Journal of Agricultural and Applied Economics (2025), **57**, 240–258
doi:10.1017/aae.2025.6

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Exploring the Impact of China's Retaliatory Tariffs on US Soybean Exports with Machine Learning Techniques

Anastasia W. Thayer[1], Pengyan Sun[2], Hernan A. Tejeda[3] and Man-Keun Kim[1,4]

[1]Department of Agricultural Sciences, Clemson University, Clemson, SC, USA, [2]School of Statistics, University of Minnesota, Minneapolis, MN, USA, [3]Department of Agricultural Economics and Rural Sociology, University of Idaho, Moscow, ID, USA and [4]Department of Applied Economics, Utah State University, Logan, UT, USA
Corresponding author: Anastasia W. Thayer; Email: awthaye@clemson.edu

Abstract
The 2018/2019 trade conflict between the United States and China impacted a broad array of agricultural products, including soybeans. Previous trade studies using gravity models fail to account for trends and complex seasonal patterns observed in the data. This study uses a machine learning (ML) approach to estimate losses in soybean export value and volume from the trade war. We find that models using ML techniques outperform traditional models and estimate losses in the value of soybean exports of $10.16 billion/year. The ML models fit the complex export trade data series well, highlighting the importance of utilizing improved modeling approaches.

Keywords: China; international trade; machine learning; retaliatory tariffs; trade conflict; soybean

JEL classifications: Q17 (agriculture in international trade); F17 (trade forecasting and simulation)

1. Introduction

Against the backdrop of tariff impositions[1] enacted in June 2018 by the Trump administration targeting Chinese imports, China responded with a series of retaliatory measures affecting a broad spectrum of US exports. These included tariffs ranging from 2.5% to 25% and marked the onset of an escalating trade dispute. Rooted in concerns over the ballooning trade deficit with China and the erosion of manufacturing jobs, the United States pursued a strategy of trade protectionism via unilateral tariff impositions to fortify its domestic economy. Given China's sizable role as one of the largest markets for United States agricultural exports and the largest market for US soybeans, commodities such as soybeans and pork bore significant retaliatory impacts from these retaliatory tariffs (Adjemian et al., 2021). Moreover, other nations responded to US tariffs on steel and aluminum imports by imposing reciprocal policies, collectively targeting approximately $450 billion worth of trade (Fajgelbaum et al., 2024), with soybean exports notably taking a severe hit.

A study by Grant et al. (2021) offers insights into the 2018/2019 trade conflict dynamics. The research studies bilateral trade relationships of agricultural products. Estimating a gravity model

[1]In 2018, the United States initiated Section 232 tariffs on steel and aluminum imports from major trading counterparts and Section 301 tariffs targeting a wide range of imports from China. In response, Canada, China, the European Union, India, Mexico, and Turkey enacted retaliatory tariffs on various US exports, including an extensive array of agricultural and food products (Morgan et al., 2022).

© The Author(s), 2025. Published by Cambridge University Press on behalf of Southern Agricultural Economics Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

with monthly panel data spanning from January 2016 to December 2019 and honing in on the repercussions of retaliatory measures on US agricultural exports, the study incorporates specific fixed effects to account for variations in trade values. Findings underscore significant trade losses, primarily attributed to retaliatory actions by China, culminating in annualized direct losses ranging from $12.6 to $17.2 billion for US agriculture in Chinese markets (Grant et al., 2021). Notably, sectors like soybeans and pork products experienced more pronounced contractions, with soybean exports facing losses of approximately $10.7 billion per year and pork products suffering losses of around $777 million per year (Grant et al., 2021). In response to projected losses, the US Department of Agriculture, under the Market Facilitation Program, instituted payments in total of $8.5 billion to soybean producers in an attempt to compensate for expected losses incurred due to trade restrictions (Adjemian et al., 2021). These findings underscore the challenges posed by retaliatory tariffs and highlight the pressing need for strategic policies to mitigate adverse effects on US agricultural exports.

A related comprehensive assessment conducted by the US Department of Agriculture (Morgan et al., 2022), following a similar gravity model econometric framework to Grant et al. (2021), emphasized the significant toll inflicted by retaliatory tariffs and found higher estimated losses. Building upon the approach of Grant et al. (2021), Morgan et al. (2022) expanded the analysis by measuring the impact of export losses at the state level to ascertain any disproportionately large effects. The study unveiled a sizeable $27 billion reduction in US agricultural exports from mid-2018 to the end of 2019. Soybeans bore the brunt of the impact, accounting for approximately 71% of the decline, which amounted to $9.4 billion in annual losses. The disproportionate effect on soybeans resonated across states heavily reliant on soybean exports, such as Iowa, Illinois, and Kansas, where gross domestic product losses totaled $3.8 billion annually. Overall, the United States incurred a trade loss of nearly $27 billion with countries that imposed retaliatory tariffs due to these tariffs, highlighting the widespread yet commodity-concentrated nature of export losses.

In this study, we reassess the impact of China's retaliatory tariffs, concentrating on soybean exports and applying machine learning (ML) methods. ML approaches are favorable for international trade studies since they are able to outperform traditional forecasting methods when the underlying time series may be heteroskedastic or have complex processes such as trend and seasonality, both frequently present in agricultural export data. The selection of soybean exports as our focal point is motivated by the proportionally large annual trade volume, associated value, and significant damage incurred. Acknowledging that the reduction in soybean exports to China had spillover effects with trade to other countries (Adjemian et al., 2021), we also consider changes in soybean exports to the rest of the world. Further, to capture not only the change in trade volumes, we also explore changes to the value of soybean exports to capture price effects.

This research makes two significant contributions to the existing body of literature. First, this work provides monthly estimates of the impact of retaliatory tariffs on soybean trade in terms of volume and value. While recent pertinent research conducted by Grant et al. (2021) and Morgan et al. (2022) has delved into the response of US agricultural exports, including soybeans, to retaliatory tariffs, these studies have applied gravity models and suffer from their associated limitations. In effect, Grant et al. (2021) and Morgan et al. (2022) employ a gravity model with dummy variables to capture the post-retaliation period. While these dummy variables are valuable for gauging the impact and estimating changes in soybean exports affected by retaliatory tariffs, they do not directly capture trends and complex seasonal patterns within the data. As such, they may not fully account for underlying large seasonal fluctuations present in addition to marked changes in the overall yearly trends in the data. Notably, coefficient estimates from gravity models fail to provide consistent coefficient estimates in the face of heteroskedasticity (Kareem and Kareem, 2014). Thus, it is plausible that the estimates produced might be underestimated, given the increasing trend observed in soybean exports to China. Others, such as Adjemian et al. (2021) and Sabala and Devadoss (2019), offer different applied methodologies and loss estimates, as shown below; however, a notable gap persists as ranges for loss estimates remain substantially

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

large and lack an enhanced approach that can address trend and seasonality. We utilize forecast models with ML techniques to address this limitation and provide improved estimates.

A second related contribution of this study is the utilization of ML techniques to address research questions in the arena of trade. Few studies exist on this topic, yet recent contributions highlight the growing interest and benefits of utilizing ML techniques given the nature of trade data (Gopinath et al., 2021; Sei et al., 2024). Critiques of ML approaches include concerns about the lack of large sample properties of estimators and the challenge of comparing results under multiple ML models or approaches (Athey and Imbens, 2019). Furthermore, shortcomings of ML methods might be overfitting, validation, and out-of-sample estimation compared to traditional econometric models used for forecasting (Athey and Imbens, 2019). Despite these challenges, ML methods are increasingly appealing to address research questions in agricultural trade (Gopinath et al., 2020). Compared to general equilibrium models or traditional econometric approaches, ML approaches may offer more comprehensive forecasts of trade volumes and values under various policies and scenarios, such as tariffs (Gopinath et al., 2021).

2. Related studies

2.1. Soybean trade and market background

Since 2001, with the introduction of the World Trade Organization, global soybean markets have experienced pronounced changes to traded volume as well as the rise of global export actors, including Brazil, Argentina, and the United States (Gale et al., 2019). Moreover, demand from China and other Asian countries gradually expanded with approximately 66% of all soybean trade in 2016/1027 marketing year being from Brazil and the United States to China (Gale et al., 2019). For the United States, top export markets during this time included countries in East Asia, Southeast Asia, and the European Union (Gale et al., 2019).

During the years preceding the 2018 trade conflict, many factors were at play that determined soybean supply, demand, and prices. In particular, US soybean production and acreage steadily grew until 2018, the same year as the trade conflict, with production levels reaching 4.43 billion bushels and 89.2 million planted acres (Bukowski and Swearingen, 2023, 2025). From 2013 through 2018, while production gradually increased, exports as a percentage of production have ranged from 44% to 50% (Bukowski and Ates, 2023). Moreover, prices fluctuated considerably, with farm prices in 2012 and 2013 garnering prices of over $13.00/bushel but falling to $8.95/bushel in subsequent years. Before the trade war, while producers had experienced wide price swings, exports as a percentage of production remained relatively stable. Conversely, with only moderate gains in domestic use, ending stocks increased from 191 million bushels in 2014 to 438 million bushels in 2017 to 909 million bushels in 2018, the year when the trade conflict began (Kansas State University, 2024).

2.2. Impact of retaliatory tariffs

In 2018, the United States imposed a series of tariffs on steel and aluminum imports from key trading partners, including China, Canada, the European Union, India, Mexico, and Turkey, under Section 232 tariffs, and imposed additional tariffs on Chinese imports under Section 301 tariffs (Morgan et al., 2022; Sukar and Ahmed, 2019; York, 2023). In response, these countries retaliated with import tariffs on US goods, escalating tariffs on globally traded goods (Morgan et al., 2022; Sukar and Ahmed, 2019; York, 2023). Numerous studies have since examined the impacts of the trade conflict on various sectors, regions, and commodities.

Williams and Hammand (2020) argued that the trade conflict and ensuing tariffs could negatively affect the US economy by increasing economic uncertainty, decreasing business investment, and hindering economic development. Williams and Hammand (2020) also predicted that escalating US tariffs would lower both US and global gross domestic product. Retaliatory

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

tariffs further exacerbated the potentially adverse effects on the US economy and specific sectors, including agriculture. In an early study, Flaaen and Pierce (2019) analyzed the domestic impacts of the trade war on US manufacturing, finding that the positive effects of tariffs were offset by higher costs from retaliatory tariffs, thereby leading to no net increase in manufacturing employment or output. Additionally, tariffs did not effectively protect US manufacturing in the short run, and domestic manufacturers may have lost competitiveness due to retaliatory tariffs.

The agricultural sector also faced significant challenges due to import tariffs imposed by trading partners. For example, Mexico imposed tariffs on cheese imports from the United States, resulting in estimated losses of $68 million in the first year, with additional indirect and induced effects (Ribera et al., 2018). An event study framework by Carter and Steinback (2020) investigated the short-run trade effects of retaliatory tariffs against US agricultural and food exports. The study found a substantial decline in US agricultural and food exports, with losses outweighing gains from increased trade with countries not imposing retaliatory tariffs on the United States of more than $14.4 billion. In general, countries that did not impose retaliatory tariffs reoriented their trade relationships, with South American and European countries benefiting the most from these adjustments. Compared to other commodities, US soybeans and meat products experienced significant redistribution effects. A study by Sabala and Devadoss (2019) used computable general equilibrium modeling approaches to estimate US losses in soybean markets of $2.72 billion with reductions in producer surplus ($5.52 billion loss) overshadowing gains in consumer surplus ($2.8 billion gain).

Grant et al. (2021), employing a monthly gravity model and utilizing panel data from January 2016 to December 2019, examined bilateral trade relationships, focusing on agricultural products. The paper integrates fixed effects to account for variations in trade policy. Findings reveal significant trade losses primarily attributed to retaliatory actions by China with, as previously mentioned, annualized direct losses ranging from $13.5 to $18.7 billion for US agriculture in Chinese markets alone. These findings highlight the challenges posed by retaliatory tariffs, emphasizing the need for strategic trade policies that mitigate adverse effects on US agriculture.

To further hone in on the impacts to specific states, Morgan et al. (2022) use a comparable econometric framework to Grant et al. (2021). They also find significant impacts of the retaliatory tariffs on US agricultural exports. They identified a $27 billion reduction in exports from mid-2018 to the end of 2019, with soybeans bearing the brunt of this decline, accounting for 71%. This sizable impact disproportionately affected states heavily reliant on soybean exports, such as Iowa, Illinois, and Kansas. According to Morgan et al. (2022), in response to reductions in US imports, China increased agricultural imports from other countries, causing a redistribution of imports of nearly $16 billion with countries such as Brazil and the European Union.

This paper focuses on soybean exports to China, given the substantial impact this commodity faced as a result of the trade conflict. Notably, studies by Carter and Steinback (2020), Grant et al. (2021), and Morgan et al. (2022) underscored this assertion. China's imposition of retaliatory tariffs on US soybeans triggered a significant decline in exports. Given the strategic importance of soybeans in US agricultural trade, particularly with China being the largest importer of US soybeans, this paper aims to provide a nuanced understanding of the reduction of US soybean exports during this time.

2.3. Machine learning approaches to model international trade

Despite growing interest and promising applications of ML approaches in agricultural economics, there remain few published studies using this methodology (Coble et al., 2018; Storm et al., 2020). A notable consideration of ML applications is that they exhibit distinct characteristics when handling either cross-sectional or time-series/spatially dynamic data (Storm et al., 2020). As mentioned previously, a common critique of ML approaches has been the lack of large sample properties of estimators or clear guidance on comparing results across multiple estimated ML models

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

or approaches (Athey and Imbens, 2019). Applications of ML to gravity models in international trade have shown promising results, with ML techniques often providing a better fit and predictive accuracy compared to traditional models (Gopinath et al., 2020, 2021). Consequently, ML approaches may be more suitable for forecasting trade volumes under various policies and scenarios, such as tariffs, than general equilibrium models or traditional econometric approaches.

In the case of ML applications to trade, a recent study by Breinlich et al. (2022) highlights the challenges of overfitting and multicollinearity when estimating the effects of trade provisions and proposes novel data-driven methods, drawing from ML and variable selection literature, to address these issues. Jošić and Žmuk (2022) employ ML algorithms to forecast Croatia's international bilateral trade flows, presenting the efficacy of ML in predicting trade patterns and providing valuable insights for economic policymakers.

In a different vein, Baier and Regmi (2023) explore the use of ML techniques to capture heterogeneity in free trade agreements, finding that more comprehensive agreements tend to result in larger estimated impacts on trade. Gupta and Kumar (2021) conduct a comprehensive literature review of ML applications in forecasting international trade patterns, emphasizing the increasing adoption of ML and deep learning techniques for economic forecasting. While traditional statistical approaches retain their importance, ML techniques are gaining prominence and often offer superior accuracy in economic forecasting tasks. This is especially poignant given that another drawback of gravity model approaches used in traditional econometric estimates has been considerable variation in mean estimates for trade (Cipollina and Salvatici, 2010).

3. Data and modeling procedure

We use historical soybean export data from the US Department of Commerce Bureau of Trade (International Trade Data. U.S. Census Bureau. https://www.census.gov/foreign-trade/data/index.html) to quantify the changes in US soybean exports in terms of volume and value to China and all other regions as a result of retaliatory tariffs. We estimate traditional forecasting models as well as forecasting models that include ML methods to predict the expected value and the volume of soybeans exported without tariffs and then compare them to the actual exports. The data, empirical approach and models, and modeling procedure are described below.

3.1. Soybean exports

Time series data of monthly values of US soybean exports to China and to other regions from January 2010 to December 2021 (144 observations) were obtained from the Global Agricultural Trade System (https://apps.fas.usda.gov/gats/default.aspx). Figure 1 illustrates the value and volume of soybean exports to China and to the other regions over this period. As shown, China is by far the largest US agricultural soybean export market (Figure 1). Before the onset of the 2018 trade war, US soybean exports to China had shown rapid growth, outpacing growth in overall global exports (Tortajada and Zhang, 2022), at $1 billion in 2000 and reaching $14 billion in 2016. Notably, 61% of US soybeans were imported by China (Gale et al., 2019) in the year before the trade conflict began.

As trade tensions escalated into a full-scale trade war in 2018, the value and volume of US soybean exports to China plummeted (Figure 1). Export value to China fell sharply from $12.3 billion in 2017 to $3.1 billion in 2018. At the same time, international soybean trade markets were disrupted as China imported soybeans from Brazil to meet the shortfall in US soybeans while the United States sought to increase soybean exports to other markets (Gale et al., 2019). This can be seen in Figure 1, with steady seasonal patterns in value and volume of exports to other regions before the trade war and increased exports after the trade war began.

One notable trend in Figure 1 is the overall steady fluctuations in export value and volumes. Seasonal patterns are evident, with peaks and troughs in export levels throughout the year,

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Figure 1. US soybean exports to China and other regions. *Note:* The dotted lines represent locally estimated scatterplot smoothing (LOESS) smoothers, which are used to visualize the underlying trends in the data. *Source*: Global Agricultural Trade System (GATS) (https://apps.fas.usda.gov/gats/default.aspx).

reflecting harvest and timing of supply where higher export values and volumes are observed during harvest seasons, while lower volumes occur during off-harvest periods. A significant event marked in the data is the onset of the trade conflict between the United States and China in June 2018, indicated by the dotted vertical line. Following this event, some disruption appears in the trend of the value and volume of soybean exports to China, with fluctuations and possibly a slowdown in export growth. More specifically, there was a substantial reduction in the value and volume of US exports to China following the implementation of retaliatory tariffs, lasting at least until the end of 2019.

In January 2020, the United States and China reached an enforceable agreement known as the Phase One trade deal. This agreement entailed structural reforms and other changes to China's economic and trade regime in various sectors, including intellectual property, technology transfer, agriculture, financial services, and currency and foreign exchange (Bown, 2022; Tortajada and Zhang, 2022). Additionally, the Phase One trade agreement outlined China's commitment to make substantial additional purchases of US goods and services over the coming years. One significant aspect of the agreement was the establishment of a robust dispute resolution system. As part of the deal, the United States agreed to modify its Section 301 tariff actions significantly. Consequently, soybean exports, which had been impacted by the trade tensions, appear to return to normal levels following the agreement (Tortajada and Zhang, 2022).

3.2. Modeling procedure

To forecast the value and volume of US soybean exports to China and other regions in the absence of retaliatory tariffs, we divided each of the 144 monthly observations into two groups. Group 1

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Figure 2. Soybean export to China showing observations for training, testing, and forecasting.

included 102 observations (from January 2010 to June 2018[2]), representing the pre-retaliatory tariff period. Group 2 consisted of the 42 monthly observations that occurred post-retaliatory tariffs (from July 2018 to December 2021). For the purposes of illustration, Figure 2 shows the forecasting approach using the value of soybean trade to China. The estimation process is repeated for soybean export volume to China, value of exports to other regions, and volume of exports to other regions. Group 1 includes the observations from the training and testing set, delineated up until the vertical dashed line, where the trade conflict began. Group 2 consists of the forecasted, delineated from the second vertical dashed line indicating the start of the trade conflict to the right.

The data observations from Group 1 were further split into two subgroups. The first subgroup comprised data from January 2010 to June 2016, used to train (or estimate) the models and then used to forecast either the volume or value of soybeans exported to China or other regions from July 2016 and June 2018. This latter is named the testing period (red solid line in Figure 2). The difference between the actual and forecasted values from June 2016 to June 2018 is tested for each model so as to find the best forecasting model. After selecting the best model according to accuracy criteria, we re-estimated exports (volume and value) using observations until June 2018 and then forecasted the post-retaliatory tariff period (from July 2018 to December 2021). The difference between these forecasted values and the volume of exports and the actual value and volume of exports reflects the change in exports due to tariffs.

To summarize, the models are trained on data, and then forecasted values are predicted using a two-step process for validation and comparison (Ramasubramanian and Singh, 2018). In the first step, the training data from January 2010 to June 2016 is used to train the models. After the training stage, estimated models are used to predict the test set or subset of observation (forecasting July 2016 to June 2018). A model's predictive abilities were evaluated using established methods to compare model fit by comparing the actual testing set data with the predicted data. In the second step, the data from Group 1 data (January 2010 to June 2018) were used to train the models a second time and then forecast forward (forecasting Group 2 from July 2018 to December 2021).

4. Empirical models

Six models were selected to forecast the volume and value of exports of US soybeans to China and other regions following the implementation of retaliatory tariffs. The purpose of using multiple

[2]Data begins in 2010. Prior to 2010, soybean prices experienced a period of high levels, and there were significant variations in domestic consumption of soybean meal (Gale et al., 2019).

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

models is to compare the forecasted value and volume of soybean exports to China and other regions (Figure 2) against the actual observed values. The six models used in this comparison were ARIMA, ARIMA XGBoost, Prophet, Prophet XGBoost, Elastic Net, and Error-Trend-Season (ETS). The different models were selected to ensure an unbiased preference in the prediction outcome and according to the range of loss estimates. The ARIMA model, a widely used time-series model, is commonly employed for forecasting purposes. The Prophet model is a nonlinear model that is appropriate for series with seasonality (Hyndman and Athanasopoulos, 2021). The Elastic Net model is categorized as a global/parametric predictor model (Mullainathan and Spiess, 2017). The ETS model is another widely used time-series model that includes specific terms to capture the error, trend, and seasonality of the underlying data. The ARIMA and Prophet models do not require ML techniques, given that they follow a defined algorithm (fully described below). For these models, the procedure of Automatic Models was applied using the modeltime package (Dancho, 2023). These diverse models allow for a more robust, informed forecasting process including a comprehensive evaluation of their performance. All statistical analysis, forecasting, and use of the aforementioned packages was computed using R.

4.1. ARIMA

The ARIMA model stands for autoregressive integrated moving average and is commonly used for analyzing time-series data. If additional seasonal terms are included, the model is known as SARIMA, represented as SARIMA(p,d,q)(P,D,Q)s, where p and q refer to the orders of autoregressive and moving average components, d is the degree of differencing, and s refers to the number of periods in each season. The uppercase P, D, and Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the model. Discussion regarding SARIMA modeling closely follows the notations and conventions introduced by Lütkepohl and Krätzig (2004). The general form of the SARIMA model is given by:

$$\phi_s(L^s)\phi(L)\Delta_s^d\Delta^d y_t = \theta_s(L^s)\theta(L)\varepsilon_t \tag{1}$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, $\phi_s(L^s) = 1 - \phi_{s1}L^s - \phi_{s2}L^s 2 - \cdots - \phi_{sP}L^{sP}$, $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, and $\theta_s(L^s) = 1 + \theta_{s1}L^s + \cdots \theta_{sQ}L^{sQ}$ where L indicates the lag operator. For instance, SARIMA(1,1,1)(1,0,1)[4] model is given by $(1-\phi_{41}L^4)(1-\phi_1 L)\Delta y_t = (1+\theta_{41})(1+\theta_1 L)\varepsilon_t$ or $\Delta y_t = \phi_1\Delta y_{t-1} + \phi_{41}y_{t-4} - \phi_1\phi_{41}\Delta y_{t-5} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_{41}\varepsilon_{t-4} + \theta_1\theta_{41}\varepsilon_{t-5}$. Deterministic terms such as a constant or a trend may be added to equation (1).

We used the default "Auto ARIMA" function to forecast data in following the ARIMA model proposed by Hyndman and Khandakar (2008). The autoarima function utilizes unit root tests and minimizes information criteria such as AIC (Akaike information criterion), AICc (corrected Akaike criterion), or Bayesian information criterion to determine the best SARIMA model. The function searches for the most suitable model within the default constraints provided. These constraints include maximum values of P, D, and Q set at 2, 1, and 2, respectively.[3]

4.2. Prophet

The second model utilized in this study was the Prophet model, as outlined by Taylor and Letham (2018, 2022). Developed by Facebook's Core Data Science team (Taylor and Letham, 2022), this time-series forecasting algorithm is particularly effective for handling seasonal data. The Prophet model decomposes time series into three main components: trend, seasonality, and holidays, represented as:

[3]These values represent the autoregressive order (P), the difference order (D), and the moving average order (Q) (Hyndman and Athanasopoulos, 2021).

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \tag{2}$$

Here, $g(t)$ describes a piecewise-linear trend, $s(t)$ represents various seasonal patterns, $h(t)$ captures holiday effects, and ε_t denotes a white noise error term. To address seasonality, we allowed the model to toggle a seasonal component to model year-over-year patterns.

Following Taylor and Letham (2018), the seasonal component of our model captures recurring patterns at fixed intervals, such as monthly or yearly cycles. To represent this component, we employed a Fourier series within the Prophet model:

$$s(t) = \sum_{n=1}^{N} (a_i \cos 2\pi it + b_i \sin 2\pi it) \tag{3}$$

Here, a_i and b_i denote the coefficients of the seasonal Fourier terms, and i represents the frequency of the seasonal component. By using the Fourier series, Prophet efficiently captures and models recurring patterns in time-series data, ensuring accurate forecasts even in the presence of complex seasonal behavior. Since our data is monthly, we opted to ignore the holiday effect in the model. Importantly, default parameters were applied in R for all other model components.

4.3. Elastic net

The third model employed in this study integrates ML techniques, utilizing a supervised learning algorithm to glean insights from the data and generate predictions (Athey and Imbens, 2019). Specifically, we employed Elastic Net, building upon the foundational work by (Firedman et al., 2010) and the extensions proposed by Xing et al. (2022). Elastic Net extends linear regression by incorporating regularization penalties into the loss function during training. By combining both Lasso (L1) and Ridge regression (L2) techniques, Elastic Net aims to prevent overfitting by controlling the amount of information used in the model. This approach ensures more robust and accurate predictions, particularly in scenarios involving high-dimensional datasets or multi-collinearity among predictor variables. Lasso regression (Least Absolute Shrinkage and Selection Operator) adds a penalty term to the linear regression based on absolute values of regression coefficient to the loss function and aims at minimizing equation (4). It is helpful for identifying influential time-series components or predictor variables in forecasting tasks.

$$\min_{\theta} \frac{1}{2m} \left[\sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)})^2 + \lambda_1 \sum_{j=1}^{n} \| \theta \|_1 \right] \tag{4}$$

where θ is the vector of coefficients, so the first part inside the bracket in equation (4) is the linear regression, which minimizes sum of squared residuals. The term $\sum_{j=1}^{n} \| \theta \|$ is the penalty. Ridge regression adds squared magnitude of coefficient as penalty term to the loss function and aims at minimizing equation (5):

$$\min_{\theta} \frac{1}{2m} \left[\sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)})^2 + \lambda_2 \sum_{j=1}^{n} \| \theta \|_2^2 \right] \tag{5}$$

When L1 (Lasso regression) and L2 (Ridge regression) regularization are used in combination, the result describes the Elastic Net method. Elastic Net aims at minimizing equation (6):

$$\min_{\theta} \frac{1}{2m} \left[\sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)})^2 + \lambda_1 \sum_{j=1}^{n} \| \theta \|_1 + \lambda_2 \sum_{j=1}^{n} \| \theta \|_2^2 \right] \tag{6}$$

Elastic Net can handle various components commonly found in time-series data, such as trend, seasonality, and autocorrelation. By incorporating appropriate lagged values and seasonal terms as

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

predictor variables, the algorithm captures temporal dependencies and patterns, enabling accurate forecasting.

4.4. Error-Trend-Season (ETS)

The Error-Trend-Season (ETS) model is a forecasting tool that combines elements of exponential smoothing with a state space framework to capture various patterns in time-series data (Hyndman et al., 2002, 2008). The ETS model comprises three main components: Error (E), Trend (T), and Seasonality (S). The Error component accounts for random fluctuations or noise in the data and is modeled using exponential smoothing. It represents the difference between observed values and predicted values based on the trend and seasonality. The Trend component captures the long-term direction or tendency in the data. It is also modeled using exponential smoothing and represents the systematic change over time. The Seasonality component accounts for periodic fluctuations in the data that occur at regular intervals, such as weekly and monthly patterns. This component is particularly useful for modeling seasonal effects, such as holiday sales spikes or weather-related fluctuations.

Mathematically, the ETS model can be expressed as a set of recursive equations that update the estimates of the Error, Trend, and Seasonality components over time. These equations incorporate smoothing parameters that control the influence of past observations on the current estimates, allowing the model to adapt to different patterns in the data. The Error component is given by:

$$\hat{e}_t = \alpha(y_t - \hat{y}_{t|t-1}) + (1-\alpha)\hat{e}_{t-1} \tag{7}$$

where $\hat{e}_t$ represents the estimated error at time t. y_t is the observed value at time t and $\hat{y}_{t|t-1}$ is the forecasted value at time t based on the previous time step. α is the smoothing parameter for the Error component, controlling the influence of the current observation on the error estimate. Trend (T) component can be represented as follows:

$$\hat{T}_t = \beta(\hat{y}_{t|t-1} - \hat{y}_{t-1|t-1}) + (1-\beta)\hat{y}_{t-1} \tag{8}$$

where $\hat{T}_t$ represents the estimated trend at time t. β is the smoothing parameter for the trend component, controlling the influence of the previous trend estimate on the current trend estimate. Seasonality (S) component is:

$$\hat{S}_t = \gamma(y_t - \hat{y}_{t|t-1} - \hat{e}_t) + (1-\gamma)\hat{S}_{t-m} \tag{9}$$

where S_t represents the estimated seasonality at time t. γ is the smoothing parameter for the seasonality component, controlling the influence of the current observation on the seasonality estimate. m is the length of the seasonal cycle (e.g., 12 for monthly data). The forecasted value $\hat{y}_{t|t-1}$ at time t is obtained by adding the estimated components together:

$$\hat{y}_{t|t-1} = \hat{T}_{t-1} + \hat{S}_{t-m} + \hat{e}_{t-1} \tag{10}$$

One intuitive way to understand the ETS model is to think of it as decomposing the time-series data into its underlying components: a smoothed trend line, seasonal fluctuations around this trend, and random noise. By explicitly modeling these components, the ETS model can provide more accurate forecasts and better insights into the underlying dynamics of the data.

4.5. ARIMA and Prophet XGBoost

Hybrid ML models, also known as combined predictor models, integrate automated algorithms with ML techniques (Mullainathan and Spiess, 2017). In this study, we adopt the ARIMA and Prophet models and enhance them with the XGBoost algorithm (Chen and Guestrin, 2016; Chen et al., 2024). XGBoost, short for extreme Gradient Boosting, belongs to the ensemble learning

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

algorithm family. In ensemble learning, weak learners (individual models) are sequentially constructed and aggregated to form a more robust predictive model. XGBoost operates iteratively, with each subsequent model rectifying the errors of the previous ones, thereby enhancing overall performance (Mullainathan and Spiess, 2017). To implement the ARIMA and Prophet Boost models, we first model the data using ARIMA or Prophet to capture its intrinsic patterns. Next, we employ the XGBoost algorithm to regress the residuals obtained from the ARIMA or Prophet model. This two-step approach harnesses the strengths of both techniques, enabling us to capitalize on the predictive capabilities of ARIMA or Prophet's time-series modeling and the ensemble learning framework XGBoost. Importantly, XGBoost is renowned as a scalable tree boosting system (Chen and Guestrin, 2016). It stands out as a rapid and efficient tool for constructing accurate models and is widely recognized as the algorithm of choice for classification or regression problems in ML.

Intuitively, XGBoost operates by iteratively improving upon the predictive capabilities of weaker models. It does so by training subsequent models to correct the errors made by earlier ones, gradually refining predictions with each iteration. This iterative refinement process, coupled with its ensemble learning approach, enables XGBoost to capture intricate patterns and relationships within data, ultimately leading to superior predictive performance.

4.6. Forecast accuracy measures

Forecast accuracy can be measured using metrics such as mean absolute error (MAE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), or root mean squared error (RMSE) (Hyndman and Athanasopoulos, 2021). Suppose the soybean export dataset is denoted by $y_1, y_2, \cdots, y_T$, and we split it into two sections: the training data $y_1, y_2, \cdots, y_N$ and the test data $y_{N+1}, y_{N+2}, \cdots, y_T$. To assess the accuracy of our forecasting method, we estimate the parameters using the training data and forecast the next $T-N$ observations. These forecasts are then be compared to the test data. The h-step-ahead forecast can be denoted as $\hat{y}_{N+h|N}$. The "hat" notation indicates that it is an estimate rather than an observed value, and the subscript indicates that we are estimating y_{N+h} using all the data observed up to and including time period N. The forecast errors are the differences between the actual values in the test set and the forecasts produced using only the data in the training set. Thus:

$$e_t = y_t - \hat{y}_{t|N}, \quad \text{for } t = N+1, \cdots, \mathrm{T} \tag{11}$$

MAE is defined as $\mathrm{MAE} = \mathrm{mean}(|e_t|)$, MAPE as $\mathrm{MAPE} = \mathrm{mean}\left(\left|\frac{100e_t}{y_t}\right|\right)$, and RMSE as $\mathrm{RMSE} = \sqrt{\mathrm{mean}(e_t^2)}$. When comparing forecast methods on a single dataset, the MAE is popular as it is easy to understand and compute (Hyndman and Athanasopoulos, 2021).

In addition to these, we employ SMAPE, as defined by:

$$\mathrm{SMAPE} = \mathrm{mean}\left(\frac{|y_t - \hat{y}_{t|N}|}{(|y_t| + |\hat{y}_{t|N}|)/2}\right) \tag{12}$$

SMAPE addresses asymmetry in error values, where the discrepancy between predicted and actual values may yield different outcomes depending on their relative magnitudes. This issue is particularly notable in the MAPE, which some researchers caution against due to its differential treatment of forecast errors above versus below the actual observations. To mitigate this concern, a symmetric MAPE has been proposed by Goodwin and Lawton (1999).

Lastly, we may use the coefficient of determination (R^2), which measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It provides insight into how well the model explains the variability of the data, which is given by:

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_{t|N})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{13}$$

where $\bar{y}$ is the mean of the observed values.

5. Results and discussion

We trained the six models discussed above using data from January 2010 to June 2016 and tested the models from July 2016 to June 2018 for export value and volume to China and export value and volume to other regions. Using the estimates from the testing set, we calculated five different criteria metrics to evaluate model fit: MAE, MAPE, SMAPE, RMSE, and R-squared (Table 1). Each model's performance is detailed in the table provided. For MAE, MAPE, SMAPE, and RMSE, the model with the lowest value for each metric is considered the best, whereas a higher R-squared value is preferred. Based on these six models, the Prophet XGBoost model consistently achieves the lowest MAE, MAPE, and RMSE or highest R-squared value, indicating superior relative performance. From Table 1, the Prophet XGBoost model achieves the best accuracy metric when forecasting the value of exports to China and the volume of exports to other regions. It also obtains the lowest MAPE and SMAPE across all trade measures forecasted. While the Prophet XGBoost model is the most accurate model when forecasting the volume of exports to China, the SARIMA(2,0,0)(2,1,0)[12] XGBoost model also appears to perform well and shows a higher R-squared value and lower RMSE compared to the Prophet XGBoost.

It is worth noting that integrating the XGBoost algorithm enhances the performance of both the SARIMA and Prophet models for specific measures of model accuracy compared to the same model without XGBoost. For example, the SARIMA(2,0,0)(2,1,0)[12] XGBoost model indicates improved performance in terms of MAE, MAPE, SMAPE, and RMSE for volume and value of exports to China and all other regions. Similarly, the Prophet XGBoost model outperforms the standalone Prophet model, displaying higher R-squared and lower MAE, MAPE, SMAPE, and RMSE. While this suggests the effectiveness of integrating XGBoost to refine predictions and reduce errors in both SARIMA and Prophet models, there is one case in this table where the metric was not improved with the integration of XGBoost. This suggests that while integrating XGBoost into traditional forecasting models could improve accuracy, caution is recommended.

After re-estimating the value and volume of exports to China and other regions, with data spanning from January 2010 to June 2018, a forecast from July 2018 to December 2021 is estimated to quantify changes as a result of tariffs and policies enacted during the trade conflict. Figure 3[4] shows the results, where the differences between the actual (gray line) and predicted models (red dotted line) and the value and volume of exports to China and volume and value of exports to other regions are compared. Over the entire test period, the Prophet XGBoost model performed well, as shown with the red dotted line (predicted) closely tracked by the gray line (actual value and volume of exports).

From Figure 3, the large seasonal swings in value and volume of exports are largely captured by the predicted models in the pre-trade conflict period. From the start of the trade conflict, the actual value and volume of exports to China have fallen below the predicted values. Conversely, the actual value and volume of exports increased above the predicted values in summer 2019, continuing above through the Phase One Trade Deal; the actual value still remains within the 95% confidence bounds. Then, as the trade war cooled, the predicted volume and value of exports closely followed the actual values from late spring to summer 2020. A similar pattern in the value and volume of exports to other regions is observed, with the predicted model largely capturing the patterns in the data observed before the trade war. After July 2018, the model fails to show any statistically significant change in the value of exports to other regions, as shown by the actual value

[4]Figure 3 only shows from 2015 onward. Models were trained from 2010.

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Table 1. Accuracy table

Value of exports to China					
Model	MAE	MAPE	SMAPE	RMSE	R-sq
SARIMA(2,0,0)(2,1,0)[12]	0.329	52.93	84.83	0.393	0.880
SARIMA(2,0,0)(2,1,0)[12] XGBoost	0.317	45.77	64.41	0.379	0.881
Prophet	0.337	64.46	93.98	0.408	0.883
Prophet XGBoost *	**0.274**	**34.16**	**39.22**	**0.353**	**0.884**
Elastic Net	0.374	54.52	75.23	0.480	0.832
ETS(A,N,A)	0.312	45.97	60.93	0.380	0.870
Volume of exports to China					
Model	MAE	MAPE	SMAPE	RMSE	R-sq
SARIMA(2,0,0)(0,1,0)[12]	0.744	50.53	84.06	0.929	**0.906**
SARIMA(2,0,0)(0,1,0)[12] XGBoost	0.644	31.99	38.81	**0.825**	**0.906**
Prophet	0.712	34.65	32.80	1.003	0.873
Prophet XGBoost *	**0.655**	**30.22**	**31.43**	0.881	0.893
Elastic Net	0.882	55.35	82.04	1.131	0.856
ETS(A,N,A)	0.904	45.65	69.46	1.223	0.880
Value of exports to other regions					
Model	MAE	MAPE	SMAPE	RMSE	R-sq
SARIMA(1,0,0)(2,1,0)[12]	0.209	24.90	30.25	0.277	0.328
SARIMA(1,0,0)(2,1,0)[12] XGBoost	0.202	23.36	24.80	0.260	0.323
Prophet	**0.200**	23.74	28.42	0.270	0.353
Prophet XGBoost *	0.202	**22.92**	**24.52**	**0.258**	**0.364**
Elastic Net	0.214	26.30	30.82	0.278	0.311
ETS(M,N,M)	0.204	23.79	27.91	0.267	0.334
Volume of exports to other regions					
Model	MAE	MAPE	SMAPE	RMSE	R-sq
SARIMA(1,0,0)(2,1,0)[12]	0.555	26.81	33.47	0.725	0.324
SARIMA(1,0,0)(2,1,0)[12] XGBoost	0.454	21.34	21.60	0.593	0.326
Prophet	0.444	21.21	26.40	0.638	**0.381**
Prophet XGBoost *	**0.383**	**17.82**	**18.56**	**0.514**	**0.381**
Elastic Net	0.595	29.32	36.11	0.739	0.349
ETS(M,N,M)	0.597	27.60	31.74	0.737	0.346

Note: In the accuracy table, bold numbers indicate the best-performing model for each respective metric. The starred model indicates the best-performing model overall.

of exports staying within the 95% confidence interval of the predicted values. The actual volume exceeds the 95% confidence interval on the predicted volume, suggesting a significant increase in the volume of exports to other regions early in the trade conflict. Finally, the actual and predicted value and volume of exports to other regions aligned after the Phase One trade deal was put in place in January 2020.

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Figure 3. Actual and predicted value and volume of soybean exports using prophet with XGBoost. *Note:* Shaded area = 95% confidence intervals. The red dotted line represents forecasted exports. The solid gray line represents actual exports.

The disparity between forecasted and actual values is shown in Figure 4 to further illustrate the impact of the tariffs and trade war on the US soybean exports. Figure 4 illustrates differences between forecasted and actual exports with more severe negative values indicating periods of losses identified by actual export value or volume falling below the predicted levels. Specifically, deviations from the observed values (horizontal line) are plotted. This allows for a visual comparison of the model's accuracy by illustrating the actual values and volume of exports before the trade war. Once the trade conflict began, a drop in the value and volume of exports to China was observed, with approximate forecasted losses of up to $3 billion in value or approximately 6,000 metric tons in volume. As noted previously, the volume of exports to other regions increased slightly following the start of the trade conflict at less than 2,000 metric tons.

Confidence intervals in Figure 4 are displayed as the shaded areas. The magnitude of the difference between actual and predicted values and the associated confidence interval on point estimate vary between measures of soybean exports. As noted, there was an increase in the actual value of exports to other regions following the start of the Trade War, as well as in the actual value and volume of exports to China in the summer of 2019 before the Phase One trade deal, these increases did not fall outside of the confidence intervals. As such, while they may have somewhat helped restore the soybean trade, they did not seem to significantly offset losses from previous months.

The monthly estimated differences between the actual and forecasted values and volumes of soybean exports to China and other regions are shown in Table 2. Export losses primarily occurred during the first year of the trade conflict, from June 2018 to June 2019, with primary losses from the value and volume of exports to China. Then, in accordance with the seasonal pattern observed prior to tariffs, continued losses were observed in late 2019 and into early 2020 in terms of both the value and volume of exports to China. As noted previously, while efforts were made to increase the soybean trade value and volume to other regions, only the volume of exports

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Figure 4. Difference between actual and forecasted value and volume of soybean exports using prophet XGBoost. *Note:* The red dotted line represents the deviations from the observed values. Deviations are calculated as the difference between actual exports [value and volume] minus forecasted exports. The shaded area represents the 95% confidence intervals.

to other regions was statistically significant early on in the trade war, from July 2018 through November 2018. Notably, the volume of exports to other regions was statistically significantly lower than forecasted in May 2019.

The cumulative estimated reduction in exports during the 24-month period from July 2018 to June 2020 is presented in Table 2 and is derived from the sum of all statistically significant differences as depicted in Figure 4. In total, over the 2-year period, the reduction in exports amounts to $20.2 billion for the value of exports to China and is not offset by any (statistically) significant increases in the value of exports to other regions. Moreover, the volume of exports to China, as calculated by the sum of all statistically significant differences, amounts to a reduction in exports of 37,210 metric tons, which is offset by an increase in exports to all other regions of 4,400 metric tons. These losses represent direct economic setbacks for the industry and domestic producers due to lower-than-anticipated export volumes.[5] Results are presented over the entire trade war period, through six months following the Phase One trade deal. For this reason, annual losses are approximately $10.16 billion in total value or 16,410 metric tons in total volume. Previous studies place the estimated losses around $10.7 billion (Grant et al., 2021).

All models captured the seasonality of the data, with Prophet XGBoost most accurately estimating the large swings in monthly year-to-year exports as discussed in Table 1. Disparities between the actual data (black line) and forecasted (red line) emerged as exports fell due to the trade war (Figure 3). As seen in Figure 4, the model Prophet XGBoost was able to capture the variability and seasonality of the export volume and value data before the trade conflict, displaying little difference between the forecasted model and actual data. These forecast models utilizing ML techniques incorporate complex time-series processes and can provide precise monthly estimates of changes to export patterns for data series that vary in terms of seasonality and magnitude.

[5]However, as stated previously, these estimated reductions in export volume and value do not include payments made to producers and the soybean industry from the Market Facilitation Program.

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Table 2. Forecasted loss in soybean volume and value exported to China and other regions during trade war

	Value of export to China	Value of export to other regions	Volume of export to China	Volume of export to other regions
Month-year	Billion dollars		1000 metric ton	
Jul-18	−0.39	0.15	−0.63	1.14*
Aug-18	−0.53	0.07	−0.93	1.01
Sep-18	−0.95*	0.08	−1.87*	1.12*
Oct-18	−3.06*	0.17	−6.54*	1.78*
Nov-18	−3.08*	0.06	−6.61*	1.43*
Dec-18	−2.14*	−0.22	−4.50*	0.58
Jan-19	−1.52*	−0.42	−2.88*	−0.02
Feb-19	−0.84*	−0.58	−1.18	−0.65
Mar-19	−0.39	−0.51	−0.30	−0.54
Apr-19	−0.27	−0.49	−0.28	−0.57
May-19	0.04	−0.65	0.55	−1.07*
Jun-19	0.26	−0.56	1.12	−0.82
Jul-19	0.23	−0.41	1.04	−0.44
Aug-19	0.40	−0.33	1.63	−0.17
Sep-19	−0.63	−0.06	−1.01	0.59
Oct-19	−2.58*	−0.12	−5.31*	0.74
Nov-19	−1.53*	−0.70	−2.43*	−0.99
Dec-19	−1.19*	−0.53	−1.95*	−0.52
Jan-20	−1.25*	−0.51	−2.30*	−0.44
Feb-20	−1.37*	−0.63	−2.82*	−0.89
Mar-20	−0.83*	−0.54	−1.71	−0.73
Apr-20	−0.31	−0.55	−0.54	−0.82
May-20	−0.34	−0.49	−0.64	−0.65
Jun-20	−0.25	−0.58	−0.36	−0.97
Sum[1]	−20.32	–	−37.21	4.40

Losses are calculated as actual exports minus forecasted exports from Prophet with XGBoost errors Numbers with * indicate statistical significance at the 5% level.
[1]Sum of only statistically significant estimates.

When included, it may be inferred that changes to existing trade patterns from unprecedented policies or shocks can be modeled with increased precision through traditional forecasting procedures that include ML techniques.

6. Summary and concluding remarks

This study provides a comprehensive estimation of the effects of retaliatory tariffs by China on US soybean exports. By considering a diverse array of forecasting models and additionally including changes in soybean exports to other regions, we derive estimates of the impact of retaliatory tariffs

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

on the value and volume of soybean exports, shedding light on the impact of retaliatory tariffs on international trade patterns. The inclusion of both traditional econometric methods and advanced ML techniques mitigated potential biases, providing enhanced and robust findings. Notably, this study provides a novel effort to couple outcomes from conventional econometric forecasting and ML models to enhance and improve the measurement of loss estimates.

Results take into account substantial declines in soybean exports following the imposition of tariffs, with total estimated losses of $20.32 billion in export value or 32,810 metric tons over a 24-month period, even when the effect of exports to other regions is included. These estimates are consistent with previous studies and show the potential use of ML techniques to incorporate complex trade patterns in the estimation process. This study emphasizes the benefit of enhanced methods for estimating the impacts of trade policy and its direct consequences on all pertinent stakeholders involved.

Despite the significance of our findings, there are certain limitations in both the conceptualization of international trade patterns and the modeling approach that warrant acknowledgment. First, while this study focuses on estimating changes in US soybean exports during the recent trade conflict, it does not account for other factors that may have mitigated or influenced the impact of reduced soybean exports. Specifically, we do not incorporate the effects of market facilitation payments or other strategies designed to alleviate the impact on producers. Furthermore, our analysis does not consider export volume or value for competing commodities, nor does it explore competition for acreage among US crops. While outside the scope of this article, it is important to note that producers typically adjust their planting decisions based on relative price signals from different crops. These decisions could affect the overall availability of soybeans for export and, by extension, influence prices and international trade flows.

Additionally, while our study employed a broad range of model specifications, including both ML techniques and traditional econometric forecasting models, further refinements could help narrow the range of estimated losses. This modeling effort sought to address the challenges inherent in applied forecasting for complex, asymmetric time-series processes. As demonstrated, ML methods can offer valuable insights, but it is crucial to also consider the strengths and limitations of any modeling approach. Specifically, the methodology we used relies on past data to forecast future changes in export volume and value. This means that the estimated changes are not based on a structural model that accounts for dynamic adjustments over time. Future research could apply this approach to other commodities or policy situations where the strengths of ML approaches to handle heteroskedasticity, zero value observations, and complex time-series processes may be beneficial.

The insights gleaned from this study can serve to gain an improved understanding of estimated potential losses in the event of future tariff impositions by other nations on US exports and thus provide benefit in assessing implications in the formulation of domestic trade policies. By improving the assessment of the economic consequences that arise from trade disputes, our study contributes to an enhanced understanding of the dynamics governing international trade relations. This, in turn, underscores the imperative of formulating informed policy to mitigate adverse outcomes for domestic industries and stakeholders.

References

Adjemian, M.K., A. Smith, and W. He. "Estimating the market effect of a trade war: The case of soybean tariffs." *Food Policy* **105**(2021):102152 doi:10.1016/j.foodpol.2021.102152.

Athey, S., and G. Imbens. "Machine learning methods that economists should know about." *Annual Review of Economics* **11**,1(2019):685–725 doi:10.1146/annurev-economics-080217-053433.

Baier, S., and N. Regmi. "Using machine learning to capture heterogeneity in trade agreements." *Open Economies Review* **34**,4(2023):863–94 doi:10.1007/s11079-022-09685-3.

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Bown, C. (2022). *US-China Phase One Tracker: China's Purchases of US Goods*. Peterson Institute for International Economics (PIIE). Internet site: https://www.piie.com/research/piie-charts/us-china-phase-one-tracker-chinas-purchases-us-goods, accessed on April 9, 2024.

Breinlich, H., V. Corradi, N. Rocha, M. Ruta, J. SAntos Silva, and T. Zylkin, *Machine Learning in International Trade Research - Evaluating the Impact of Trade Agreements*.CEPR Discussion Paper No. DP17325. Center for Economic Policy Research. Internet site: https://ssrn.com/abstract = 4121518

Bukowski, M., and A.M. Ates. (2023). *U.S. soybean exports and share of production exported*. U.S. Department of Agriculture, Economic Research Service. https://www.ers.usda.gov/data-products/chart-gallery/chart-detail?chartId = 106804

Bukowski, M., and B. Swearingen. (2023). *Oil crops outlook: November 2023* (Report No. OCS-23k). U.S. Department of Agriculture, Economic Research Service.

Bukowski, M., and B. Swearingen. (2025). *Soybeans and Oil Crops-Oil Crops Sector at a Glance*. U.S. Department of Agriculture, Economic Research Service. https://www.ers.usda.gov/topics/crops/soybeans-and-oil-crops/oil-crops-sector-at-a-glance

Carter, C., and S. Steinback. (2020) "The Impact of Retaliatory Tariffs on Agricultural and Food Trade." *NBER Working Paper Series 27147*. https://www.nber.org/papers/w27147, accessed on April 3, 2024.

Chen, T., and C. Guestrin. (2016). XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Internet site: https://dl.acm.org/doi/10.1145/2939672.2939785, accessed on April 1, 2024.

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, Y. Li, and J. Yuan. (2024). xgboost: Extreme Gradient Boosting. CRAN R-Project. Internet site: https://cran.r-project.org/web/packages/xgboost/xgboost.pdf, accessed on April 1, 2024.

Cipollina, M., and L. Salvatici. "Reciprocal trade agreements in gravity models: A meta-analysis." *Review of International Economics* **18**,1(2010):63–80 doi:10.1111/j.1467-9396.2009.00877.x.

Coble, K., A. Mishra, S. Ferrel, and T. Griffin. "Big data in agriculture: A challenge for the future." *Applied Economics Perspectives and Policy* **40**,1(2018):79–96 doi:10.1093/aepp/ppx056.

Dancho, M. (2023). modeltime: The tidymodels extension for time series modeling. CRAN R-Project. Internet site: https://cran.r-project.org/web/packages/modeltime/modeltime.pdf.

Fajgelbaum, P., P. Goldberg, P. Kennedy, A. Khandelwal, and D. Taglioni. "The US-China trade war and global reallocations." *American Economic Review: Insights* **6**,2(2024):295–312 doi:10.1257/aeri.20230094.

Firedman, J., T. Hastie, and R. Tibshirani. "Regularization paths for generalized linear models via coordinate descent." *Journal of Statistical Software* **33**,1(2010):1–22 doi:10.18637/jss.v033.i01.

Flaaen, A., and J. Pierce. "*Disentangling the Effects of the 2018-2019 Tariffs on a Globally Connected U.S. Manufacturing Sector. Finance and Economics Discussion Series (FEDS)*. Washington: Board of Governors of the Federal Reserve System. DOi: 10.17016/FEDS.2019.086.

Gale, F., C. Valdes, and M. Ash. "*Interdependence of China, United States, and Brazil in Soybean Trade*. United States Department of Agriculture Economic Research Service, Oil Crops Outlook No. OCS-19F-01. https://www.ers.usda.gov/publications/pub-details/?pubid = 93389, accessed on June 20, 2024.

Goodwin, P., and R. Lawton. "On the asymmetry of the symmetric MAPE." *International Journal of Forecasting* **15**,4(1999):405–8 doi:10.1016/S0169-2070(99)00007-2.

Gopinath, M., F. Batarseh, and J. Beckman. (2020). Machine learning methods that economists should know about." *NBER Working Paper No. W27151*. https://ssrn.com/abstract = 3603781, accessed on July 28, 2023.

Gopinath, M., F. Batarseh, J. Beckman, A. Kulkarni, and S. Jeong. "International agricultural trade forecasting using machine learning." *Data and Policy* **3**,1(2021):1–20 doi:10.1017/dap.2020.22.

Grant, J., S. Arita, C. Emlinger, R. Johansson, and C. Xie. "Agricultural exports and retaliatory trade actions: An empirical assessment of the 2018-2019 trade conflict." *Applied Economics Perspective and Policy* **43**,2(2021):619–40 doi:10.1002/aepp.13138.

Gupta, V., and E. Kumar. "2021, Review on Machine Learning Techniques for International Trade Trends. *3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. DOI: 10.1109/ICAC3N53548.2021.9725585.

Hyndman, R., M. Akram, and B. Archibald. "The admissible parameer space for exponential smoothing models." *Annals of Statistical Mathematics* **60**,2(2008):407–26 doi:10.1007/s10463-006-0109-x.

Hyndman, R., and G. Athanasopoulos. (2021). *Forecasting: Principles and Practice*, 3rd edn. Melbourne, Austrailia: OTexts. https://otexts.com/fpp3/.

Hyndman, R., and Y. Khandakar. "Automatic time series forecasting: The forecast pakcage for R." *Journal of Statistical Software* **27**,3(2008):1–22 doi:10.18637/jss.v027.i03.

Hyndman, R., A. Koehler, R. Snyder, and S. Grose. "A state space framework for automatic forecasting using exponential smoothing methods." *International Journal of Forecasting* **18**,3(2002):439–54 doi:10.1016/S0169-2070(01)00110-8.

Jošić, H., and B. Žmuk. "A machine learning approach to forecast international trade: The case of Croatia." *Business System Research* **13**,3(2022):144–60 doi:10.2478/bsrj-2022-0030.

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press

Kansas State University. U.S. supply and demand (wasde) spreadsheets, 2024. Internet site: https://www.agmanager.info/grain-marketing/grain-supply-and-demand-wasde/us-supply-and-demand-wasde-spreadsheets.

Kareem, F.O., and O.I. Kareem. " *Specification and Estimation of Gravity Models: A Review of the Issues in the Literature*. European University Institute, Robert Schuman Centre for Advanced Studies Global Governance Programme, EUI Working Paper RSCAS 2014/74. Internet site: http://www.eui.eu/RSCAS/Publications/.

Lütkepohl, H., and M. Krätzig, eds. (2004). *Applied Time Series Econometrics*. New York, NY: Cambridge University Press.

Morgan, S., S. Arita, J. Beckman, S. Ahsan, D. Russell, P. Jarrel, and B. Kenner. (2022). The Economic Impacts of Retaliatory Tariffs on US Agriculture. Economic Research Report No. ERR-304. Internet site: https://www.ers.usda.gov/publications/pub-details/?pubid = 102979, accessed on April 2, 2024.

Mullainathan, S., and J. Spiess. "Machine learning: An applied econometric approach." *Journal of Economic Perspective* **31**,2(2017):87–106 doi:10.1257/jep.31.2.87.

Ramasubramanian, K., and A. Singh, *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*. 2ndedn. APress, 2018.

Ribera, L., F. Adcock, and J. Mu. (2018). *Estimated Economic Impacts of Retaliatory Tariffs by China and Mexico on U.S. Dairy Products*. CNAS Report 2018-3, Centre for North American Studies. Internet site: https://agecoext.tamu.edu/wp-content/uploads/2021/02/CNAS-pub-Estimated-Economic-Impacts-of-Retaliatory-Tariffs-September-2018.pdf.

Sabala, E., and S. Devadoss. "Impacts of Chinese tariff on world soybean markets." *Journal of Agricultural and Resource Economics* **44**,2(2019):291–310 doi:10.2307/26797558.

Sei, J., M. Gopinath, A. Kulkarni, and F. Batarseh. "Deploying machine learning methods to predict global trade patterns: the case of beef." *Journal of the ASABE* **67**,1(2024):219–32 doi:10.13031/ja.15619.

Storm, H., K. Baylis, and T. Heckelei. "Machine learning in agricultural and applied economics." *European Review of Agricultural Economics* **47**,3(2020):849–92 doi:10.1093/erae/jbz033.

Sukar, A., and S. Ahmed. "Rise of trade protectionism: The case of US-Sino Trade War." *Transnational Corporations Review* **11**,4(2019):279–89 doi:10.1080/19186444.2019.1684133.

Taylor, S., and B. Letham. "Forecasting at scale." *The American Statistician* **72**,1(2018):37–45 doi:10.1080/00031305.2017.1380080.

Taylor, S., and B. Letham. (2022). prophet: Automatic forecasting procedure. CRAN R-Project. Internet site: https://cran.r-project.org/web/packages/prophet/prophet.pdf.

Tortajada, C., and H. Zhang. "Policies and politics: Effects on US-China soybean trade." *Georgetown Journal of International Affairs, Walsh School of Foreign Service, Georgetown University*, 2022. Internet site: https://gjia.georgetown.edu/2022/10/26/policies-and-politics-effects-on-us-china-soybean-trade/, accessed on Feb 21, 2024.

Williams, B., and K. Hammand. (2020). *Escalating U.S. Tariffs: Affected Trade. IN10971*. Congressional Research Service. Internet site: https://sgp.fas.org/crs/row/IN10971.pdf, accessed on Mar 29, 2024.

Xing, Y., D. Li, and C. Li. "Time series prediction via elastic net regularization integrating partial autocorrelation." *Applied Soft Computing* **129**(2022):109640 doi:10.1016/j.asoc.2022.109640.

York, E. "Tracking the economic impact of U.S. tariffs and retaliatory actions." *Tax Foundation Research*, 2023. Internet site: https://taxfoundation.org/research/all/federal/tariffs-trump-trade-war/, accessed on February 21, 2024.

Cite this article: Thayer, A.W., P. Sun, H.A. Tejeda, and M.-K. Kim (2025). "Exploring the Impact of China's Retaliatory Tariffs on US Soybean Exports with Machine Learning Techniques." *Journal of Agricultural and Applied Economics* **57**, 240–258. https://doi.org/10.1017/aae.2025.6

https://doi.org/10.1017/aae.2025.6 Published online by Cambridge University Press