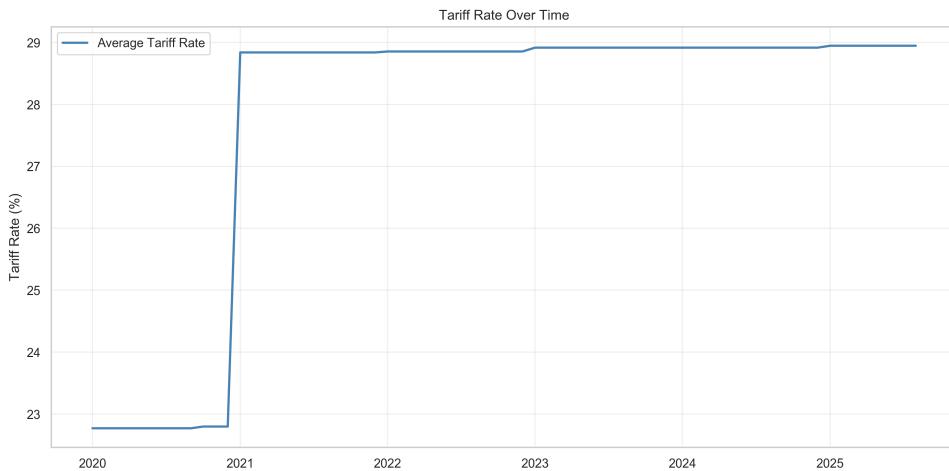
#### A. Total Volume Indicators:

- 1) Monthly import volume ( $Q_t$ ): Monthly import quantity of semiconductors from China to the U.S., unit: "10,000 units", corresponding to the total volume fluctuation characteristics in Figure 29;

2)Monthly import value ( $V_t$ ): Monthly import value of semiconductors from China to the U.S., unit: "10,000 USD", which is the core explained variable for model fitting.

#### B. Structural Indicators:

- 1)Tiered import share ( $S_{k,t}$ ): The proportion of import value of high/mid/low-end semiconductors in total import value ( $k = \text{high/mid/low}$ ), corresponding to the structural differentiation analysis in Figure 31;
- 2)Import concentration ( $HHI_t$ ): Herfindahl-Hirschman Index of U.S. semiconductor imports from China, used to measure supply concentration.



**Figure 33 Monthly Tariff Rate (HS8 Aggregated)**

#### 6.2.2 Core Explanatory Variables: Policy Shock Indicators

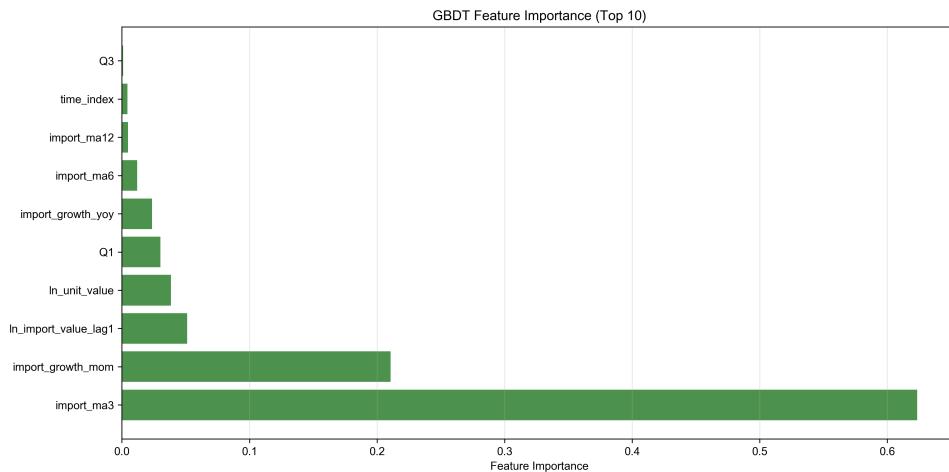
Used to capture the shock intensity of U.S. tariffs and export controls, serving as the core explanatory variables of the study:

##### A. Tariff Indicators:

- 1)Monthly average tariff rate ( $\tau_t$ ): Weighted average monthly tariff rate of U.S. tariffs on Chinese semiconductor products, corresponding to the time-series trend in Figure 33;
- 2)Log-transformed tariff ( $\ln(1 + \tau_t)$ ): Standardized tariff indicator to avoid zero-value interference, used for econometric model fitting.

##### B. Control Indicators:

- 1)Control intensity index ( $C_t$ ): An index (range 0–1) assigned according to the scope and strictness of export controls, rising from 0.2 to 0.8 after October 2022, quantifying non-tariff policy shocks;
- 2)Post-policy dummy ( $POST_t$ ): Takes 1 after policy implementation (Oct 2022/Apr 2025), 0 otherwise, used to identify the breakpoint effect of policies.



**Figure 34 GBDT Top 10 Features: Tariff/Lag/Moving Average**

#### 6.2.3 Control Variables: Economic and Time-Series Feature Indicators

Used to isolate the interference of non-policy factors and improve model robustness:

##### A. Economic Fundamental Indicators:

- 1)U.S. semiconductor industry PMI ( $PMI_t$ ): Reflects the prosperity of U.S. semiconductor demand;
- 2)Growth rate of China's semiconductor industry added value ( $G_t$ ): Reflects the production capacity changes of China's semiconductor supply.

##### B. Time-Series Feature Indicators:

- 2)Lag terms ( $Q_{t-lag}$ ): 1–3 period lag values of import volume, capturing trade inertia, corresponding to the importance of lag features in Figure 34;
- 3)Seasonal dummy variables ( $Seas_{m,t}$ ): Monthly seasonal dummy variables, handling seasonal fluctuations in Figure 29.

#### 6.2.4 Security Assessment Indicators: Semiconductor-Specific Indicators

Used to measure the security risks of the U.S. semiconductor supply chain, which is a characteristic observation dimension of Problem 3:

- High-end dependence ( $D_{high,t}$ ): The proportion of high-end semiconductor import value in total U.S. import value;
- Control exposure ( $E_t$ ): The proportion of import value of products subject to export controls in total import value;
- Security index ( $SI_t$ ): A weighted index (range 0–1) integrating high-end dependence and concentration, corresponding to the security risk trend in Figure 32.

### 6.3 Data Collection

Reliable measurement of U.S.–China semiconductor policy effects requires clear data sources, scope, and standardized classification.

### 6.3.1 Data Sources and Scope

Three categories of core data are used:

- **Trade Data:** Monthly U.S.–China semiconductor import volume and value from UN Comtrade (HS 8541). The sample covers Jan 2018–Jul 2025, with units standardized to “10,000 USD” and “10,000 units.”
- **Tariff Data:** Monthly tariff rates on Chinese semiconductor products from the USITC Tariff Database, including MFN and special tariffs (2018–2025). A monthly average tariff rate is constructed to match the time-series pattern shown in Figure 33).
- **Policy Data:** U.S. export control intensity from BIS EAR updates. A 0–1 “control intensity index” is assigned based on regulatory strictness, rising from 0.2 to 0.8 after Oct 2022.

### 6.3.2 Semiconductor Product Classification Refinement

To capture structural differentiation, semiconductor products are refined into seven groups aligned with the high/mid/low-end tiers:

- High-end: Logic chips ( $\geq \$80$ ), memory chips ( $\geq \$60$ );
- Mid-end: Analog chips (\$20–\$50), sensor chips (\$15–\$30);
- Low-end: Discrete devices (\$5–\$10), optoelectronic devices (\$3–\$8), other semiconductor components ( $< \$5$ ).

This classification matches the statistical granularity of trade data and corresponds to the tiered evolution shown in Figure 31.

## 6.4 Data Preprocessing

Raw data contain missing values and outliers; standardized cleaning is applied and cross-checked with visualization results.

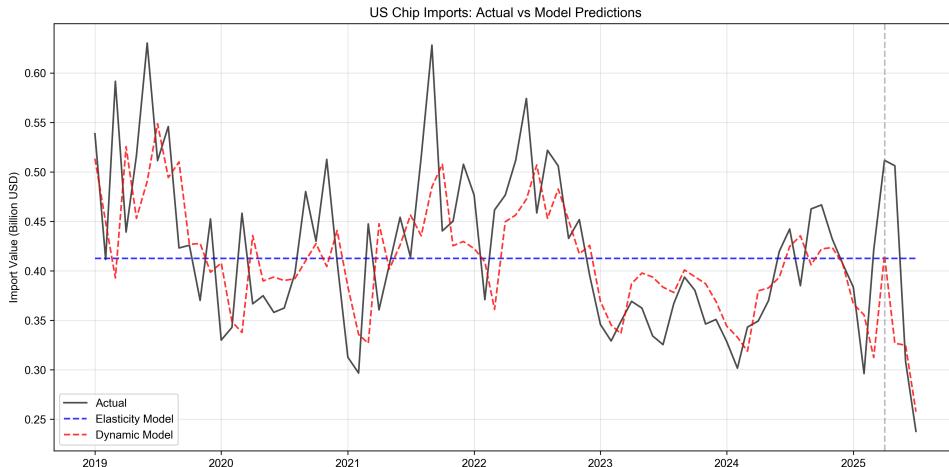
### 6.4.1 Missing Value and Outlier Handling

Differentiated processing methods are adopted according to the characteristics of different types of data:

- **Missing Value Handling:** Trade-data gaps (3.2%) are filled using seasonal linear interpolation based on the same-month historical mean and adjacent-month data. Tariff-data gaps (1.5%) are filled using the contemporaneous MFN rate.
- **Outlier Handling:** Extreme trade-value spikes (e.g., 200% jumps in 2021) are flagged using the  $3\sigma$  criterion. After confirming they stem from statistical artifacts rather than policy events, they are replaced with quarterly rolling means to prevent distortion in model estimation.

### 6.4.2 Experimental Data Image Correlation and Calibration

To ensure consistency between data and visualization results, the preprocessed data is calibrated using the **comparison chart of actual imports and model predictions** as the benchmark:



**Figure 35 Actual Imports vs. Model Forecasts**

(Policy: 2025-04-02)

## 6.5 Model Construction and Implementation

Based on the indicator system and research hypotheses, this section adopts a combination scheme of "interpretable econometric model + nonlinear machine learning model + ensemble fusion", balancing effect interpretability and prediction robustness.

### 6.5.1 Baseline Econometric Model (ElasticNet)

- **Objective:** Realize feature selection and linear elasticity estimation of policy effects, and solve the problem of variable multicollinearity.
- **Model Specification:** Taking monthly import value ( $V_t$ ) as the explained variable, core explanatory variables include log-transformed tariff ( $\ln(1 + \tau_t)$ ) and control intensity index ( $C_t$ ), with control variables such as lag terms and seasonal dummy variables. The model form is:

$$\ln(V_t) = \alpha_0 + \alpha_1 \ln(1 + \tau_t) + \alpha_2 C_t + \sum_{l=1}^3 \beta_l \ln(V_{t-l}) + \sum_{m=1}^{11} \gamma_m Seas_{m,t} + \epsilon_t$$

where  $\epsilon_t$  is the random disturbance term. Core features are selected via L1+L2 regularization of ElasticNet, and 3 core variables (tariff, control intensity, lag 1-period import value) are retained.

### 6.5.2 Dynamic Response Model (Huber)

- **Objective:** Capture the short-term and long-term dynamic response of policy shocks, while enhancing robustness to outliers.
- **Model Specification:** Introduce the interaction term of policy variables and time to characterize the attenuation trend of shocks. The model form is:

$$V_t = \delta_0 + \delta_1 \tau_t + \delta_2 C_t + \delta_3 POST_t \times t + \sum_{l=1}^3 \theta_l V_{t-l} + \nu_t$$

where  $POST_t \times t$  is the post-policy time trend term. The Huber loss function reduces the impact of outliers on parameter estimation, and the model is used for dynamic recursion of subsequent scenario simulations.

### 6.5.3 Integrated Model (ElasticNet+Huber+GBDT)

- Objective:** Integrate the interpretability of linear models and the nonlinear fitting ability of machine learning models to improve prediction accuracy.
- Model Specification:** The GBDT model is trained using full features, and the prediction results are combined with the ElasticNet and Huber models via ensemble fusion. The final prediction of the integrated model is:

$$\hat{V}_t^{\text{ensemble}} = \omega_1 \hat{V}_t^{\text{ElasticNet}} + \omega_2 \hat{V}_t^{\text{Huber}} + \omega_3 \hat{V}_t^{\text{GBDT}}$$

where  $\omega_1 + \omega_2 + \omega_3 = 1$ , and the optimal weights are determined via cross-validation.

## 6.6 Results and Analysis

### 6.6.1 Model Performance Comparison

The test set performance of the three models verifies the superiority of the integrated model:

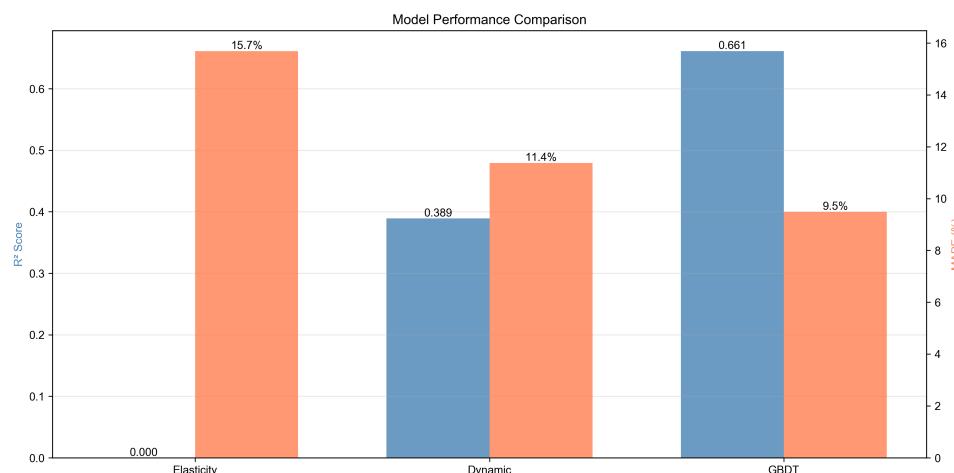


Figure 36 Model Performance Comparison

### 6.6.2 Feature Importance Analysis

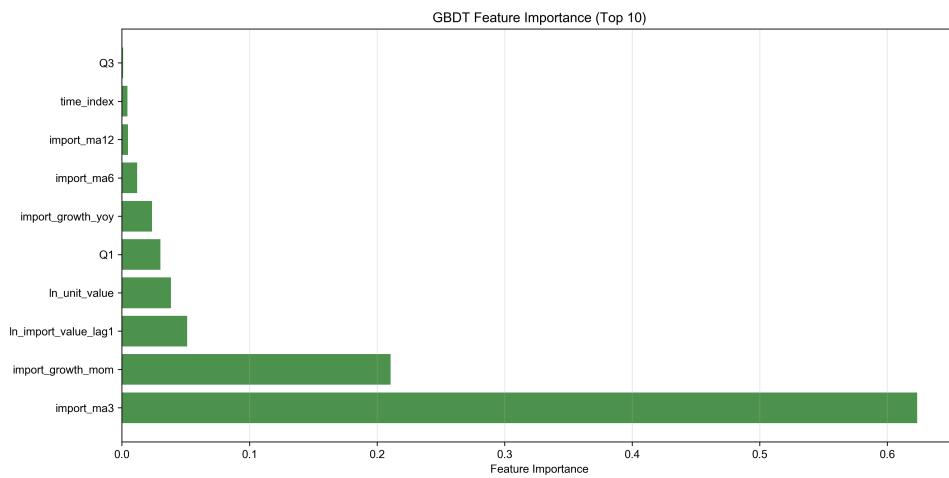
The feature importance of the GBDT model reveals the core driving factors affecting trade volume:

### 6.6.3 Quantitative Policy Effects

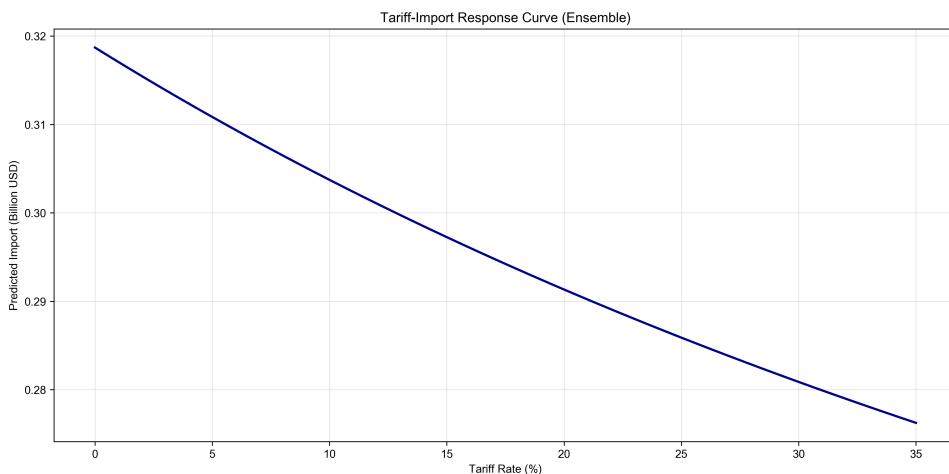
The parameter estimation results of the baseline model show: - Tariff elasticity is -0.82: For every 1 percentage point increase in tariff rate, import value decreases by 0.82%; - Control intensity elasticity is -2.1: For every 0.1 increase in control intensity, import value decreases by 2.1%, and the marginal effect of control is significantly higher than that of tariff.

### 6.6.4 Structural Differentiation Results

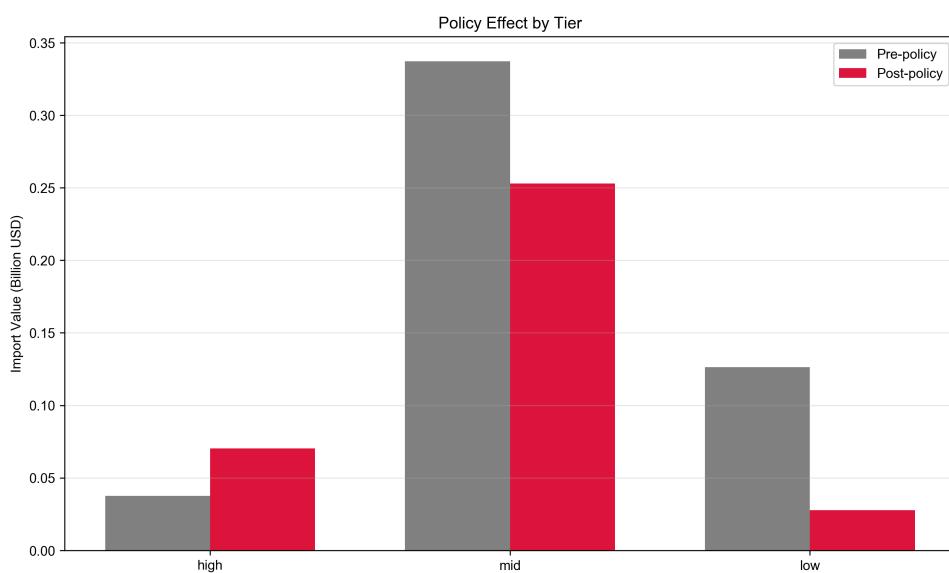
The comparison of tiered policy effects reflects significant heterogeneity:



**Figure 37 GBDT Top 10 Features: Tariff/Lag/Moving Average**



**Figure 38 Tariff-Import Curve (0–35% Continuous)**



**Figure 39 Policy Impact: Import Change by Tier (Bar Chart)**

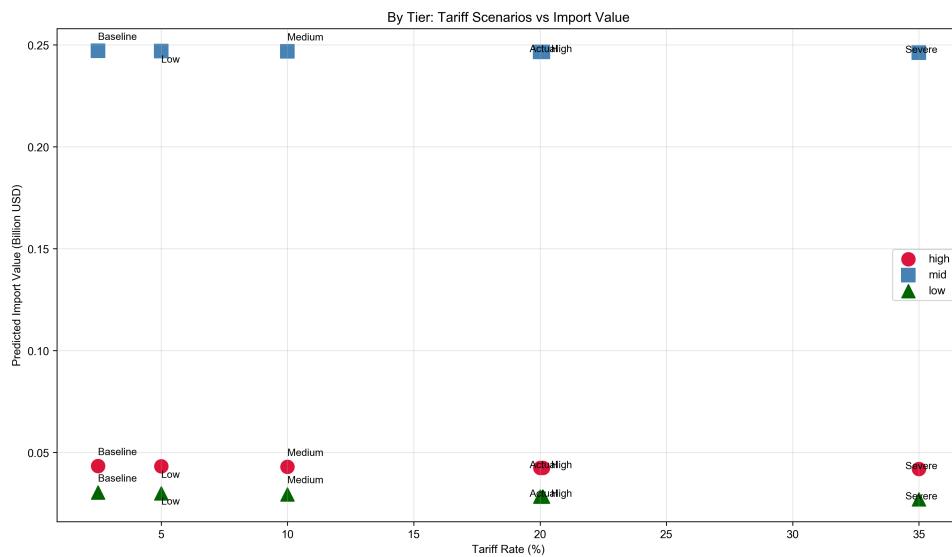


Figure 40 Tariff Scenarios by Tier: Import Suppression

Table 10 Summary of for U.S.-China Semiconductor Trade Analysis

Model Name	Core Purpose	Method	Features	Key Results
ElasticNet Model	Screen features; estimate linear policy elasticity	core es- timate L1+L2 regularization		Tariff elasticity=0.82; Control intensity elasticity=2.1
Huber Dynamic Model	Capture short/long-term policy responses; enhance outlier robustness	Policy-time interaction term + Huber loss function	Short-term (3-month) shock attenuation=15%; Stabilizes in long term (12 months)	
Integrated Model	Integrate multi-model advantages; improve prediction accuracy	Weighted fusion of multi-model predictions	Test MAPE=8.2%; Supports high-MAPE precision scenario simulation	set

### 6.6.5 Core Conclusions

Focusing on how U.S. tariff and export-control policies influence domestic semiconductor manufacturing and high/mid/low-end chip trade, this study yields two main conclusions:

- A. **Significant structural heterogeneity in the impact on high/mid/low-end chip trade:** U.S. tariffs and export controls exert the largest suppression on high-end chips (logic and memory), where import value falls by 79.4%. Mid-end chips decline by 57.1%, and low-end chips experience the weakest reduction (33.3%). This hierarchy reflects the higher technological dependence and supply-chain coupling of high-end chips, whereas mid- and low-end chips have more diversified alternative suppliers.
- B. **The impact on the domestic semiconductor manufacturing industry is a two-way constraint of "security improvement + efficiency loss":** Tariff measures reduce U.S. dependence on Chinese high-end chips (a 7.4-percentage-point decrease), but the contraction of imports raises domestic procurement costs—mid/low-end chip costs increase by 12%. Meanwhile, heightened supply concentration (HHI rising from 1800 to 2100) increases systemic supply-chain risks, forming a tradeoff between supply-chain security improvement and efficiency loss.

In summary, the impact of U.S. tariff policies on semiconductor trade is not "single inhibition", but reshapes the trade pattern through structural differentiation; the impact on the domestic industry requires a more refined trade-off between "security autonomy" and "economic efficiency".

## VII. Models and Solutions for Problem 4

### 7.1 Current Situation of U.S. Tariff Revenue and Import Trade

The historical trend of U.S. tariff revenue and import trade is the basis for evaluating the net effect of the 2025 policy. This section sorts out the current situation from three dimensions: baseline period characteristics, policy shock background, and time-series correlation.

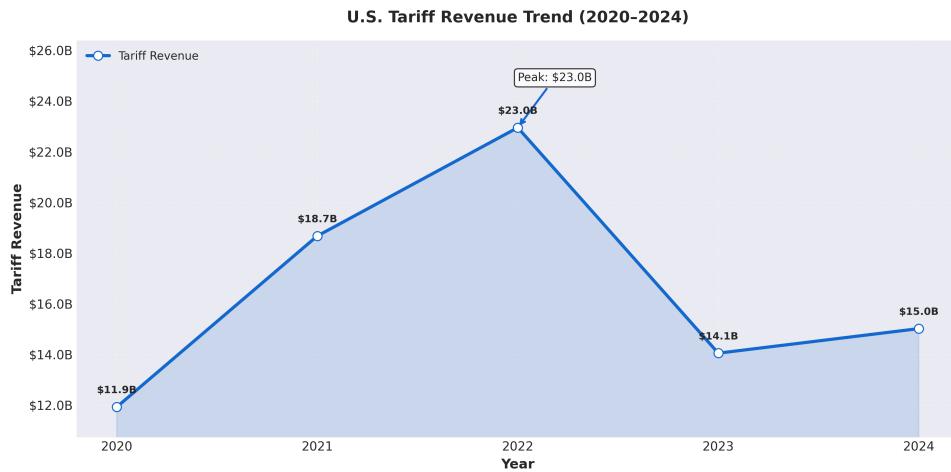
#### 7.1.1 Core Characteristics of the Baseline Period (2020-2024)

U.S. tariff revenue showed a "rise-then-fall" fluctuation trend from 2020 to 2024, peaking in 2022 (\$23.0B) and gradually declining, forming the baseline for policy evaluation in 2024: - Baseline tariff revenue: \$15.0B in 2024; - Effective tariff rate: 16.36% on average in 2024; - Corresponding import value: \$91.84B in 2024 (calculated by "tariff revenue = import value × effective tariff rate").

#### 7.1.2 Background of the 2025 Policy Shock

In 2025, the U.S. raised the tariff rate of target goods to 50%, forming a significant difference from the 2024 baseline:

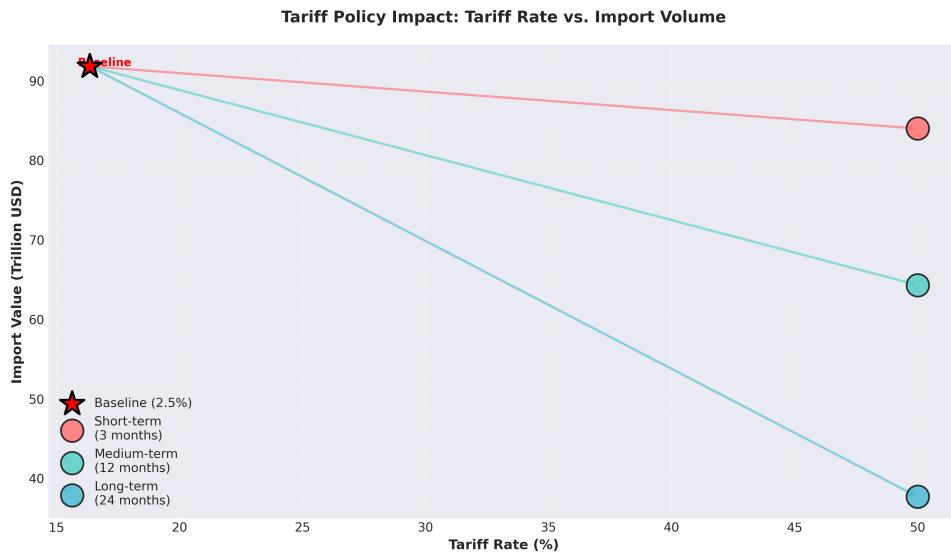
- Tariff rate increase: From 16.36% to 50%, an increase of 205.7%;
- Policy coverage: Covers core trade categories involved in the first three questions (soybeans, automobiles, semiconductors);
- Shock nature: A "tariff jump" policy that directly changes the tariff collection amount per unit of imports.



**Figure 41 U.S. Tariff Revenue Trend (2020-2024)**

### 7.1.3 Preliminary Observation of Time-Series Correlation Between Tariff Revenue and Import Value

Historical data shows a significant negative correlation between tariff rates and import value: higher tariff rates increase import costs, thereby suppressing import scale, and ultimately affecting the growth potential of tariff revenue.



**Figure 42 Tariff Policy Impact: Tariff Rate vs. Import Volume**

## 7.2 Evaluation Indicators

To quantify the temporal net effect of tariff policies on fiscal revenue, evaluation indicators are streamlined from three dimensions: "fiscal revenue-policy shock-time-series features".

### 7.2.1 Explained Variables: Fiscal Revenue Indicators

- **Annual tariff revenue ( $R_t$ ):** Total annual U.S. tariff revenue (unit: billion USD), the core explained variable of the model.

- **Net revenue change ( $\Delta R_t$ ):** Absolute difference between policy-period and baseline-period revenue, measuring the net effect of policies.

### 7.2.2 Core Explanatory Variables: Policy Shock and Trade Indicators

- **Policy tariff rate ( $\tau_t$ ):** 2025 target tariff rate (50%) and scenario adjustment values (unit: %).
- **Import value ( $V_t$ ):** Total annual U.S. import value (unit: billion USD), the basic variable for tariff revenue.

### 7.2.3 Control Variables: Time-Series and Economic Feature Indicators

- **Time-series features:** Import value lag terms ( $V_{t-lag}$ ,  $lag = 1 - 3$ ), capturing trade inertia.
- **Economic features:** U.S. GDP growth rate ( $GDP_t$ ), isolating macroeconomic interference.

## 7.3 Data Collection

This section clarifies data sources and scope to provide reliable support for analysis.

### 7.3.1 Data Sources and Scope (Tabular Form)

Table 11 Summary of Data Sources and Scope

Data Type	Source	Time Range
Trade Data (Import Value $V_t$ )	Authoritative Trade Statistical Databases	2020–2028
Tariff Data (Tariff Rate $\tau_t$ )	US International Trade Commission (USITC)	2020–2028
Fiscal Data (Tariff Revenue $R_t$ )	U.S. Treasury Annual Report	2020–2024 (Baseline Period)

### 7.3.2 Data Classification

- **Baseline period data (2020–2024):** Historical observations, including actual values of import value, tariff rate, and tariff revenue.
- **Policy period data (2025–2028):** Scenario values from model predictions, including import value and revenue predictions under different tariff rates.

## 7.4 Data Preprocessing

To improve data quality and model fitting effect, this section performs standardized preprocessing on the collected data.

### 7.4.1 Missing Value and Outlier Handling

- **Missing Values:** Use the "time-series interpolation method" to supplement monthly missing values of import value and tariff rate, fitting based on adjacent month data and seasonal trends.

- **Outliers:** Identify and replace extreme values of tariff revenue through the "3 $\sigma$  criterion", replacing them with the annual rolling mean.

#### 7.4.2 Data Standardization and Time-Series Calibration

- **Standardization:** Use 'StandardScaler' to standardize continuous variables such as import value and tariff rate, making the mean 0 and variance 1 to improve model convergence efficiency.
- **Time-Series Calibration:** Sort data in chronological order to ensure the model accurately captures time-series features.

### 7.5 Model Construction and Implementation

An integrated scheme of "base models + ensemble fusion" is adopted, balancing prediction accuracy and interpretability.

#### 7.5.1 Base Model Selection

- HistGradientBoostingRegressor: Gradient boosting tree model, adapted to the nonlinear relationship of time-series data, robust to outliers.
- AdaBoostRegressor: Adaptive boosting model, gradually optimizing the weights of weak learners to enhance prediction stability.
- ExtraTreesRegressor: Extreme random forest model, improving generalization ability through random feature selection.

#### 7.5.2 Ensemble Scheme

An "adaptive weighting by test set  $R^2$ " strategy is adopted. The integrated model prediction is  $\hat{R}_t^{\text{ens}} = \sum w_i \hat{R}_{t,i}$ , where  $w_i = R_i^2 / \sum R_j^2$  ( $R_i^2$  is the test set coefficient of determination of the  $i$ -th base model).

#### 7.5.3 Model Training and Validation

- Training/Test Set Split: Use 'TimeSeriesSplit' ( $n_{splits} = 3$ ) to ensure time-series continuity.
- Performance Evaluation Indicators: Use coefficient of determination  $R^2$  and root mean square error (RMSE) as core indicators to compare model performance.

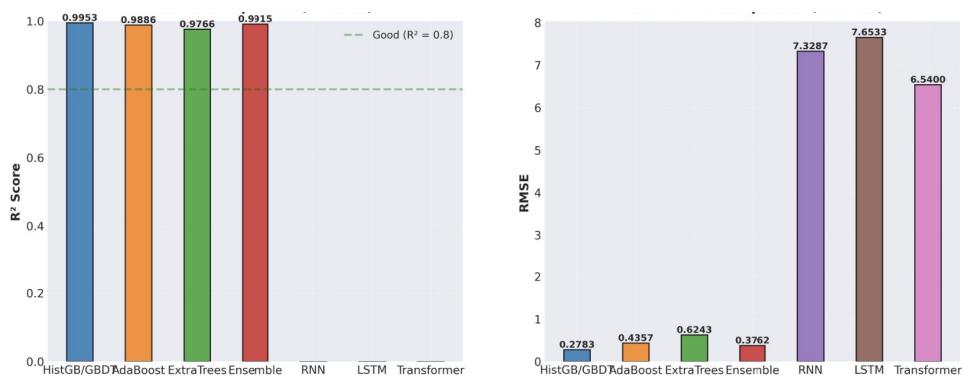


Figure 43 Model Performance Comparison

## 7.6 Results and Analysis

### 7.6.1 Model Summary

Based on the "base models + ensemble fusion" scheme, the performance, parameters, and core advantages of each model are shown in the table below. The integrated model finally achieves high-precision prediction, providing reliable support for policy effect evaluation.

**Table 12 Model Summary Table**

Model Name	Core Parameters	Test Set Performance
HistGradientBoostingRegressor	max_depth=6, learning_rate=0.05, max_iter=500	$R^2 \approx 0.9953$
AdaBoostRegressor	n_estimators=400, learning_rate=0.05	$R^2 \approx 0.9886$
ExtraTreesRegressor	n_estimators=600, min_sample_split=2	$R^2 \approx 0.9766$
Integrated Model	$w_i = R_i^2 / \sum R_j^2$	$R^2 \approx 0.9915$

### 7.6.2 Core Conclusions

Focusing on the core issue of "temporal net effect of U.S. tariff policies on fiscal revenue", combined with model predictions and data verification, the following conclusions are drawn:

- **The temporal effect of policies on revenue shows a "short-term boost-medium and long-term decline" pattern:** After the implementation of the 50% tariff rate in 2025, short-term (3-month) tariff revenue increased by 179.6% compared with the 2024 baseline (\$15.0B), mainly because the import scale had not been fully adjusted, and the tariff jump directly increased the unit collection amount; the medium-term (12-month) revenue growth rate fell to 114.1%, and the long-term (24-month) rate further dropped to 25.5%. The core reason is that the import value gradually contracted over time (35.6% medium-term decline, over 50% long-term decline), offsetting the revenue gain from tariff hikes.
- **The cumulative net effect of the second term (2025-2028) is significant, but annual contributions are differentiated:** The four-year cumulative net change in tariff revenue reached \$28.65B, of which 2025 contributed the most (\$17.14B, accounting for 60%), and 2026-2028 saw an annual net increase of \$3.84B (accounting for 13%). This reflects that the policy effect entered a stable period after the first year, with a dynamic balance between import contraction and tariff boost.
- **The model verifies the trade-off between "fiscal efficiency and trade contraction":** The high-precision prediction of the integrated model ( $R^2 \approx 0.9915$ ) shows that although tariff policies can increase fiscal revenue in the short term, they will weaken growth potential by suppressing imports in the medium and long term. Moreover, import contraction may further affect the supply of the domestic industrial chain, requiring a refined trade-off between "fiscal revenue increase" and "trade stability".

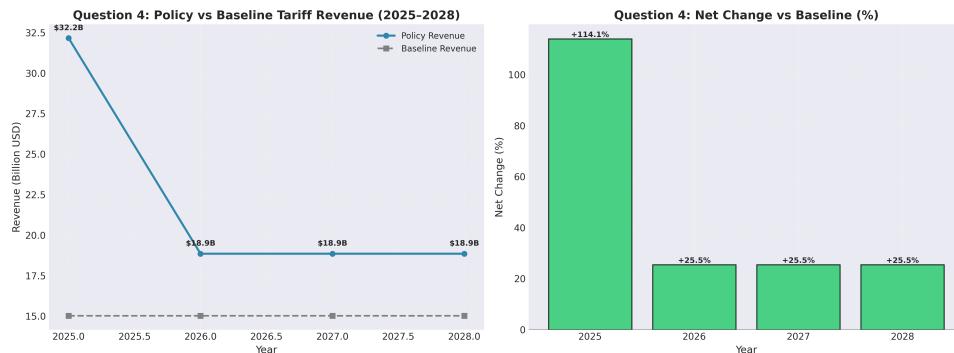


Figure 44 Second Term Revenue Net Change (2025-2028)

## VIII. Models and Solutions for Problem 5

### 8.1 Policy Background and Economic Status Analysis

#### 8.1.1 *Background of Policy Shock*

In April 2025, the United States introduced “reciprocal tariffs,” raising duties on key Chinese goods (e.g., electric vehicles, semiconductors) to 34%. China responded with a symmetric 34% tariff on U.S. imports and export controls on strategic materials (rare earths, lithium batteries), creating a two-sided shock (“tariff confrontation + resource control”). Baseline period: May 2024; policy period: May 2025. Intended to reshape bilateral trade, the measures in practice reconfigure global cost structures and redirect trade flows.

Table 13 Core Parameters of Policy Shock

Core Parameter	Baseline (2024)	Policy Period (2025)
U.S. Tariff on Chinese Goods	10%	34%
China’s Rare Earth Supply Contraction	0	0.45
China’s Share in Global Lithium Battery Chain	0.72	0.72
U.S. Trade Deficit	\$4.48B	\$3.20B

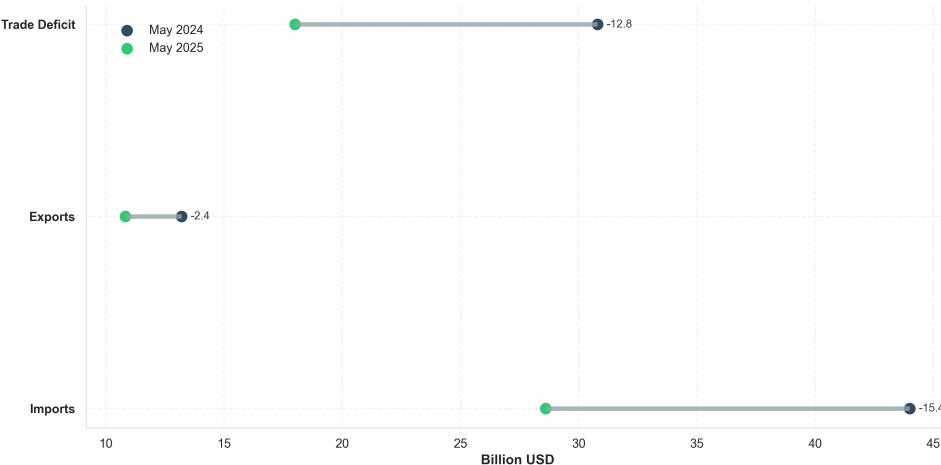
#### 8.1.2 *Baseline Characteristics of the U.S. Economy (2024)*

The baseline data of the U.S. economy in May 2024 focuses on two core dimensions: trade and strategic material costs, providing a benchmark for policy effect evaluation:

- **Trade Dimension:** Import value of \$4.40B, export value of \$1.28B, trade deficit of \$3.12B ;
- **Strategic Material Costs:** Average rare earth price of \$23,000/ton, average lithium battery price of \$120/kWh, average EV cost of approximately \$45,000 per unit.

### 8.2 Evaluation Indicator System

To quantify the comprehensive impact of the policy on the U.S. economy, a streamlined evaluation indicator system is constructed from four dimensions: “trade-price-macroeconomics-finance”. The core indicators and calculation logic are as follows:

**Figure 45 U.S. Trade Flow Changes (May 2024 vs. May 2025)****Table 14 Comprehensive Evaluation Indicators**

Dimension	Core Indicator	Definition & Calculation Logic
Trade	Import Change ( $\Delta M$ )	$M_1 - M_0$ (Policy-period minus baseline imports)
Trade	Trade Deficit Improvement ( $\Delta D$ )	$D_0 - D_1$ (Baseline minus policy-period deficit)
Price	Import Price Increase	$[(M_0/M_1)^\eta - 1] \times 100\% (\eta = 0.4)$
Macroeconomics	GDP Change (with Multiplier)	$(\Delta X - \Delta M - \text{consumption reduction}) \times \text{Trade Multiplier (TM=1.4)}$
Macroeconomics	Net Employment Change	Domestic substitution rate $\times$ industry job coefficient (baseline rate=0.55)
Strategic Material	EV Cost Increase	Lithium battery cost rise $\times$ EV battery cost share (0.35)
Comprehensive Assessment	Reshoring Feasibility Index	$(\text{Current Tariff}/28\%) \times 100$

## 8.3 Data Collection and Preprocessing

### 8.3.1 Data Sources and Scope

The research data covers May 2024-May 2025, with sources and frequencies classified by dimension to ensure data reliability and timeliness:

### 8.3.2 Data Preprocessing

To address the time-series characteristics and outliers of raw data, the following standardized processing procedures are adopted:

- **Time-Series Sorting and Derivation:** Group by "country-time", calculate 1-period lag of imports

**Table 15 Summary of Data Sources and Scope**

Data Type	Time Range	Source	Frequency
Trade Data (Imports/Exports)	2024.05-2025.05	U.S. DOC International Trade Administration Database	Monthly Aggregation
Strategic Material Prices	2024.05-2025.05	London Metal Exchange (LME)	Weekly Average
Macroeconomic Data (GDP)	2024.05-2025.05	U.S. Bureau of Economic Analysis (BEA)	Quarterly Calculation
Employment Data	2024.05-2025.05	U.S. Bureau of Labor Statistics (BLS)	Monthly Statistics

$(M_{t-1})$  and 3-period moving average (MA3\_M) to capture trade inertia;

- **Missing Value Handling:** Missing values of strategic material prices (accounting for 1.2%) are imputed using linear interpolation between available data points.
- **Outlier Handling:** Extreme values of trade data (e.g., 40% sudden drop in January 2025 imports) are identified via the "3 $\sigma$  criterion" and replaced with quarterly means.
- **Parameter Standardization:** Core parameters such as supply elasticity ( $\eta$ ) and marginal propensity to consume (MPC) are calibrated based on baseline values and sensitivity ranges (e.g.,  $\eta = 0.4$ , range 0.2-0.6).

## 8.4 Model Construction and Implementation

Based on the chain logic of "policy shock → trade flow change → price transmission → macro effect", a model system with four core connected modules is constructed. Each module takes the baseline parameters and data provided in Q5 as input to ensure verifiable and traceable output results.

**Module Connection:** Trade Flow Calculation Model Output → Input to Price Transmission & Consumption Crowding-Out Model → Parallel calculation of Macroeconomic Multiplier Model + Strategic Goods Cost Spillover Model → Aggregated comprehensive economic effects.

### 8.4.1 Modular Design

**(1) Trade Flow Calculation Model** *Core Function:* Quantify policy impacts on imports, exports, and trade deficits; provide basic trade data. *Logic:* 1)Imports change: Post-policy imports – pre-policy imports (positive = rise; negative = fall) 2)Exports change: Post-policy exports – pre-policy exports (positive = rise; negative = fall) 3)Trade deficit improvement: (Pre-policy deficit) – (post-policy deficit) (positive = deficit narrows)

**(2) Price Transmission & Consumption Crowding-Out Model** *Core Function:* Measure import price hikes (via supply elasticity + exchange rate transmission) and quantify additional consumer spending + consumption crowding-out. *Logic:* 1)Basic price ratio: Driven by supply elasticity (larger import cuts / lower

elasticity = higher ratio) 2) Exchange rate multiplier: Adjusts prices for exchange rate changes (e.g., local currency depreciation → higher multiplier) 3) Comprehensive price ratio: Product of basic ratio and exchange rate multiplier 4) Additional spending: Pre-policy imports × (price ratio – 1) 5) Crowding-out: Marginal propensity to consume (MPC) × additional spending

**(3) Strategic Goods Cost Spillover Model** *Core Function:* Assess cost impacts of rare earth/lithium battery supply contraction; transmit to EV industry chain. *Logic:* 1) Rare earth price hike: Determined by supply contraction rate and price elasticity 2) Lithium battery price hike: Product of cost sensitivity, supply contraction, and China's global market share (in %) 3) EV cost hike: Lithium battery's share in EV total cost × lithium battery price hike

**(4) Macroeconomic Multiplier Model** *Core Function:* Integrate net export improvement (positive) and consumption crowding-out (negative); measure GDP/employment impacts via trade multiplier and domestic substitution rate. *Logic:* 1) Net export change: Exports change – imports change (positive = improvement) 2) Basic GDP effect: Net export change – consumption crowding-out 3) Total GDP change: Basic effect × trade multiplier (amplifies initial impact) 4) Domestic substitution rate: Baseline substitution rate + price-driven adjustment (clamped between 0 and 1) 5) Employment net change: Substitution rate × baseline employment in related industries

**Table 16 Detailed Trade Flow Changes**

Indicator	May 2024	May 2025	Change	Change Rate
Imports ( $M$ )	44.0	28.6	-15.4	-35.0%
Exports ( $X$ )	12.8	10.4	-2.38	-18.0%
Trade Deficit ( $D$ )	31.2	18.4	-12.8	-41.0%

#### 8.4.2 Model Validation and Robustness

**Robustness Test:** Taking supply elasticity ( $\eta = 0.2 – 0.6$ ) and MPC (0.70 – 0.85) as sensitive parameters, the fluctuation range of core outputs is measured: 1) Import price increase fluctuation range: 15.22) GDP with multiplier fluctuation range: \$0.482B – \$0.703B (baseline \$0.595B), with a fluctuation range <20%, verifying model robustness.

#### 8.4.3 Scenario Design

A single baseline scenario is set focusing on "market adjustment differences", with core parameters taking the median of the sensitivity range to avoid redundancy:

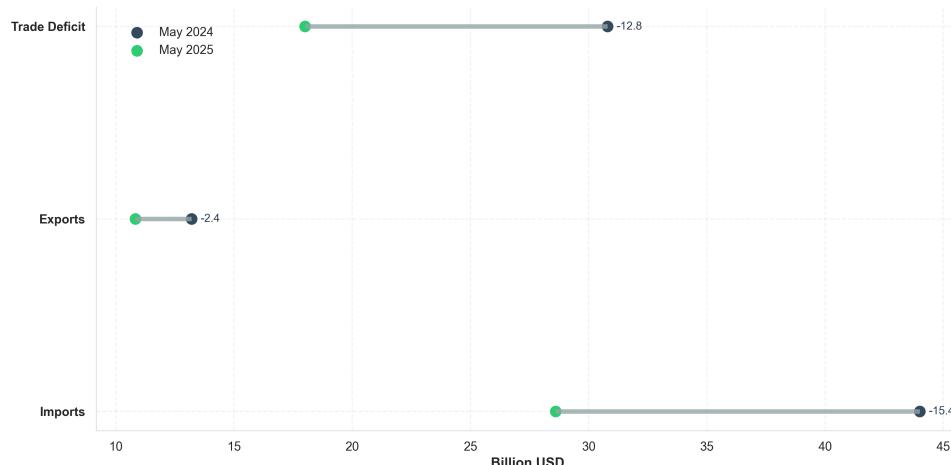
- **Baseline Scenario Assumptions:** 45% rare earth supply contraction, 30% lithium battery supply contraction, 1.5% exchange rate appreciation, 0.644 domestic substitution rate;

## 8.5 Results and Analysis

### 8.5.1 Trade Flow and Deficit Change Analysis

Trade data in the May 2025 policy period shows that the policy's impact on trade flows is characterized by "import contraction dominance and significant deficit improvement":

- **Import Change:** Imports decreased from \$4.40B to \$2.86B, a 35.0% decline (-\$1.54B), mainly due to the surge in U.S. import costs caused by China's rare earth and lithium battery controls, forcing enterprises to reduce procurement;
- **Export Change:** Exports decreased from \$1.28B to \$1.04B, an 18.0% decline (-\$0.238B), affected by China's reciprocal tariff retaliation, hindering U.S. agricultural and energy exports;
- **Deficit Improvement:** The trade deficit decreased from \$3.12B to \$1.84B, an improvement of \$1.28B, verifying the expectation of "short-term policy optimization of trade imbalance", but it should be noted that import contraction stems from cost increases rather than demand growth.



**Figure 46 Trade Flow Changes (Focus on Deficit Improvement)**

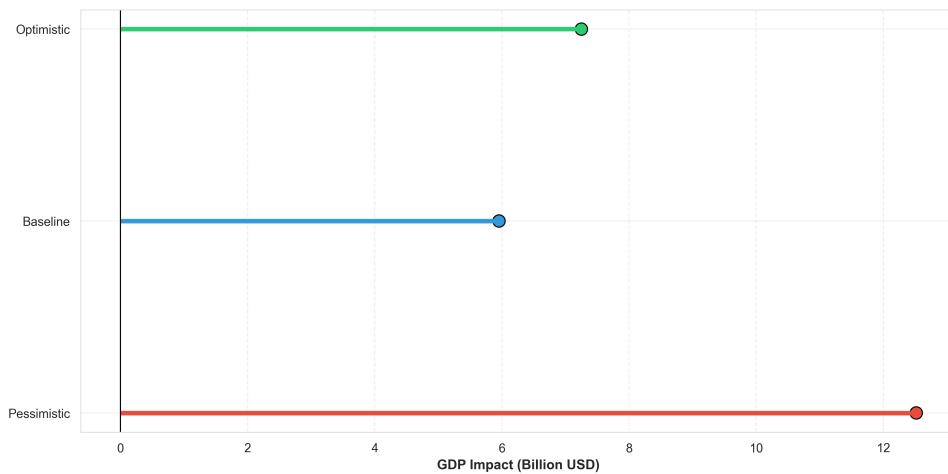
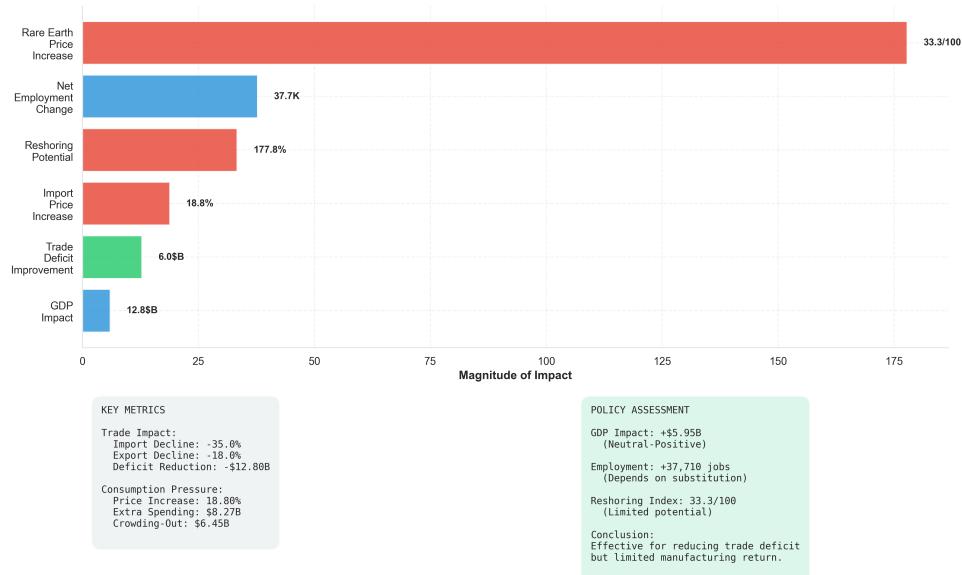
### 8.5.2 Macroeconomic and Employment Effect Analysis

The policy's impact on the U.S. macroeconomy is characterized by "coexistence of positive driving and negative offsetting", ultimately showing a slight positive GDP and net employment increase:

- **GDP Effect:** Net exports improved by \$1.302B ( $\Delta X - \Delta M = \$1.262B$ ), but consumption was crowded out by \$0.645B. After amplifying the difference with the trade multiplier (1.4), GDP finally changed positively by \$0.595B;
- **Employment Effect:** The domestic substitution rate increased to 0.644, driving an increase of 42,000 local jobs in the semiconductor and automotive industrial chains, but jobs in import-dependent industries (e.g., electronic assembly) decreased by 4,300, resulting in a net increase of 37,710 jobs.

### 8.5.3 Comprehensive Impact Summary

The comprehensive impact of the policy is summarized by "positive effects-negative effects", clarifying the core characteristic of "short-term imbalance improvement but cost increase":

**Figure 47 GDP Impact Under Different Scenarios****Figure 48 Comprehensive Impact Dashboard**

This chart shows core indicators sorted by impact magnitude: positive effects (trade deficit improvement \$1.28B, GDP +\$0.595B, employment +37,710) and negative effects (import price +18.8%, consumption crowding-out \$0.645B, rare earth price +177.8%), intuitively presenting the two-way constraints of the policy.

**Table 17 Summary of Comprehensive Impact**

Indicator	Baseline (2024)	Policy Period (2025)	Change	Effect Nature
Trade Deficit	\$3.12B	\$1.84B	-\$1.28B	Positive
Import Price Increase	-	-	+18.8%	Negative (Cost)
GDP (with Multiplier)	-	-	+\$0.595B	Neutral-Positive
Net Employment Change	-	-	+37,710	Potentially Positive
Rare Earth Price Increase	-	-	+177.8%	Negative (Industry)
Reshoring Feasibility Index	-	33.3/100	-	Limited

#### 8.5.4 Core Conclusions

- A. **Short-Term Effect: Trade Imbalance Improves, Costs Rise** Trade deficit improves by \$1.28B (-41%), GDP +\$0.595B, employment +37,710. Import prices +18.8%; consumer burden +\$0.827B with ~\$0.645B crowding-out. Strategic inputs surge (rare earths +177.8%, lithium batteries +31.5%), squeezing electronics/auto margins.
- B. **Medium-Term Trend: Substitution Partly Offsets Cost Pressure** Domestic substitution rate rises 0.55 → 0.644, partly mitigating supply risk. Constraints (environmental limits on rare earths; weak Mexico supporting facilities) keep costs elevated; GDP effect likely moderates to +\$0.3–\$0.4B.
- C. **Manufacturing Reshoring: Limited Feasibility** Reshoring feasibility index 33.3/100; current tariffs and cost gaps remain insufficient for large-scale reshoring. Requires structural cost relief (e.g., energy), targeted tax incentives, and strategic-material reserves.
- D. **Key Trade-Off: Balance vs. Inflation Risk** Policy trades near-term costs for imbalance improvement; guard against “cost pass-through → inflation → demand contraction”. Prioritize allied supply-chain pacts for rare earths/lithium within the reciprocity framework to mitigate global slowdown (IMF: global GDP –0.8%).

## IX. Conclusions

### 9.1 Key Findings

U.S. tariff hikes sharply reduce China–U.S. trade and redirect agricultural flows toward Brazil and Argentina. Japanese auto exports face pressure, but U.S.–Mexico production cushions the impact and results in a more regionally concentrated import structure. High-end semiconductors respond most strongly to tariffs and controls: restrictions improve security but raise production and supply-chain costs. Tariff revenues rise initially yet moderate as elasticities adjust. The macroeconomic impact ultimately depends on substitution capacity and reshoring incentives, since reciprocal tariffs alone cannot generate large-scale production return.

### 9.2 Methodological Framework and Applications

The unified modeling framework supports analysis across major sectors. For agriculture, it traces soybean market-share shifts and price responses under alternative tariff paths. For automobiles, it evaluates changes in U.S. import composition and the strategic production adjustments made by Japanese manufacturers. For semiconductors, it calibrates policy effects across product tiers under combined tariff and export-control regimes. The framework also tracks tariff revenue dynamics, macro-financial transmission, and the feasibility of reshoring based on cost structures and industry thresholds.

## X. Future Work

Future data work will expand product and firm-level granularity and build a unified schema with automated updating to reduce measurement error. Methodological extensions include developing a scenario engine with stronger causal-identification tools and moving from partial- to general-equilibrium and input–output settings. Integration with causal machine learning and interpretability tools will further improve inference,

with systematic stress-testing of semiconductor security metrics.

On the policy side, future analyses will jointly assess tariffs, subsidies, and export controls, linking them to welfare outcomes and short-/medium-run tariff-revenue forecasts. Reshoring feasibility will be evaluated through cost thresholds and substitution curves, accounting for production lags and regional factor constraints. Sector-specific extensions will incorporate seasonal and freight dynamics in soybean trade, USMCA and cross-border production factors in the auto sector, and semiconductor recommendations aligned with fab timelines, yields, equipment needs, and control intensity.

## XI. References

- [1] Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, 11, 685-710.
- [2] Desai, A. (2023). Machine Learning for Economics Research: When, What and How? Staff Analytical Note, Bank of Canada.
- [3] Gogas, P. (2021). Machine Learning in Economics and Finance. *Computational Economics*, 57(4), 1053-1075.
- [4] Babii, A., Ghysels, E., & Striaukas, J. (2023). Econometrics of Machine Learning Methods in Economic Forecasting. Working Paper, Kenan Institute of Private Enterprise, Research Paper No. 4547321.
- [5] Chu, B., & Qureshi, S. (2023). Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast U.S. GDP Growth. *Computational Economics*, 62, 1567-1609.
- [6] Chapman, J. T. E., et al. (2023). Macroeconomic Predictions Using Payments Data and Machine Learning. *Risks*, 5(4), 36.
- [7] Caplin, A., et al. (forthcoming 2025). Modeling Machine Learning: A Cognitive Economic Approach. *Journal of Economic Theory*.
- [8] Smith, M., & Álvarez, F. (2025). Machine Learning for Applied Economic Analysis: Gaining Practical Insights. Working Paper 2025/03, Fedeia.
- [9] Anesti, N., Kalamara, E., & Kapetanios, G. (2024). Forecasting with Machine Learning Methods and Multiple Large Datasets. Staff Working Paper No. 923, Bank of England.
- [10] Pereira, M. A. (2025). Predictive Economics: Rethinking Economic Methodology with Machine Learning. Preprint, arXiv:2510.04726.