Application of Machine Learning to International Market and Trade Analysis

by

Sei Jeong

(Under the Direction of Gopinath Munisamy)

ABSTRACT

The main focus of the three studies in this dissertation is the application of advanced techniques including machine learning methods and finer data to model economic relationships: price, demand and trade patterns. Agricultural products and food are absolutely necessary for human survival, and unexpected shocks pose a threat to food availability and access in many developing countries. Along with the growth of population and income, trade in agricultural commodities has witnessed dramatic increases in recent years. In the first chapter, the role of international price volatility and inventories on domestic market price dynamics in the case of agricultural commodities is investigated using Least Absolute Shrinkage and Selection Operator (LASSO) methods. The second chapter attempts to explain global beef trade by comparing traditional econometric methods, i.e., Poisson Pseudo Maximum Likelihood (PPML), and ML techniques. The third chapter employs finer data, Harmonized System 6-digit products, to model import demand by major sources in a commodity (beef) subjected to frequent disease outbreaks.

INDEX WORDS: International trade, Machine learning, Agricultural commodities,

Application of Machine Learning to International Market and Trade Analysis

by

Sei Jeong

B.A. in Economics, Konkuk University, 2013

M.A. in Economics, Konkuk University, 2015

M.S. in Agricultural Economics, Iowa State University, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

Application of Machine Learning to International Market and Trade Analysis

by

Sei Jeong

Major Professor:	Gopinath Munisamy
Committee:	Chen Zhen
	Berna Karali

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

DEDICATION

I dedicate this to my mother, Sunjung Yu, who I love the most in the world. Her prayers and dedication have brought me here. Thank you, Lord, for being with me in every situation.

ACKNOWLEDGEMENTS

I appreciate my major advisor, Gopinath Munisamy, for his invaluable support and guidance during the years of my Ph.D. program. He always gave me the opportunity to pursue exciting research ideas and methods. I would also like to thank Dr. Chen Zhen and Berna Karali for insightful feedbacks.

I would like to express my deepest gratitude to my family. I could not have made it without their love, prayer and support.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The main focus of the three studies in this dissertation is the application of machine learning methods to the analysis of international markets, trade and demand. Agricultural products and food are absolutely necessary for human survival, and unexpected shocks pose a threat to food availability and access in many developing countries. Along with the growth of population and income, demand for and trade in agricultural commodities has witnessed dramatic increases in recent years. Given the increasingly globalized agricultural markets, changes in the production or consumption of agricultural commodities in one place can significantly affect other economies including the risk of price (income) volatility and food security (Schierhorn et al., 2016). Each country needs to be account for external shocks as well as trade patterns for its stable supply and demand of agricultural commodities and foods.

Over the past decade, conventional statistical and econometric techniques have been challenged in modeling economic relationships due to the emerging volume of finer data and complexities in such relationships (Varian, 2014). Machine learning (ML) has been offered as an alternative to address many of these challenges including dealing with the sheer volume of data (James et al., 2013; Qiu et al., 2016; Batarseh and Yang, 2017; L'heureux et al., 2017; Athey et al., 2019; Gopinath et al., 2021). In each of the three chapters, ML techniques are introduced and compared with conventional methods in the context of international market, trade and demand analysis.

In the first chapter, the role of international price volatility and inventories on domestic market price dynamics in the case of agricultural commodities is investigated using Least Absolute Shrinkage and Selection Operator (LASSO) methods. With a first-stage LASSO estimator to identify the best instrument set, a nonlinear approach is used to estimate price dynamics. The expectation is that international market information plays a critical role in domestic market price dynamics. Both international price volatility and international inventories are likely to shape domestic prices, but with differences in the magnitude of their effects.

The second chapter attempts to explain global beef trade by comparing traditional econometric methods, i.e., Poisson Pseudo Maximum Likelihood (PPML), and ML techniques. Then alternative policy scenarios are modeled to predict future patterns of trade. This study makes a modest attempt to compare the advantages and disadvantages of PPML methods and ML models. The main advantage of the PPML appears to be its strong predictive power, but only with the inclusion of a large number of fixed effects. ML models have the advantage include selecting explanatory variables when a long list of explanatory variables is to be considered and out-of-sample predictions.

The third chapter analyzes the world's beef import demand at a disaggregated level of beef commodities through a multi-stage approach. The multi-stage budgeting approach begins with an estimate for the demand function for imported beef on world imports level by including cattle diseases such as Foot and Mouth Disease (FMD) and Bovine Spongiform Encephalopathy (BSE) as well as FTA into the demand system. All three studies are expected to provide insights into the advantages of finer data and advanced techniques in modeling economic relationships: price, demand and trade.

CHAPTER 2

INTERNATIONAL MARKET INFORMATION AND AGRICULTURAL COMMODITY PRICE DYNAMICS[1]

[1] Jeong, S., and Gopinath, M. Accepted by *Journal of Agribusiness in Developing and Emerging Economies* on August 29th 2022. It will be published soon, and the article is open access.

Abstract

This study investigates the role of international price volatility and inventories on domestic market price dynamics in the case of agricultural commodities. A structural model is employed to uncover relationships among commodity price, price volatility, inventories, and convenience yield. Monthly producer price data along with annual data on trade, consumption, inventories, and tariffs for 71 countries and 13 commodities covering 2010–2019 are assembled to estimate the model. With a first-stage Least Absolute Shrinkage and Selection Operator (LASSO) estimator to identify the best instrument set, a nonlinear approach is used to estimate the model. Results show that international market information plays a critical role in domestic market price dynamics. International price volatility has a stronger effect on domestic prices than that of international inventories. Although hypotheses exist that global market information (volatility and inventories) helps countries manage domestic commodity prices, there have been limited studies on this relationship, especially with a structured model and cross-country data.

Introduction

Price fluctuations of agricultural commodities commonly arise in the context of many external shocks such as weather patterns, sudden changes to income or wealth, and political conflicts. However, large and unexpected fluctuations can have a negative impact on the food security of consumers, income of farmers, and welfare of entire nations. For some developing countries, food price volatility is a major concern since it undermines their perceived national food self-sufficiency. Moreover, price volatility can be disruptive to supply and demand chains in essential commodity markets.

Figure 2.1 shows that the global food price index surged in 2008 and along with its sub-indices became more volatile relative to those in the early 2000s (FAO, 2020). Also, according to U.S. Department of Agriculture, the average farm price of the top three U.S. field crops – corn, wheat, and soybeans – regularly rises or falls by over 10 percent from year to year (USDA, 2021). Not surprisingly, there is a sizable literature examining agricultural price dynamics or volatility.

Gouel (2012), in reviewing studies on agricultural price dynamics, offered two major explanations: cobweb logic and rational expectations. The cobweb logic is based on expectation errors, which make for complex price dynamics and possibly, chaos. Fortunately, the empirical evidence favors the rational expectations tradition of dynamics driven by real shocks. Muth (1961), one of the earliest contributions to the rational expectations literature, introduced storage in a dynamic model, where greater demand for inventories to smooth production arises when prices become volatile. Thus, an increase in volatility can lead to inventory build-ups and raise prices in the short run. Pindyck (2004) defines equilibrium inventory behavior as the solution to a stochastic dynamic problem through a structural-form inventory model. Many studies have extended the storage model to examine commodity price volatility. Deaton and Laroque (1992), employing annual price data on 13 commodities, found that there is a cut-off price above which the next period price is no longer linked to the current price. Beck (2001) tested the hypotheses that inventory carryover creates an autoregressive conditional heteroscedasticity (ARCH) process in prices and found that a serially correlated price variance process exists for storable but not for non-storable commodities. Pietola et al., (2010) also tested a similar autoregressive process and found that wheat inventories impacted price volatility, which in turn, affected prices. In addition to understanding price dynamics, many studies introduced additional factors affecting price volatility and storage. For example, Miranda and Glauber (1993) incorporated both private and government stockholding into a rational expectations model, which performed better than ARCH-type models in explaining price dynamics. Osborne (2004) found that advance information or news about future production affected competitive storage behavior and prices in the Ethiopian grain market. Lence and Hayes (2002) and Gouel (2016) examined the role of U.S. farm and Indian food policies, respectively, on price stability. Akanni (2020) examines the linkage between volatility of exchange rate and domestic food price.

As Gouel's (2012) review suggested, the majority of empirical studies, focused on individual commodities or domestic markets, favored the competitive storage model with rational expectations. However, attention to international trade in agricultural commodities, which has historically played a crucial role in reducing food insecurity and stabilizing domestic markets, has been limited in understanding commodity price dynamics (Gillson and Fouad, 2014; Matthews, 2014). For instance, China imported more than \$138 million of pork and pork products from the United States in July 2019, compared with \$41 million in January of the same year because of the former's African swine fever outbreak (USDA, 2019). This increase in imports not only lowered the pork price in China, but also raised it in the U.S. market. In 2021, the surge in Chinese feed demand and poor weather forecasts in South America led to sharp price increases in some commodity markets, while Argentina attempted to lower internal prices of commodities to ensure food security. Thus, events or information in overseas markets, e.g. inventories of foreign producers or a sudden price increase or decrease, have been found to impact commodity prices in domestic and foreign markets (Osborne, 2004). Akanni (2019) also demonstrates that the volatility of exchange rates tends to have spillover effects on domestic food prices.

The objective of this study is to investigate the effect of international market information – especially price volatility and inventories – on domestic market's commodity price dynamics. For this purpose, a storage-trade model is considered, where the equilibrium is achieved when with optimal (economic) responses of consumers and producers. In this study's context, the production decision arises from the maximization of expected profit. A storage model has been considered to be a better representation of world price dynamics than simple dynamic process (Cafiero et al., 2011; Gouel et al., 2016). Furthermore, an extension of the Pindyck's (2004) structural model of price dynamics can directly identify the relationships among price, price volatility, inventories, and convenience yield, i.e. value of the marginal unit of inventory. In particular, an augmentation of the marketing cost function of Pindyck's (2004) model allowed for capturing the effect of trade openness (i.e., tariffs limiting exposure to international markets), and other countries' price volatility and inventories (i.e., information available in domestic markets) on commodity price dynamics. The resulting empirical model includes equations for convenience yield and price formation as functions of inventories (domestic and international), price volatility (domestic and international), trade openness, and other controls. Monthly producer price data from FAO along with data on trade, consumption, inventories, and tariffs for 71 countries and 13 commodities, covering 2010-2019, are used to estimate the augmented model. The next section presents the methodological approach followed by a description of data, discussion of results and summary and conclusions.

Approach

Theoretical Model

The theoretical framework for assessing the impact of international market information on domestic market price dynamics is an extension of Pindyck's (2004) model relating the dynamics of inventories, spot and futures prices, and the level of volatility. For any given country i, the total economic cost of agricultural commodity production (C^p), marketing (C^m), and storage (kn_t) is given by:

$$TC_t = C^p(x_t) + C^m(p_t, \sigma_t^2, n_t) + kn_t \quad (1)$$

where x_t refers to the output, p_t is the price of the output, σ_t^2 is the price volatility, and n_t refers to the inventory level[2]. The first and last components of equation (1) are the direct production cost and the cost of storage, respectively. The total marketing cost, C^m, includes actual or opportunity costs of activities facilitated by inventories, e.g., costs of adjusting production over time, delivery scheduling, and stock-out avoidance. The per-unit storage cost (k) is assumed to be constant. With the objective to assess the effects of international market information on domestic price dynamics, the marketing cost component in equation (1) is rewritten as:

$$TC_t = C^p(x_t) + C^m(p_t, \sigma_t^2, n_t, tariff_t, \sigma_{-it}^2, n_{-it}) + kn_t \quad (2)$$

[2] Pindyck's (2004) total economic cost has four components including opportunity cost of producing now, rather than waiting to see how prices evolve. This component is not considered here because current agricultural production does not necessarily increase the next period's production costs and does not create option values comparable to those of depleting nonrenewable natural resources (see also Pietola et al, 2010).

For country i, the marketing cost (C^m) now includes two sets of variables: domestic information on price, price volatility and inventories and other countries' price volatilities (σ^2_{-it}) and inventories (n_{-it}). The international market information including inventories and price volatilities are assumed to be exogenous to country i. Equation (2) now hypothesizes that information on international markets, e.g. reported in World Agricultural Supply and Demand Estimates (WASDE), affects domestic markets primarily through the marketing-cost component. Moreover, it is highly likely that international market information will affect domestic prices depending on country i's openness to trade. Hence, some information on a country's trade openness is needed in equation (2) to allow for international volatility and inventories to affect commodity price dynamics. Unlike the anticipated effects of p_t, σ^2_t and n_t as detailed below, the impact of σ^2_{-it} and n_{-it} cannot be assigned *a priori*, but the anticipation is that they play a role similar to that of domestic volatility and inventories, but conditional on being open to trade.

Assuming that price-taking and risk-neutral producers maximize the present value of the expected flow of profits by choosing output (x_t) and inventory (n_t) levels:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t (p_t q_t - TC_t) \quad s.t. \;\; \Delta n_t = x_t - q_t \; and \;\; n_t \geq 0 \quad for\; all\; t, \tag{3}$$

where β^t is the discount factor and q_t is sales or domestic consumption. Holding n_t fixed, the change in x_t equals to the change in q_t, hence the first order conditions for maximizing with respect to x_t is:

$$p_t = \frac{\partial TC_t}{\partial x_t} \tag{4}$$

Equation (4) captures the traditional relationship that price equals marginal cost of production.

Note that the marketing cost (C^m_t) is the cost associated with delivering goods to customers, which is facilitated by inventories. Hence, it includes the cost of adjusting production, delivery scheduling and stock-out avoidance. So, maximizing equation (3) with respect to n_t and using the Euler equation yields:

$$\beta E_t[p_{t+1}] = p_t + \frac{\partial TC_t}{\partial n_t} = \beta E_t[p_{t+1}] = p_t + \frac{\partial C_t^m}{\partial n_t} + k \tag{5}$$

Equation (5) equates the expected next-period price to the sum of current price, marginal marketing cost of inventories, and storage cost. Including international market information – and assuming that the marketing cost is iso-elastic in price, the variance of log price changes, and the total inventory level as in Pindyck (2004) – the marketing cost can then be specified as:

$$\begin{aligned} &C_t^m(p_t, \sigma_t^2, n_t, tariff_t, \sigma_{-it}^2, n_{-it}) \\ &\qquad = \frac{1}{a_4 - 1}\exp(b_0) p_t^{a_1} (\sigma_t^2)^{a_2} (\sigma_{-it}^2)^{a_3 TO_t} n_t^{(1-a_4)} (n_{-it})^{a_5 TO_t}, \end{aligned} \tag{6}$$

where $a_4 > 1$, and TO_t is a measure of trade openness described in the next section. Then, the value of the marginal unit of inventory also referred to as marginal convenience yield is $\psi_t(p_t, \sigma_t^2, n_t, tariff_t, \sigma_{-it}^2, n_{-it}) = -\partial C_t^m / \partial n_t$. The equation for the *T*-period (net of storage costs) marginal convenience yield of a commodity is written as:

$$\psi'_{T,t} = (1 + R_{T,t}) P_t - F_{T,t}, \tag{7}$$

where $R_{T,t}$ refers to monthly interest rates, P_t refers to spot price, and $F_{T,t}$ is the futures price. The expectation for the domestic market information is that the convenience yield increases with volatility but decreases with inventories (Pindyck, 2004; Gorton et al., 2013; Sevi, 2015). That is, when price volatility is higher, it is highly likely that inventories are not sufficient, hence spot prices increase relative to future prices. On the other hand, high inventory levels imply a low probability of a stock-out. As noted earlier, the effects of international inventory and price volatility likely depend on country's engagement in global markets (*TO*) and are empirical in nature.

Empirical Strategy

While alternative methods can be employed in assessing the role of inventories and volatility on commodity price dynamics, this study follows Pindyck's (2004) structural approach to estimate domestic price and convenience yield equations (Pietola et al., 2010). First, the price equation for estimation purposes, given equation (4) is specified as:

$$p_{ij,t} = c_1 x_{ij,t} + c_2 TO_{ij,t} + c_3 Year_t + u_{ij} + \epsilon_{ij,t}, \tag{8}$$

where $x_{ij,t}$ refers to the total output of commodity j in country i. According to the storage model, the market equilibrium is defined as: $Production_t + Import_t = Demand_t + \Delta Stock_t + Export_t$. With the objective of investigating the impact of international information on price dynamics, x_t is defined in this study also as the summation of the change of inventory level ($\Delta n_{ij,t}$), domestic demand and export minus import. Since x_t refers to the calculated production rather than actual production, it can have negative values especially when the amount of import is high. When the calculated production (x_t) is negative, then it is replaced with zero[3]. The variable $TO_{ij,t}$ denotes trade openness of country i for commodity j. For estimation purposes, when the tariff rate of commodity j imposed by country i is less than the average of other countries' tariff rate (for the same commodity), then $TO_{ij,t} = 1$, otherwise, $TO_{ij,t} = 0$. Since international supply plays an important role in stabilizing domestic prices, the relative tariff rate is likely a determinant of its price depending on a country's trade openness. $Year_{ij,t}$ is a trend variable, and u_{ij} refers to a set of commodity-country fixed effects, which control for differences across commodities and countries.

[3] Empirically, production cannot be negative. Note also that $Demand_t$, defined as consumption or disappearance, is an accounting identity. Of the total 1,639 observations, 57 had negative values. Analysis including the negative values did not significantly affect reported results.

The marginal convenience yield is given by the first derivative of the marketing cost function with respect to inventories ($\psi_t = -\partial C_t^m / \partial n_t$). A log-transformation of equation (6) leads to the following specification for convenience yield:

$$\log \psi_{ij,t} = a_1 \log p_{ij,t} + a_2 \log \sigma_{ij,t}^2 + a_3 TO_{ij,t} * \log \sigma_{-ij,t}^2 - a_4 \log n_{ij,t} + a_5 TO_{ij,t} * \log n_{-ij,t} + u_{ij} + \varepsilon_{ij,t} \quad (9)$$

Note the negative sign before a_4 for the value of the marginal unit of inventories and the specification in equation (9) conditions the impact of international market information on a country's trade openness for that commodity. In essence, the argument is that a country which has a lower tariff than the rest of the world may rely more on international markets and their informational developments. Hence, the trade openness dummy ($TO_{ij,t}$) interacts with other countries' price volatilities ($\sigma_{-ij,t}^2$) and inventories ($n_{-ij,t}$) in the estimation of the effects of international information on the marginal convenience yield. The international price volatility ($\sigma_{-ij,t}^2$) is derived by calculating the average price volatility of ($N-1$) countries except country i for commodity j, where N is the total number of countries in the sample. Likewise, international inventory of country i ($n_{-ij,t}$) is the summation of the inventory data from the sample countries excluding the country indexed by i. Since not all countries provide accurate and high-quality information, it is impossible to know the "real" changes in price and inventory levels around the world. By using the average price volatility and the summation of the inventory data from the sample, it is assumed that decision makers in each country make decisions by using information available to them.

Note that equation (5) can be rewritten as $\beta E_t[c_1 \Delta x_{ij,t+1} + c_2 \Delta TO_{ij,t+1} + c_3 + \psi_{ij,t} - k_{ij} + \Delta\epsilon_{ij,t+1}] = 0$ using the defined price ($p_{ij,t}$) and convenience yield ($\psi_{ij,t}$) equations above.

Substituting the equation for convenience yield ($\psi_{ij,t} = -\partial C_{ij,t}^{m}/\partial n_{ij,t}$) into the rewritten equation results in:

$$\begin{aligned}0 = {} & c_1 \Delta x_{ij,t+1} + c_2 \Delta TO_{ij,t+1} + c_3 \\ & + \exp(u_{ij})\, p_{ij,t}^{a_1} \left(\sigma_{ij,t}^2\right)^{a_2} \left(\sigma_{-ij,t}^2\right)^{a_3 * TO_{ij,t}} \left(n_{ij,t}\right)^{-a_4} \left(n_{-ij,t}\right)^{a_5 * TO_{ij,t}} - k_{ij} \\ & + \Delta\epsilon_{ij,t+1} + \varepsilon_{ij,t}\end{aligned} \tag{10}$$

The expectation operator is dropped, instead the actual values of variables dated at $t+1$ are used. The next section presents details on price data, computation of volatility and inventories – both domestic and international.

Data

Price, Volatility and Tariff Data

Multiple sources are employed to assemble data on commodity prices, consumption, inventories, trade, and tariffs at the 4-digit Harmonized System (HS) level. The choice on commodities – almond, coffee, wheat, barley, oats, maize (corn), rice, sorghum, soybean, rapeseed, sunflower seed, palm oil, and cotton – is based on availability of price data as well as a futures market for forwards and options as reported by Bloomberg. Livestock commodities are not considered because of matching issues, e.g. meat of bovine (0201) reported at the HS4 level in prices and production do not match the available futures price from the Bloomberg for frozen beef.[4] The chosen time period – 2010 through 2019 – is due to the availability of monthly producer price data from United Nations' Food and Agricultural Organization (FAO). Such data are available for 81 countries, but many are missing data on not only prices but also inventories and consumption. After merging data from different sources, the sample data contain 71 countries

[4] This matching problem leads to large gaps between producer and futures prices yielding unrealistic convenience yields for livestock commodities.

and 13 commodities during 2010–2019. Most of data for European Union (EU) countries are not individually reported, hence EU is considered as one country in the analysis. The list of countries and corresponding price data availability are in Table A1.

Monthly and annual producer price data are obtained from FAO. The monthly data are used to calculate annual price volatility. In specific, the annual price volatility is computed using the standard deviation of the monthly producer price, and the calculated price volatility is then normalized by the mean. The monthly producer price is reported only in local currency. As will be mentioned later, all empirical equations are estimated simultaneously in this study, hence exchange rates from the Global Economic Monitor are used to convert the locally denominated price to U.S. dollars per metric ton to maintain consistency among the data sets.

The raw data of these monthly producer prices have several problems, e.g. unusually large price values for some years. For example, the monthly producer price data show that the maize price in Bangladesh in January 2011 is ৳75,580 in Bangladeshi taka (BDT), which is $1,038 in US dollars when the exchange rate is applied, while the annual producer price data (2011) indicate that the price of maize in local currency and US dollar is ৳13,650 and $184.1, respectively. Hence the ৳75,580 are replaced with ৳15,580 to match the average price during 2011. When the price data have abnormally high or low values, the values are replaced comparing the annual average to other monthly prices within that year. Ecuador's missing data are for a few months are filled using the corresponding inflation rate.

Annual trade (import and export), consumption, and inventory data are obtained from USDA's Foreign Agricultural Service. The inventory refers to the beginning stock data, and the unit for every commodity is converted to metric ton. Tariff data for each commodity and each country are obtained from the Market Access Map based on HS4 level. The tariff has missing data for some years, so the missing data are replaced by the data from the observable preceding period following the approach of Jayasinghe et al. (2010).

Convenience Yield and International Market Information

To compute the convenience yield described in equation (7), futures price data for the 13 commodities and interest rate data of each of the 71 countries are taken from Bloomberg and International Financial Statistics (IFS), respectively. Missing interest rates are replaced with the average interest rate of all countries in the sample.

The futures price is monthly data from the nearest futures contracts for each commodity. For example, contract for corn in CBT is held in March, May, July, September, and December. The contract with the closest settlement date is called the nearby futures contract, hence the futures price of corn in January 2010 is the settlement price of futures contracts in March 2010. Not every country in the sample has a futures market for commodities considered in the study. Therefore, the futures price from the geographically closest futures market is used to derive the convenience yield data for each country. For instance, in the case of barley, futures data from ICE Futures Canada are used for Canada and the United States. Those from Matba-Rofex, Argentina are used for most of South American countries including Bolivia, Colombia, Ecuador and Uruguay. If there are several futures markets in a country, a rare instance, then the average prices from these markets are used to represent of the country's global contracts. The monthly producer price data are used to proxy spot price data. Then, a simple average is used to convert the monthly convenience yield to annual data. As in Pindyck (2004), the constant storage cost, k, is estimated as $\hat{k} = |\min \psi'_{T,t}|$, and the gross marginal convenience yield is computed as $\psi_{T,t} = \psi'_{T,t} + \hat{k}$.

Table 2.1 shows the summary of statistics of sample data. The total number of observations is 1,639 with 13 agricultural commodities and 71 countries covering 2010–2019. The average sample length among the 13 commodities is 5.3 years. The mean of convenience yield is 300.49, which appears relatively larger than expected. The high convenience yield values in Table 1 are primarily due to inclusion of almond, and Japan and South Korea, which have higher producer prices than other countries. For example, the price of rice (1006) in those two countries in 2018 was $2,186.15 and $2,008.39, respectively[5]. Moreover, the sample average of rice price for the two countries is $2,069.56 per metric ton, while that of all other countries is $457.83. In addition, the two countries do not have their own rice futures markets. The closest futures market is in China, and the Chinese average rice price is also much lower than them with $442.22 during the same period. Excluding data from Japan and Korea, the average and maximum of convenience yield for all commodities drops to 252.6 and o 2,798.74, respectively. The wide variation in convenience yield motivated empirical analysis with and without Japan and Korea detailed in the next section. In addition to the whole sample – 13 commodities for 71 countries during 2010–2019 – two sub-samples are considered for the empirical analysis: grains and corn. Corn has the largest number of observations among all commodities for a separate analysis.

Estimation Methods

The equation (8) and (9) contain four endogenous variables: output (x), convenience yield (ψ), prices (p), and inventories (n). With international market information, the set of instruments available to handle this endogeneity is fairly large. In addition to the instruments from Pindyck

[5] Another factor influencing the high average convenience yield in Table 1 is the almond market. Almond price data are available for European Union and Turkey, while its futures market is in the United States.

(2004) – exchange rate index, trade-weighted U.S. dollar index, price indexes, bond yields, 3-month treasury bill, 1-year treasury bill, one- and two-period lagged endogenous variables – this study employed CPI, inflation, real interest rate of countries in the sample, and one- and two-period lagged values of those variables. When using many instrumental variables, conventional asymptotics may provide poor approximations to the sampling distributions of the resulting estimators (Hansen et al., 2008). The most often prescribed solution is to use the Least Absolute Shrinkage and Selection Operator (LASSO), which minimizes the sum of squared results but with a penalty for model size through the sum of the absolute values of the coefficients.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \tag{11}$$

Belloni et al. (2012) showed that LASSO-based procedures yielded first-stage predictions that provided good approximations to the optimal instruments when the number of available instruments is large. Another advantage of the use of LASSO as a variable selection technique is that any prior knowledge about the identity of the most relevant instruments is not necessary. Once LASSO provides the first-stage predictions, alternative approaches are employed to test and identify the impact of international market information on agricultural price dynamics.

Results

Table 2.2 holds the results from estimating the nonlinear models for three cases – 13 commodities, grains subsample, and corn subsample – using the two-stage least squares approach with fixed effects and the first-stage estimation by LASSO. As noted earlier, LASSO is employed to identify the best instrument set separately for each endogenous variable. Then, the nonlinear SUR procedure is employed to estimate the three equations – (8), (9), and (10) – with predicted values from LASSO replacing actual values for each of the endogenous variables.

Given the panel setting with three equations, correcting for contemporaneous correlation if any should improve the precision of the estimated parameters (Greene, 2012). In addition, the nonlinear SUR is flexible to set restrictions on coefficients. The coefficient of domestic inventory (a_4) should be greater than 1 to define the marketing cost and marginal convenience yield as the derivative of marketing cost with respect to inventories ($\psi_{ij,t} = -\partial C^m_{ij,t} / \partial n_{ij,t}$)[6]. The price volatility measure is normalized by using its mean, and similarly, other variables including price, output, convenience yield, domestic inventory and international inventory are also normalized to avoid any non-stationary problem[7]. The above procedure is repeated for the grains and corn subsamples. First differencing of data together with log transformation of right-hand-side variables as in equation (10) reduce the number of observations available for estimation: 1,000 (all 13 commodities), 899 (grains), and 208 (corn).

In the Table 2.2, the first three coefficients represented by c are the coefficients for the price equation (8), and the coefficients represented by a are the coefficients for the convenience yield equation (9). Focusing first on the results of price equations, the results show that domestic price increases along with output as expected, and this effect is statistically significant for all the cases. In other words, the marginal cost of production is increasing. Also, the coefficient on the trade openness dummy is negative and significant. This implies that allowing international supply by imposing a lower tariff rate compared to other countries, lowers domestic price.

For the convenience yield estimation, the results for all the 13 commodities show that most coefficients are statistically significant and have the expected signs of a well-behaved value

---

[6] Note the minus sign in front of $a4$ in equation (9).

[7] Even though panel unit root tests for the variables reject the nulls, it is hard to say that all the variables are stationary because the data are unbalanced and tests for many of panels could not be computed. In addition, the nonlinear SUR was not working with non-scaled variables. Some variables might have stationary issue (Pietola et al, 2010).

of marginal inventory (convenience yield) function. The exception being the coefficient on price (a_1) in the convenience yield equation, which is insignificant. The results from column 1 for the 13-commodity case show that convenience yield is increasing with its own price volatility as expected (Pindyck, 2004; Gorton et al., 2013). Interestingly, international price volatility also has a positive effect on a country's convenience yield. This result points to the potential that decisions on inventories and marketing of commodities consider volatility in both international and domestic markets. Surprisingly, the coefficient on international price volatility is larger than that on domestic volatility, which might arise if countries have become increasingly integrated in agricultural markets, e.g., China, Japan, and the Middle East now rely heavily on major agricultural producers such as USA, Brazil, and Australia.

Equation (9) has a negative sign before a_4, hence the positive value of a_4 refers to a negative relationship between inventories and convenience yield. The estimated coefficient of domestic inventory is 1.12, which also satisfies the restriction from equation (6), i.e., ($a_4 > 1$), and statistically significant. Although the coefficient of international inventory level is smaller than that on domestic inventory (-0.01), it is statistically significant. In general, when a country is open to the world trade, international volatility as well as international inventory have significant but opposite effects on convenience yield. To be specific, increases in the international price volatility cause additional increases in the convenience yield, while increases in international inventory cause additional decreases in the convenience yield in the countries with lower relative tariff rates, i.e., more open to trade.

Column 2 of Table 2.2 shows results for the grains subsample where the price coefficient is positive and significant as expected, but the negative coefficient on domestic inventory (a_4) does not satisfy the assumption ($a_4 > 1$) of the convenience yield equation. Therefore, estimation with a restriction on a_4 was attempted as in Pindyck (2004). Specifically, the same nonlinear model is re-estimated with a_4 simply constrained to equal 1.1. Column 3 in Table 2.2 shows the results of the restricted model for the grains subsample. With the restriction, domestic and international price volatility now have positive and significant effects on convenience yield, but the coefficient of international inventory is not statistically significant. The estimation for corn subsample shows that coefficients on price, international price volatility and domestic inventory are statistically significant and have the expected signs. Corn is produced and traded by many countries in the world (59 out of 71 countries in the sample are engaged in corn trade). In addition to the large volume of trade (average import volume in the sample is 1.4 million tons), tariffs are zero in several countries during 2010–2019. For these reasons, international information likely has a distinct influence on corn price analysis, and especially, international volatility appears more important than domestic volatility.

As mentioned earlier, Japan and Korea have much larger convenience yield values compared to other countries. Table 2.3 reports the estimation results of an attempt to analyze the nonlinear models except for outliers from Japan and Korea. Most effects seen in the Table 2.2 remain significant with the same sign in the Table 2.3. For all the three different groups, output has positive effects on prices and trade openness has negative effects. Focusing on the convenience yield estimation results, now an increase in domestic inventory leads decreases in convenience yield for all the three cases, and the coefficients are significant. Above all, since the estimated a_4 (domestic inventory coefficient) is greater than 1 even in the grains subgroup, there is no need to impose restrictions. With regard to the effects of international market information, international price volatility significantly affects marginal convenience yield for all three cases, while international inventory is insignificant for the grains subsample. Domestic inventory significantly affects marginal convenience yield for all the cases with theoretically consistent signs, but domestic price volatility again does not have a statistically significant effect in the corn subsample.

To sum up on Table 2.2 and 2.3 results, regardless of whether data from Japan and Korea are considered or excluded, the effects of international price volatility on marginal convenience yield are unambiguous in the estimated models: positive and statistically significant for all 13 commodities, the grains and corn subgroups. The estimated coefficients of international price volatility are bigger than those of domestic price volatility. In addition, the coefficient in the grains subgroup is greater than that in all 13 commodities, and it is the largest in the corn subgroup. Usually, grains, especially wheat, corn, and soybean are known to be produced and traded a lot worldwide. Even in countries with low relative tariffs on each agricultural commodity, the average trade volume of grains is much higher than the average trade volume of non-grain, and the average trade volume of corn is higher than that of grain. Hence, international price volatility can have a greater impact on domestic price dynamics.

After excluding Japan and Korea from the sample, the coefficients of the domestic inventory are significant and theoretically satisfy the assumption. However, international inventories did not significantly affect marginal convenience yield in the models for grains sub-sample. It is likely that volatility spills across markets easily, but inventories are not necessarily shared even when available, especially during time periods of high volatility. Nonetheless, international market information appears to play a critical role in domestic commodity price dynamics.

Conclusion

This study investigates the effect of international market information – especially price volatility and inventories – on domestic market commodity price dynamics. For this purpose, a structural model of price dynamics is extended to directly identify the relationships among price, price volatility, inventories, and convenience yield. The resulting empirical model includes equations for convenience yield and price formation as functions of inventories (domestic and international), volatility (domestic and international), trade openness, and other controls.

With price, convenience yield, output, and inventories as endogenous variables and 71 countries in the sample, the potential instrument set is large. Hence, LASSO, a regularization and variable-selection method, is used in the first stage to identify the best instrument set for each endogenous variable. Then, a nonlinear approach is applied to estimate the second stage of the model.

The estimation results, particularly those for the 13 commodities, show that both domestic and international price volatility positively and significantly affect the convenience yield, while domestic and international inventory levels have the opposite effects. Specifically, the result that convenience yield increases as domestic price volatility increases points to a well-behaved marginal value of storage function as in earlier studies. Similarly, this study has identified alike effects of international price volatility as well. Similar effects of domestic and international volatility are also observed for the major grains subsample, especially when imposing a theoretical restriction on the effect of inventory and when excluding the data from Japan and Korea. Results imply that international price volatility plays a critical role in domestic market price dynamics. With regard to international inventories, the model yields expected effects: convenience yield decreases with an increase in global inventories. However, while international price volatility has a significant effect on convenience yield for all the analyses (grains, corn, and non-Japan and non-Korea subgroups), the effect of international inventory is only found to be statistically significant in a few cases. As noted earlier, price volatility likely spills across markets easily, but inventories are not necessarily shared. Although exporters have increased their capability, a number of port rejections may arise hindering inventory sharing (Yoshida, 2017). In spite of these, cultural and political distances between trading partners are still the most profound challenges in trade (Knoll et al., 2018). Nonetheless, the importance of international market information on domestic commodity price dynamics, especially for making decisions on commodity inventory level and marketing has been demonstrated in this study. In particular, international market information is likely as important as domestic market information, if not more, in the case of importers.

The interconnection between domestic price dynamics and international markets information mandates proactive policy measures to secure price stability in domestic markets. For instance, for progress in agriculture predictions of future economic conditions are critical. Therefore, stakeholders seek information on changing patterns of trade patterns and the underlying economic factors. The results of this paper show the utility of international market information including WASDE (World Agricultural Supply and Demand Estimates) in securing stability in the context of domestic prices. Moreover, the results suggest the need for additional cross-country information on inventories. Given the increasingly globalized agricultural markets, the importance of international market information in determining domestic price movements is greater now than before. Future efforts with finer data and a larger sample (time and countries) will help in further understanding of the role of trade in commodity price dynamics.

Table 2.1. Summary Statistics

Variable		Mean	SD	Min	Max	1st Quartile	3rd Quartile	Observation
Price (p)	All	531.15	732.48	55.5	18,093.4	240.7	519.4	
	Grain	391.28	307.5	55.5	2,872.1	225.5	433.6	
	Corn	309.34	176.63	85.5	1,191.4	196.77	361.6	
Production (x)	All	10.5 (MMT)	36.4 (MMT)	0 (MMT)	389 (MMT)	92,000	4.93 (MMT)	
	Grain	11.7 (MMT)	38.4 (MMT)	0 (MMT)	389 (MMT)	0.14 (MMT)	6.28 (MMT)	
	Corn	21.6 (MMT)	67.2 (MMT)	0 (MMT)	389 (MMT)	0.39 (MMT)	7.73 (MMT)	
Domestic Inventory (n)	All	2.32 (MMT)	12.1 (MMT)	0	212 (MMT)	18,000	0.73 (MMT)	All: N=1,639 n=309 $\bar{T}$=5.30
	Grain	2.57 (MMT)	12.8 (MMT)	0	212 (MMT)	24,000	0.85 (MMT)	
	Corn	4.67 (MMT)	22.3 (MMT)	0	212 (MMT)	56,000	0.83 (MMT)	
Tariff (r)	All	0.18	0.69	0	21.88	0.00	0.15	
	Grain	0.19	0.72	0	21.88	0.00	0.14	Grain: N=1,459 n=270 $\bar{T}$=5.40
	Corn	0.14	0.32	0	4.25	0.00	0.14	
Domestic Price volatility (σ^2)	All	0.10	0.09	0	1.42	0.05	0.13	
	Grain	0.10	0.07	0	0.80	0.05	0.13	
	Corn	0.11	0.08	0	0.80	0.05	0.14	
Convenience Yield (ψ)	All	300.49	541.28	0.68	5,252.88	44.72	319.76	
	Grain	260.69	507.63	0.68	5,252.88	41.59	254.48	Corn: N=313 n=59 $\bar{T}$=5.31
	Corn	205.17	307.31	1.89	1,958.19	35.70	274.20	
International Price Volatility (σ^2_{-i})	All	0.10	0.03	0.03	0.26	0.08	0.11	
	Grain	0.10	0.02	0.04	0.15	0.08	0.11	
	Corn	0.11	0.01	0.09	0.14	0.10	0.12	
International Inventory ($n_{-\mathrm{i}}$)	All	3.12 (OMT)	21.2 (OMT)	1.85 (MMT)	392 (OMT)	146 (TMT)	4.24 (SMT)	
	Grain	3.46 (OMT)	22.4 (OMT)	1.85 (MMT)	392 (OMT)	159 (TMT)	6.62 (SMT)	
	Corn	8.65 (OMT)	39.8 (OMT)	105 (MMT)	392 (OMT)	421 (TMT)	9.41 (SMT)	

Notes: SD – Standard deviation; MMT – million metric tons; TMT – trillion (10^{12}) metric tons; SMT – Septillion (10^{24}) metric tons; OMT – octillion (10^{27}) metric tons

Table 2.2. Non-Linear SUR Estimates of Price and Convenience Yield Equations

Variable/ Parameter	13 commodities Column 1	Grains Column 2	Grains_ Restriction Column 3	Corn Column 4
Output (c_1)	3.44*** (1.01)	3.06*** (0.72)	3.22*** (0.81)	1.85* (0.97)
Tariff (c_2)	-5.23*** (1.62)	-2.14* (1.11)	-4.93*** (1.28)	-3.69** (1.47)
Year (c_3)	0.12 (0.30)	0.14 (0.28)	0.26 (0.28)	0.26 (0.40)
Price (a_1)	-0.07 (0.32)	5.18*** (0.47)	6.12*** (0.50)	4.06*** (0.35)
Domestic price volatility (a_2)	0.56*** (0.13)	0.01 (0.20)	1.17*** (0.14)	-0.01 (0.20)
International price volatility (a_3)	0.87*** (0.08)	0.49*** (0.16)	1.46*** (0.12)	1.84*** (0.25)
Domestic inventory (a_4)[a]	1.12*** (0.04)	-0.46** (0.20)	1.1	1.19*** (0.12)
International inventory (a_5)	-0.01* (0.01)	-0.01 (0.02)	0.01 (0.01)	-0.01 (0.03)
Observations	1,000	899	899	208

Notes: Standard errors in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.
[a] Equation (9) has a negative sign before a_4, hence the positive value of a_4 refers to negative relationship between inventories and convenience yield

Table 2.3. Non-Linear SUR Estimates (Japan and Korea Excluded)

Variable/ Parameter	13 commodities Column 1	Grains Column 2	Corn Column 3
Output (c_1)	3.10***	2.70***	1.92*
	(0.88)	(0.66)	(0.98)
Tariff (c_2)	-4.65***	-2.68***	-3.62**
	(1.44)	(1.01)	(1.53)
Year (c_3)	0.08	0.12	0.19
	(0.26)	(0.24)	(0.40)
Price (a_1)	-0.11	4.63***	3.94***
	(0.33)	(0.26)	(0.36)
Domestic price volatility (a_2)	0.54***	0.88***	0.02
	(0.14)	(0.10)	(0.18)
International price volatility (a_3)	0.88***	1.30***	1.83***
	(0.08)	(0.10)	(0.27)
Domestic inventory (a_4)[a]	1.12***	1.14***	1.07***
	(0.05)	(0.05)	(0.13)
International inventory (a_5)	-0.01*	0.01	-0.01*
	(0.01)	(0.01)	(0.03)
Observations	966	865	208

Notes: Standard errors in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

[a] Equation (9) has a negative sign before a_4, hence the positive value of a_4 refers to negative relationship between inventories and convenience yield

Table A1. List of Countries and Data Availability

Country	Commodity	Time
Albania	Wheat, barley, oats, maize	2010, 2011, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Argentina	Wheat, oats, maize, rice, sorghum, soybean, sunflower seed, cotton	2010, 2011, 2012, 2013
Armenia	Wheat, barley	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Azerbaijan	Cotton	2017, 2018, 2019
Bangladesh	Wheat, maize, rice, soybean, rapeseed	2010, 2011, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Belarus	Wheat, oats, maize	2019
Benin	Maize, rice, sorghum	2011, 2016, 2018
Bhutan	Wheat, barley, maize	2010, 2011
Bolivia	Coffee, wheat, barley, maize, rice, sorghum, soybean, sunflower seed	2016, 2017, 2018
Bosnia and Herzegovina	Wheat, barley, oats, maize, soybean, sunflower seed	2018, 2019
Brazil	Coffee, maize, rice, soybean cotton	2015, 2016
Brunei	Rice	2017, 2019
Burundi	Maize	2019
Cabo Verde	Maize	2012, 2014, 2015, 2016, 2017, 2018, 2019
Cambodia	Maize, rice	2012
Cameroon	Maize, rice, sorghum, oil palm	2012
Canada	Wheat, barley, oats, maize, soybean, rapeseed, sunflower seed	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Chad	Wheat, maize, rice, sorghum	2011, 2014, 2017, 2018
Chile	Wheat, maize, rice	2013, 2017, 2018, 2019
China	Wheat, barley, oats, maize, rice, sorghum, soybean, rapeseed, sunflower seed, oil palm, cotton	2010, 2012, 2013, 2014, 2015, 2016, 2019
Colombia	Coffee, wheat, barley, maize, rice, sorghum, soybean, oil palm, cotton	2010, 2011, 2012, 2013, 2015, 2016, 2019
Costa Rica	Maize, rice	2010, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Dominican Republic	Coffee, maize, rice, sorghum,	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2018
Ecuador	Coffee, wheat, barley, maize, rice, soybean, cotton	2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Egypt	Wheat, maize, rice	2017
El Salvador	Maize, rice, sorghum	2010, 2011, 2012, 2018

Ethiopia	Wheat, barley, maize, rice, sorghum, soybean, rapeseed	2010, 2015
European Union	Almond, wheat, barley, oats, maize, rice, sorghum, soybean, rapeseed, sunflower seed, oil palm, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Gambia	Maize, rice, sorghum	2013, 2019
Ghana	Maize, rice, sorghum, oil palm	2011, 2018, 2019
Indonesia	Coffee, maize, rice, soybean	2012, 2014, 2015, 2017, 2018, 2019
Israel	Almond, wheat, barley, maize	2013, 2014, 2015, 2016, 2017, 2018, 2019
Japan	Wheat, barley, rice, soybean	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Jordan	Wheat, barley, maize	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2018
Kazakhstan	Wheat, barley, oats, maize, rice, rapeseed, sunflower seed, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Kenya	Barley, maize, rice, sorghum	2010, 2016
Kuwait	maize	2016, 2017, 2018, 2019
Kyrgyzstan	Wheat, barley, oats, maize, rice, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2019
Lebanon	Barley	2012
Madagascar	Rice	2010, 2011
Malawi	Maize, rice	2013, 2014, 2015, 2016, 2017
Malaysia	Maize, rice, oil palm	2011, 2012, 2013, 2014, 2016, 2017, 2018
Mali	Maize, rice, sorghum	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Mongolia	Wheat	2010, 2011, 2013, 2016, 2018, 2019
Mozambique	Wheat, maize, rice	2017
Namibia	Wheat, maize	2018
Nicaragua	Maize, rice, sorghum, soybean	2012
Niger	Rice, sorghum	2010, 2011
Panama	Coffee, maize, rice	2010, 2011, 2013, 2014, 2015, 2016, 2017
Paraguay	Wheat, maize, rice, sorghum, soybean, rapeseed, cotton	2011, 2018
Peru	Coffee, wheat, barley, maize, rice, sorghum, soybean, cotton	2010, 2011, 2012, 2013, 2014, 2016, 2017, 2018, 2019

Philippines	Coffee, maize, rice, soybean	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Russia	Wheat, barley, oats, maize, rice, soybean, rapeseed, sunflower seed	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Rwanda	Coffee, wheat, maize, rice, sorghum	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Saudi Arabia	Wheat	2013, 2014, 2015
South Africa	Wheat, barley, oats, maize, sorghum, soybean, sunflower seed, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018
South Korea	Barley, maize, rice, sorghum, soybean	2010, 2017, 2018, 2019
Sri Lanka	Coffee, maize, rice	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Sudan	Wheat, sorghum	2013
Suriname	Rice	2012
Tajikistan	Wheat, maize, rice	2019
Thailand	Coffee, maize, rice, sorghum, soybean, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Togo	Coffee, maize, rice, sorghum, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Trinidad and Tobago	Rice	2012, 2013
Turkey	Almond, wheat, barley, oats, maize, rice, soybean, rapeseed, sunflower seed	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2018, 2019
Ukraine	Wheat, barley, maize, soybean, rapeseed, sunflower seed	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
United States	Wheat, barley, oats, maize, rice, sorghum, soybean, sunflower seed, cotton	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2019
Uruguay	Wheat, barley, oats, maize, rice, sorghum, soybean	2017
Vietnam	Coffee, maize, rice, soybean, cotton	2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019
Zambia	Maize, rice, sorghum	2014, 2015, 2018, 2019
Zimbabwe	Wheat, maize, sorghum, soybean, cotton	2013, 2014, 2016, 2017

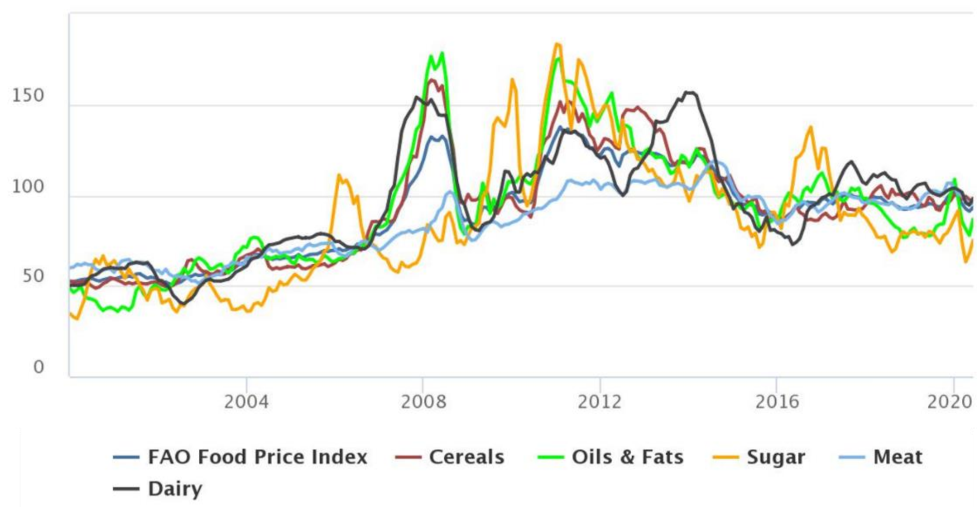


Figure 2.1. FAO Food Price Index

Source: Agricultural Market Information System (AMIS), FAO

CHAPTER 3

COMPARING MACHINE LEARNING AND MAXIMUM LIKELIHOOD METHODS TO ESTIMATE GRAVITY MODELS: THE CASE OF GLOBAL BEEF TRADE[8]

[8] Sei Jeong, Gopinath Munisamy, and Feras Batarseh. To be submitted to Agricultural Economics.

Abstract

The gravity model was introduced to approximate the size of bilateral trade flows between any two countries with their economic size and distance to each other as main features. The Poisson Pseudo Maximum Likelihood (PPML) was developed recently to estimate the gravity model to overcome the potential bias caused by the large proportion of zero observations in bilateral trade data. However, conventional statistical and econometric techniques have been challenged in modeling the gravity relationship, especially the need for a large number of pair-wise fixed effects. There is an opportunity for recent computational advances, such as Machine Learning (ML) together with big data technologies and high-performance computing to unravel, quantify, and understand trade patterns. This study makes a modest attempt to compare the advantages and disadvantages of PPML methods and ML models. The main advantage of the PPML is strong predictive power, but only with the inclusion of a large number of fixed effects. ML models have the advantage in selecting explanatory variables when a long list of explanatory variables is to be considered. Although the ML models have some limitations such as the lack of a theoretical explanation, they have sufficient potential for development that can be used in various analysis in the future.

Introduction

Trade among countries and its growth in the past two decades have been critical to economic well-being in developed and developing countries. In particular, expanding international trade in agricultural commodities, e.g., meat, soybean, has played a more critical role in serving the food and nutritional needs of the global populace. Given the increasingly globalized agricultural markets, changes in the production or consumption of agricultural commodities in one place can now drastically affect other economies even if not directly linked to trade (Schierhorn et al., 2016). Therefore, the relationship between trade flow and economic factors has been a research priority in recent years.

Tinbergen (1962) advanced one of the first empirical framework to relate trade to economic factors, i.e., the gravity model, where the magnitude of bilateral trade flows between any two countries is directly (inversely) proportional to their combined economic scale (distance between them). Anderson (1979) and Bergstrand (1985) provide theoretical foundations for the gravity model, augmenting the distance effect to include other trade costs such as tariffs and logistics. Eaton and Kortum (2002) further advanced the theory behind gravity models by using a Ricardian model of heterogeneous firms in each country, while Anderson and van Wincoop (2003) developed an endowment-economy model along the lines of Anderson (1979). Most recently, Allen et al. (2020) provided a universal framework for understanding the general equilibrium forces in gravity trade models.

Parallel to the advances in the theory of gravity models is innovative empirical approaches to quantify the trade relationship to economic fundamentals. A key challenge in empirics has been the fact that many countries do not trade with many others. Silva and Tenreyro (2006) proposed the Poisson Pseudo Maximum Likelihood (PPML) to deal with a potential bias arising from zero observations in bilateral trade data, the left-hand-side of the gravity equation. The PPML model is robust to different patterns of random error structure and provides a natural way to address sample selection issues arising from the omission of zero trade flows (Silva and Tenreyro, 2006; Yotov et al., 2016; Grant et al. 2018). Following that advance, gravity models have focused on the effect of distance (Melitz, 2007; Anderson, 2014), migration (Ghatak et al, 2009; Figueiredo et al, 2020), currency unions (De Sousa, 2012; Glick, 2017; Larch et al, 2019), trade facilitation and related policies (Kepaptsoglou et al, 2010; Felipe and Kumar, 2012; Kuik et al, 2019; Lopez, 2019).

Despite the empirical popularity of the gravity model, the use of conventional statistical and econometric techniques has been recently challenged (Martin, 2020, Varian, 2014). For instance, Figueiredo et al. (2020) indicated that with product level data (Harmonized System, HS, 6-digits) conventional techniques, including the PPML estimator, face convergence issues due to the large number of estimated coefficients. Martin (2020), using Monte Carlo simulations, demonstrated the biases inherent in PPML and related techniques. Most applications of PPML are limited to three or four-digit HS codes, partly due to the fact that disaggregated data create convergence and collinearity challenges. Imposing pair-wise fixed effects, often in three dimensions (exporter, importer and time) likely exacerbates the collinearity problem.

With advances in data availability, e.g. monthly trade data has become available more recently, there is an opportunity now for recent computational advances, such as Machine Learning (ML) together with big data technologies and high-performance computing to unravel, quantify, and understand economic relationships (Mullainathan, 2014; Liakos et al, 2018; Gopinath et al., 2021). One of the main advantages of the ML techniques, often categorized into three main groups; supervised learning, unsupervised learning, and reinforcement learning, is that they are capable of autonomously solving large non-linear problems using large datasets from multiple sources (James et al., 2014; Chlingaryan et al, 2018; Liakos et al., 2018; Sharma et al., 2020).

Estimating the bilateral trade of beef among all countries at the HS-6 level using conventional and ML methods – to explain and compare factors driving trade patterns – is the primary objective of this study. Meat, especially beef, has been a key contributor to the dramatic growth in agricultural trade in the last two decades aiding in increased nutritional content of diets in many developing countries. In addition to comparing conventional and ML methods, this study evaluates their predictive ability in alternative policy scenarios. Specifically, the PPML estimation with high-dimensional fixed effect (HDFE) is applied to allow for year and country-type (exporter or importer) effects and their interactions. In the case of ML, supervised learning as in Gopinath et al. (2021), which has classification and regression as its main subcategories, is selected to investigate the relationship between trade flows and economic factors.

Gravity Model and Data

Poisson Pseudo Maximum Likelihood (PPML)

The basic gravity model is given by:

$$y_{ijt} = \exp\left(\beta * \boldsymbol{X}_{ijt} + \delta_{it} + \delta_{jt} + \delta_{ij}\right), \tag{1}$$

where y_{ijt} refers to the value of imports in US dollars of country i from j at time t and $\boldsymbol{X}_{ijt}$ is a set of explanatory variables varying by i, j and t as well. The δ_{it}, δ_{jt}, and δ_{ij} refer to importer-year, exporter-year, and importer-exporter fixed effects, respectively. In the gravity literature, the use of time-varying exporter and importer fixed effects account for multilateral resistance - the other barriers to trade between a pair of countries – implied by the theory (Anderson and van Wincoop, 2003; Feenstra, 2004: Baldwin and Taglioni, 2007). Time-invariant pair-wise fixed effects have been also included to absorb unobservable frictions to trade and to address the endogeneity of trade policy variables (Baier and Bergstrand, 2007). The PPML estimation with high-dimensional fixed effect (HDFE) is applied to allow multiple fixed effects and interactions, and the standard errors are clustered on the pair of exporter and importer countries.

Explanatory Variables and Data

The dependent variable, monthly bilateral import data of beef disaggregated at HS 6-digit level from January 1990 to December 2020 was obtained from Global Trade Atlas (GTA)[9]. There are six beef commodities at the HS6 level including fresh or chilled bovine meat – 020110 (Bovine carcasses and half carcasses), 020120 (Bovine cuts bone in), 020130 (Bovine cuts boneless)– and frozen bovine meat – 020210 (Bovine carcasses and half carcasses), 020220 (Bovine cuts bone in), 020230 (Bovine cuts boneless). The dynamic gravity data collected by Gurevich and Herman (2018) is the main source for the gravity variables. The data is available from 1990 to 2019, but current gross GDP obtained from the World Development Indicators replaced the GDP data from the source due to its absence after 2015. The GDP is in nominal rather than real terms to proxy economic sizes (Baldwin and Taglioni, 2006; Shepherd, 2013; Gurevich and Herman, 2018; Kuik et al., 2019). Most of the gravity data are time invariant variables, and the only annual GDP for as many as countries is available, so the monthly gravity data were derived using the annual data repeatedly. Overall, monthly data from January 1990 to December 2019 is the time dimension of the data used for the analysis.

Tariff rate was the first considered explanatory variables in the current context of trade wars with tit-for-tat tariffs. The tariff data was obtained from TRAINS during the time period of 1989 through 2014, and recent tariff data after 2014 was obtained from Market Access Map

[9] International bilateral trade flow data can be accessed with subscription or purchasing the data from the GTA.

(MAM). Missing data of the tariff rate was replaced by using several ways including using the data from the observable preceding period and the importer countries' average tariff rate toward the rest of the world. For the data of each EU country, each country's EU entry year was considered to solve the missing data issue.

Explanatory variables widely used in the previous literature were then considered to determine the international trade flow of beef commodities. For instance, GDP, distance, common language and contiguity are introduced in many gravity studies (Anderson, 2014; Kuik et al., 2019; Conconi et al., 2020; Figueriredo et al., 2020; Lopez ,2020). Furthermore, trade polices such as customs unions (CU) and free trade agreements (FTA) were also chosen to verify the effects of trade facilitation variables.

As noted earlier, the PPML HDFE model using disaggregated data encountered multicollinearity problems requiring some transformations of the independent variable. For example, the summation of the GDP of the two trading partners replaced the GDP of each exporter and importer countries in the estimation model. The north-south distance between trading partners, defined as $NS_{ij} = abs(Latitude_i - Latitude_j)$ in Anderson's (2014) study, replaced the latitude of each country. The north-south distance potentially takes into account of different life cycles of beef production between Southern and Northern Hemispheres. Also, inclusion of the difference in longitude between importer and exporter countries, population, and importer-exporter fixed effects produced results suggesting severe multicollinearity (lack of significance of most variables).

Machine Learning Techniques

As its name implies, ML methodologies involve a learning process with the objective to learn from experience to perform a task. For example, the trained model can be used to classify, predict, or cluster new examples using the experience after the learning process. An advanced statistical analysis is applied to the same data through the ML to identify patterns and model relationship, hence the performance on a specific task or forecasts are progressively improved (James et al. 2014; Liakos et al, 2018; Mohri et al., 2018).

Usually, ML methods are categorized into three main types; supervised learning, unsupervised learning, and reinforcement learning (James et al. 2014; Liakos et al., 2018; Sharma et al., 2020). In supervised learning, data are presented with example inputs and the corresponding outputs with an objective to construct a relationship between inputs and outputs. A predictive model is developed using the example data with the prior knowledge of the input and the corresponding output variables during training. Then, future values are predicted applying the same knowledge to new input data. In other words, the relationship between trade and independent variables is derived from "training" data sample then applied to a "test" data sample to compute prediction through the supervised ML models.

Classification and regression are the main two subcategories of the supervised learning. Unlike the unsupervised learning, which does not have pre-defined response variable with unlabeled datasets, in the supervised learning, the dataset is labeled (Jordan and Mitchell, 2015; Traore et al., 2017; Liakos et al., 2018; Gopinath et al., 2021). Boosting algorithms such as Light GBM, XGBOOST, and Extra Trees further help in refining the supervised learning process.

Decision trees are classification or regression models formulated in a tree-like architecture. They are segmented into branches (splits) and leaves (nodes). Each leaf node contains a simple regression model and represents the final decision or prediction taken after following the path from root to leaf (Belson, 1959; Rodriguez-Galiano et al., 2015; Ahmad et al., 2018). In this procedure, each training sample of the decision tree is dependent on previous trees, and they are chosen sequentially. The boosting techniques - Light GBM, XGBOOST, and Extra Trees - collects several weak prediction models and improves the classification. The key difference between LightGBM and XGBoost is that the XGBoost splits the tree nodes one level at a time (level-wise), while LightGBM splits leaf-wise. The leaf-wise algorithm can reduce information loss relative to the level-wise algorithm and hence can yield better accuracy (Khandelwal, 2017; Ke et al., 2017; Gopinath et al., 2021). Extremely randomized trees (extra trees) algorithm is a type of ensemble learning technique, developed as an extension of random forest algorithm (Geurts et al., 2006). It is similar with random forest in that it uses a random subset of features to train each base estimator. The difference between extra trees and random forest is that the best feature along with the corresponding value for splitting the node is randomly selected in extra tree. The entire training dataset is used to train each regression tree in extra trees while the random forest uses a bootstrap replica to train the model (John et al., 2015; Ahmad et al., 2018).

Estimation Results

PPML Estimation

The empirical PPML estimation model using the selected 8 explanatory variables is following:

$$\begin{aligned} y_{ijt} = \exp\Bigg(&\alpha + \beta_1 \ln\left(GDP_{ijt}\right) + \beta_2 Contiguity_{ijt} + \beta_3 \ln\left(Distance_{ijt}\right) \\ &+ \beta_4 \ln\left(1 + \frac{Tariff_{ijt}}{100}\right) + \beta_5 Common\ language_{ijt} + \beta_6 CU_{ijt} \\ &+ \beta_7 FTA_{ijt} + \beta_8 NS_{ijt} + \delta_{it} + \delta_{jt} + \delta_{tt}\Bigg) \end{aligned} \quad (2)$$

Three fixed effects such as importer-year (δ_{it}), exporter-year (δ_{jt}), and year-month (δ_{tt}) are included. Even though the PPML is advantageous in naturally accounting for zero trade pairs in the data, adding 1 to the trade values minimizes the missing observations, then logarithms of both sides of the equation were taken (Figueriredo et al, 2020).

The PPML with HDFE estimation results of the gravity model are shown in Table 3.1. Overall, the PPML estimation of beef trade models show that R^2 scores are higher than 0.8. Results in table, the traditional gravity model, are comparable both in significance and signs to previous literature in most cases. GDP, common language, CU, and FTA have significant and positive coefficients for most of the six beef commodities. The positive effects of GDP have already been foreseen in the fact that the meat market has grown significantly in many countries with economic growth, as GDP provides economic snapshots of the country. The positive coefficients of FTA and CU agreements suggest the important role of trade facilitation in the case of the beef market (Moïsé and Sorescu, 2013). As Melitz (2007) and Anderson (2014) find, there is a possibility that north-south distance can have significant and positive impact on trade flows. For example, the coefficients of the north-south distance (NS) for commodity 020130 (Fresh or chilled bovine cuts boneless) and 020210 (Frozen carcasses and half-carcasses) are positive and significant. The positive coefficients suggest that the difference in life cycle between northern and southern hemisphere due to the opposition of seasons significantly encourages international trade.

Distance and tariff rate have significant and negative relationship with trade, as expected. The importer i's tariff rate applied on imports from j has negative effects consistent with the previous literature (Fontagné et al., 2018; Figueiredo et al., 2020). Also, the impact of contiguity, i.e., sharing a common land border, on trade is positive and significant at 10% for some commodities – 020110 (Fresh or chilled bovine carcasses and half carcasses) and 020220 (Frozen bovine cuts bone in) (Gurevich and Herman, 2018; Figueiredo et al., 2020; Lopez, 2020). The sparse evidence for contiguity across the six HS codes can be attributed to observed data that most trading partners with the highest cumulative trade values (over the sample period) do not share a border.

Supervised ML Results

Table 3.2 presents the performances of the three supervised ML models measured by R^2, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), and the sizes of training and test data sample for six commodities also can be seen in the table. ML models with the highest R^2 also have lower MAE or RMAE scores usually. Based on the R^2 scores, the LightGBM has the best performance for three commodities (020130, 020210, and 020230), while extra tree regression does for two commodities (020110, 020120, and 020220).

Table 3.3 presents feature importance scores from the explanatory variables in the best fitted models. The values of information gain were normalized based on the highest score values, meaning that 100 indicates the variable with the highest feature importance score, and other values are relative information gain values compared to the highest value. The ranking of each variable is in parentheses. GDP and Tariff rate are the top information provider for three commodities, respectively. The relative importance shows that distance related explanatory – Distance and North-South Distance – also have great influence on learning of trade flows. On the other hand, the trade policy facilitation such as customs union and FTA provide relatively lower feature importance scores.

In both PPML and ML analysis, GDP and tariff rate significantly affect the international beef trade flows. The main differences between the PPML estimation and the supervised ML models are that the estimated coefficients of FTA and CU in the former were significant for most of the commodities and were much higher than that of north-south distance, while ML models suggest that north-south distance is more important in learning the trade flows of beef commodities.

Comparison: PPML vs. ML

In this section, three criteria are introduced to evaluate the traditional PPML and new ML techniques to identify the pros and cons of the two approaches. Easy-to-see first is statistical criteria or performance scores. As seen in Table 3.1 and 3.3, four criteria including R^2, MAE, RMSE, RMAE, and RRMSE were calculated. It is seen that PPML has higher R^2 scores for all the six commodities when using HS6 level data. However, when comparing other performance scores such as MAE and RMSE, ML analysis shows better performance scores than the PPML analysis. In addition, it is observed that the estimation results of the PPML have a lot of missing observations. For example, for 020110, the ML analysis uses 3,833,628 test samples, while the observation number of PPML analysis is only 406,728. Many of observations are dropped because they belong to FE groups for which the dependent variable (value of imports) is always zero. For example, as explained above, an importer-exporter fixed effects were added into the estimation. If only zero values are observed for the entire sample for a particular trade pair, then it is not possible to include these observations as the fixed effects that corresponds to them is not defined.

Observing predictions or in-sample forecasts is the next step to evaluate the two methods in analyzing the international beef trade flows. Figure 3.1 and 3.2 show the aggregated trade values of each commodity from all over the countries measuring by PPML and ML models, respectively. Predicted import values of the PPML estimation tracks the actual import values very closely, while there are bigger differences between the predictions of the supervised ML and the actual values. The gravity equation has been found to have very high predictive power

(Yotov, 2020). It delivers very strong fit and plausible estimates on the standard set of gravity variables (Martin, 2020). However, it has been noticed that the powerful predictive power of the gravity model mostly comes from fixed effects in this analysis. Three fixed effects such as importer-year (δ_{it}), exporter-year (δ_{jt}), and year-month (δ_{tt}) are included as explained above. The predictions without these fixed effects are shown in figure 3. Charbonneau (2017) finds appropriately controlling for multiple fixed effects has a substantial effect on the estimated parameters of interest relative to models without fixed effects or models inappropriately controlling for fixed effects. For instance, the estimation of nonlinear panel models with fixed effects encounters a challenge in the form of a incidental parameters problem (IPP) derived from the estimation of the unobserved time invariant heterogeneity or unrestricted individual specific effects (Neyman and Scott, 1948; Bai, 2009; Charbonneau, 2017; Fernandez-Val and Weidner, 2018; Weidner and Zylkin, 2021). In the gravity model, importer and exporter fixed effects absorb market sizes and multilateral resistances, and unobserved trade costs enter the country-specific terms. The country-specific parameters are noisily estimated, raising concerns about a possible IPP. Estimates can suffer from bias due to the IPPs, affecting the validity of inference and rendering the estimates inconsistent. In addition, when common shocks have homogenous effects on the output, the model collapses to the usual time effect. It is the heterogeneity that gives rise to a factor structure (Bai, 2009). Therefore, interactive effects or multiple fixed effects play a major role in reducing bias.

There are two major benefits associated with using pair fixed effects in gravity estimations. First, the pair fixed effects are able to account for the endogeneity of trade policy variables (Baier and Bergstrand, 2007). Second, the pair fixed effects provide a flexible and comprehensive account of the effects of all time-invariant bilateral trade costs. Pair fixed effects have been shown to carry systematic information about trade costs in addition to the information captured by the standard gravity variables (Egger and Nigai, 2015; Agnosteva et al., 2014). However, the downside of using pair fixed effects is that one cannot identify the effects of any time-invariant bilateral determinants of trade flows, because the latter will be absorbed by the pair fixed effects. In addition, the pair fixed effects lead to the exclusion of many observations, in some cases 90 percent of the sample.

Figure 3.4 and Figure 3.5 represent in-sample forecasts for the top1 and 2 trade pairs. Specifically, the top 1 and top 2 trade partners for the beef commodities are following:

1) 020110 (fresh or chilled bovine carcasses and half carcasses): Netherlands – Italy, France – Italy
2) 020120 (fresh or chilled bovine cuts bone in): France – Italy, Germany – Italy
3) 020130 (fresh or chilled bovine cuts boneless): Australia – Japan, Canada – United States
4) 020210 (frozen bovine carcasses and half carcasses): Ukraine – Russia, Belarus - Russia
5) 020220 (frozen bovine cuts bone in): United States - South Korea, Australia – Korea
6) 020230 (frozen bovine cuts boneless): Australia – United States, New Zealand – United States

It is also seen that the predictions from the PPML estimation with fixed effects are very close to the actual values of imports. Surprisingly, for the commodity 020210, the accuracy of prediction from the PPML estimation without fixed effects seems to be better than those from the machine learning analysis. This indicates that PPML without fixed effects may not cause serious bias for individual estimation when the observation number is not huge. On the other hand, as seen in the

Figure 3.3, the estimation can cause serious bias problems as the sample size is bigger. Similar results can be seen in Figure 3.5, which shows the predictions of top 2 trade pairs.

Finally, it was attempted to compare out-of-sample predictions using the GDP forecasts from 2020 to 2026 obtained from the IMF, World Economic Outlook Database. Due to the lack of monthly data, the annual data is repeated on a monthly for the same year. Assuming the explanatory variables except GDP are time invariant, the data in 2019 was repeated to build another data sample set from 2020 to 2026. Then estimating the predictions was carried out under three simulations to investigate which policy change would have the greatest impact on increasing trade flows, hence to find a desirable direction for implementing policy. Three simulations were adapted as following:

- Case 1: 10% decrease in tariff
- Case 2: 10% decrease in distance
- Case 3: 10% increase in GDP forecasts

The simulation results are mostly in line with the results of PPML estimation and feature importance of ML models. Figure 3.6 shows trade forecasts under the three simulation cases by applying PPML. It shows that a 10% decrease in distance (case 2) has the great impacts on trade flows compared to other variables. A 10% reduction in tariffs (case 1) has the smallest effects on the trade flows among the three cases. As reported in table 1, the elasticity of distance was higher than that of GDP for all the six commodities from the PPML estimation results. Therefore, a decrease in distance dominates other variables and plays an important role in increasing trade flow. The forecasted patterns of all six commodities are very similar as the coefficients of the variables were comparable among the six commodities from the PPML estimation results.

Overall, the simulation results imply that reducing the distance between trade partners is the key factor to increase the trade flows.

Increasing in transportation costs is a well-known explanation why distance matters in the international trade context. The further the distance, the higher the costs of transporting the product, and therefore the cost of imports and exports. The building of Suez Canal is a good example to show the effort to shorten trade distance between trade partners. The most obvious benefit using the canal was the saving in distance, and the new channel led to significant shifts in the patterns of Eastern and Australasian trade. The canal gave a great impetus to the building of large, fast, economical steamships, so its opening and closure have provided great influence on the bilateral trade between partners (Fletcher,1958; Feyrer, 2009). If considering shifts to trading costs, increasing the load also can be an alternative to reduce distance.

Under the supervised ML models, the training data is from 1990 to 2010 as before, and the new test data is from 2020 to 2026. Extra tree regression was selected for 020110 and 020120, and 02220, and LightGBM was used for 020130, 020210, and 020230. The out of sample predictions are shown in Figure 3.7. The effect of increasing GDP (Case 3) has the great impact on increase in trade flows for most of the commodities. Interestingly, tariff rate is highest information provider for frozen commodities such as 020210, 020220, and 020230, but a 10% decrease in tariff rate (Case 1) leads the largest increase in the trade forecasts only for 020210. The effects of 10% decrease in tariff rate is the smallest for most of the commodities.

Overall, the out-of-sample predictions with three-case simulation show that a 10% decrease in distance (Case 2) has the great impacts on trade flows compared to other variables when using PPML estimation. The ML analysis predicts a 10% increase in GDP (Case 3) leads the biggest increases in beef trade flows for most commodities.

Conclusion

Agricultural trade encourages investment and promotes economic growth. For instance, agricultural exports generate 7,500 full-time, civilian jobs for every $1 billion in farm exports in the United States (USDA-ERS, 2018). Predicting trade flows in agricultural markets is critical in context of economic growth in both developed and developing economies. This study investigates the bilateral trade of beef using PPML methods and ML analysis and compares their applicability and prediction quality. The PPML methods have been widely used to estimate trade flows. Machine Learning (ML), especially, three supervised ML models –LightGBM, XGBoost, and Extra tree regression- were implemented to compensate for the limitations of the PPML methods such as multicollinearity problem especially when using disaggregated trade data. It is clearly seen that the PPML estimation has higher R^2 score than the ML models, which implies the PPML estimation fits better than the estimations of other ML models. In addition, its in-sample prediction tracks the actual values very closely. As the PPML method has been used in many studies over a long period of time, the predictability of the method is quite powerful. The estimation results are also consistent with other previous literature results and theories. For example, GDP and trade facilitation such as CU and FTA have positive and significant effects on the trade flows, while distance and tariff rate have negative effects on it. Elasticities and predictions can be easily calculated through these numerated estimated values.

Meanwhile, several Machine Learning (ML) methods were employed to overcome some of the limitations of PPML. One of the main advantages of the ML methods is that it can review large volume of data and identifies specific trends and patterns easily. As explained in appendix, 32 gravity variables were selected to analyze the beef trade flows in this study. Given lots of explanatory variables, it is difficult to select which variables can be considered as important among them. It is very important to choose meaningful and appropriate variables to establish an equation. As mentioned earlier, the PPML method has experienced severe multicollinearity issue. In addition, the estimation results vary greatly depending on which variables are included into the equation. The ML analysis especially, comparing the feature importance of the variables not only greatly helped select the explanatory variables when estimating using the conventional econometrics, but also allows to examine the effects of each of the variables without collinearity issue. In addition, ML models have better performance scores when comparing MAE, RMSE, RMAE, and RRMSE. A lot of observations are dropped in the PPML estimation. Especially, inclusion of pair fixed effects exacerbated the drop in observations. The powerful predictive power of the gravity model mostly comes from fixed effects rather than the explanatory variables themselves.

Based on the analysis of the two different methods, simulations under three different cases were also carried out. The simulation results were in line with the results from the analysis part of this study, showing that a 10% decrease in distance (case 2) has the greatest influence on increase in trade flow in PPML and that a 10% increase in GDP (case 3) leads to the largest increase in trade flows in ML models.

The ML models show another advantage in doing out-of-sample prediction. Even though the PPML method with multiple fixed effects increase bias correction and prediction power, it does not do out-of-sample prediction. In this study, the fixed effects estimates were copied to the out-of-sample observations. If the fixed effects do not go through big changes, copying the estimates would be an optimal option to fix the issue. However, when the fixed effects change over time, then the out-of-sample predictions can be biased because PPML estimation relies heavily on the fixed effects to obtain predictive power. In addition, as mentioned earlier, if some observations belong to FE groups for which the dependent is always zero, then they are dropped. Therefore, the PPML analysis cannot be possible if the observations are insufficient. For example, the same analysis for each of the top 5 exporters was tried using both PPML and ML methods, but the PPML analysis was impossible due to the insufficient observations. Meanwhile, the predictions of ML models track the actual import values closely (see Appendix Figure 3).

As the gravity model using PPML methods has been used in many studies for a long time, the estimation results of the model show very high accuracy in predict the bilateral trade flows. The ML techniques have recently come into the spotlight with their capability of autonomously solving large non-linear problems using datasets from multiple sources. Even though the methods utilize a lot of data, it is insufficient to explain the results theoretically. However, the ML models present the potential for finding accurate and robust methodologies to specify complex economic relationships in this study. It also would be useful to predict future trade patterns more accurately. Since the gravity model was first introduced in 1962 by Tinbergen, it has gone through a lot of development and been used as the most powerful method to analyze trade flows. On the other hand, it has not been long since machine learning techniques were used for the trade analysis. Thus, ML techniques have sufficient room for improvement in their adaptation to economic analysis.

Table 3.1. PPML Estimation of the Gravity Model (Monthly Data)

	020110	**020120**	**020130**	**020210**	**020220**	**020230**
$\ln(GDP)$	1.05+	0.54+	0.76**	1.30***	0.80**	0.44
	(0.60)	(0.30)	(0.25)	(0.34)	(0.30)	(0.34)
$\ln(Distance)$	-0.30	-1.47***	-1.67***	-1.58***	-1.45***	-0.83**
	(0.33)	(0.20)	(0.17)	(0.28)	(0.30)	(0.32)
$\ln\left(1+\frac{Tariff_{ij}}{100}\right)$	-1.99*	0.44	-2.10**	-1.84+	-1.42+	-1.91***
	(0.83)	(1.27)	(0.77)	(1.03)	(0.86)	(0.40)
$Contiguity$	0.98**	0.15	0.26	-0.31	0.51+	-0.15
	(0.31)	(0.21)	(0.33)	(0.31)	(0.30)	(0.35)
NS	-0.05	-0.02	0.02**	0.03+	-0.01	0.01
	(0.04)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
$Common\ language$	0.86*	0.73**	0.46	2.55***	1.78***	0.08
	(0.39)	(0.23)	(0.37)	(0.28)	(0.30)	(0.29)
CU	7.91***	3.87***	1.78**	7.48***	4.35***	2.01***
	(1.22)	(0.33)	(0.60)	(0.68)	(0.59)	(0.56)
FTA	5.12***	1.94***	1.08***	0.55	0.82*	0.86*
	(0.76)	(0.33)	(0.26)	(0.49)	(0.36)	(0.42)
$Constant$	-24.86	4.84	50.18***	-13.17	2.40	9.66
	(19.20)	(8.83)	(7.27)	(8.87)	(9.12)	(10.91)
N	406,728	781,896	1,130,701	330,540	957,948	1,563,096
Pseudo R2	0.90	0.93	0.89	0.91	0.90	0.82
MAE	48,265	39,802	113,842	4,851	13,113	147,196
RMSE	379,549	340,724	943,118	57,801	186,775	1,666,063
RMAE	63.31	49.56	60.93	49.90	64.59	99.22
RRMSE	497.88	424.25	504.73	594.46	919.94	1,123.04

Table 3.2. Performance Scores of the Supervised ML Analysis (Monthly Data)

		020110	**020120**	**020130**	**020210**	**020220**	**020230**
LightGBM	R^2	0.24	0.39	0.57	0.00	0.17	0.30
	MAE	3,934.34	7,362.96	34,985.57	595.31	4,086.11	53,954.60
	RMSE	117,175.6	209,051	625,456.4	25,668.5	208,497.2	1,084,477.3
	RMAE	140.10	119.32	132.71	232.35	128.78	153.52
	RRMSE	4,172.52	3,387.73	2,372.47	10,018.25	6,571.01	3,085.76
XGBoost	R^2	-0.06	0.25	0.40	-2.47	0.07	0.28
	MAE	6,225.42	8,940.63	41,147.90	1,123.96	5,244.65	54,747.75
	RMSE	138,716	232,365.8	739,931.8	47,896.1	219,816.8	1,103,007.8
	RMAE	221.68	144.89	156.08	438.67	165.29	155.78
	RRMSE	4,939.54	3,765.55	2,806.69	18,693.6	6,927.76	3,138.49
Extra trees Regression	R^2	0.53	0.48	0.48	-1.43	0.17	0.28
	MAE	2,681.52	5,638.30	30,137.77	564.53	3,875.79	44,583.66
	RMSE	92,305	192,787	691,237	40,082	208,414	1,098,985
	RMAE	95.49	91.37	114.32	220.33	122.15	126.86
	RRMSE	3,286.90	3,124.17	2,621.98	15,643.87	6,568.42	3,127.04
Observations	Train	6,316,788	6,316,788	6,756,528	6,316,788	6,316,788	6,316,788
	Test	3,833,628	3,833,628	4,111,069	3,833,556	3,833,508	3,833,508

Table 3.3. Relative Importance of Variables and Ranking (Monthly Data)

	020110	020120	020130	020210	020220	020230
GDP	100 (1)	100 (1)	100 (1)	87.87 (2)	80.26 (2)	40.67 (4)
Distance	98.75 (2)	54.79 (2)	67.84 (4)	71.59 (3)	44.00 (4)	58.81 (3)
Tariff rate	80.02 (4)	51.83 (4)	69.13 (2)	100 (1)	100 (1)	100 (1)
Contiguity	7.20 (7)	10.39 (6)	8.81 (5)	0.89 (8)	0.22 (7)	0.13 (8)
Common Language	9.80 (5)	19.30 (5)	8.29 (6)	2.34 (6)	3.03 (5)	1.94 (6)
Customs Union	8.23 (6)	9.11 (7)	1.90 (8)	1.06 (7)	0.09 (8)	0.13 (7)
FTA	6.11 (8)	7.09 (8)	4.54 (7)	5.31 (5)	0.64 (6)	2.60 (5)
North-South distance	91.12 (3)	53.16 (3)	68.14 (3)	64.27 (4)	50.51 (3)	59.55 (2)

Figure 3.2. PPML Estimation Predictions

Figure 3.2. Supervised ML Prediction

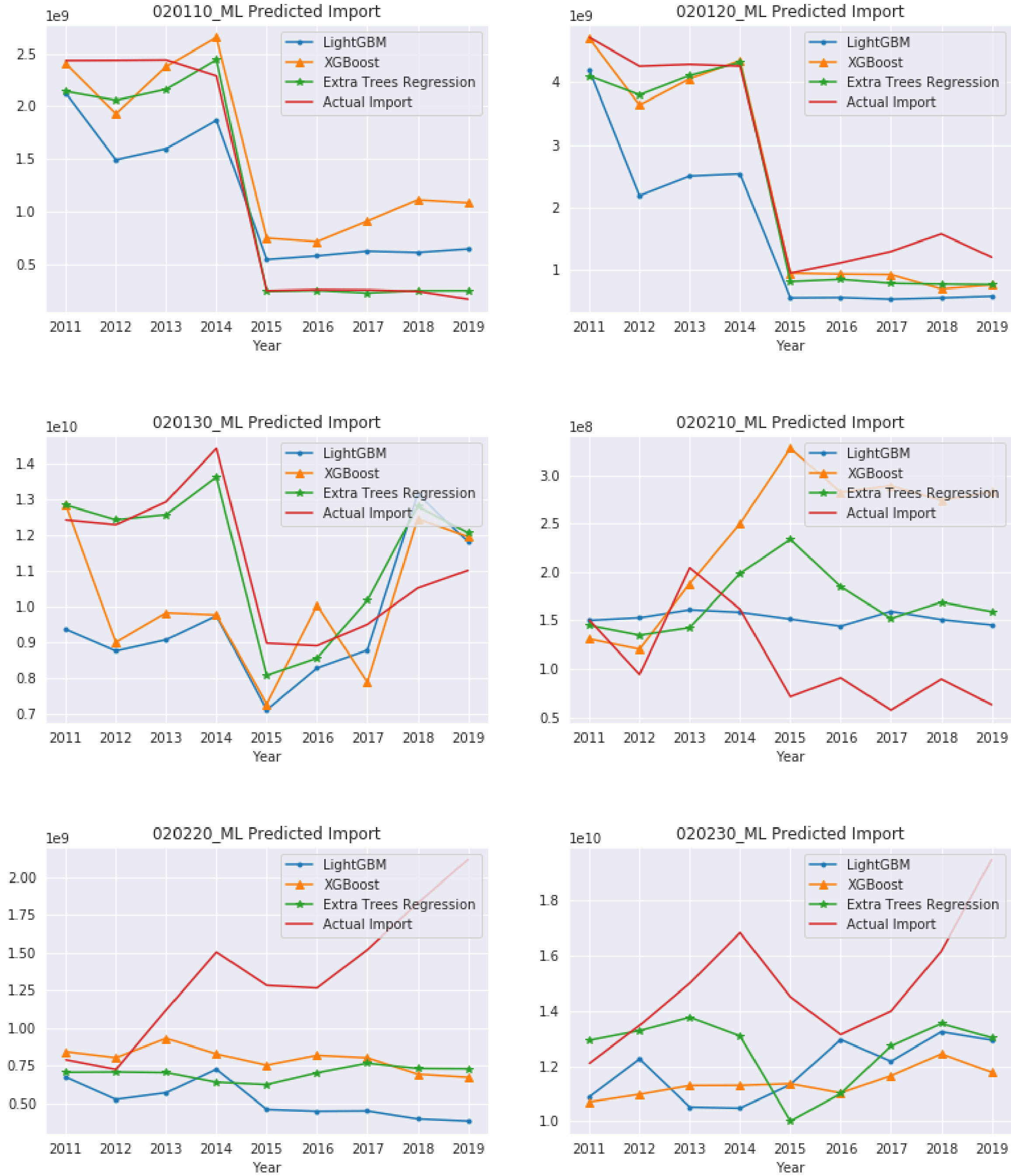


Figure 3.3. Predictions of ML and PPML without Fixed Effects

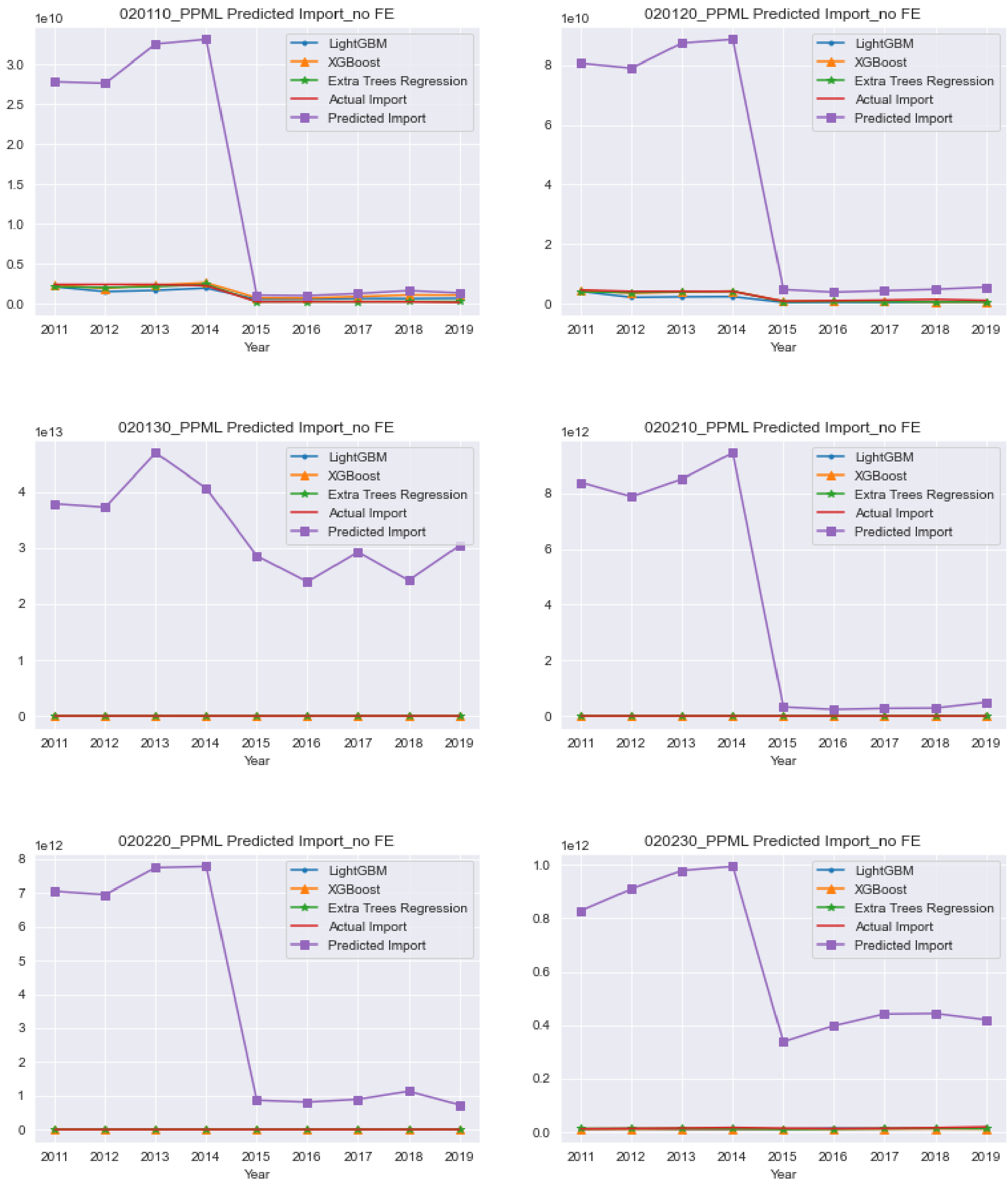


Figure 3.4. Predictions of Top 1 Trade Pair

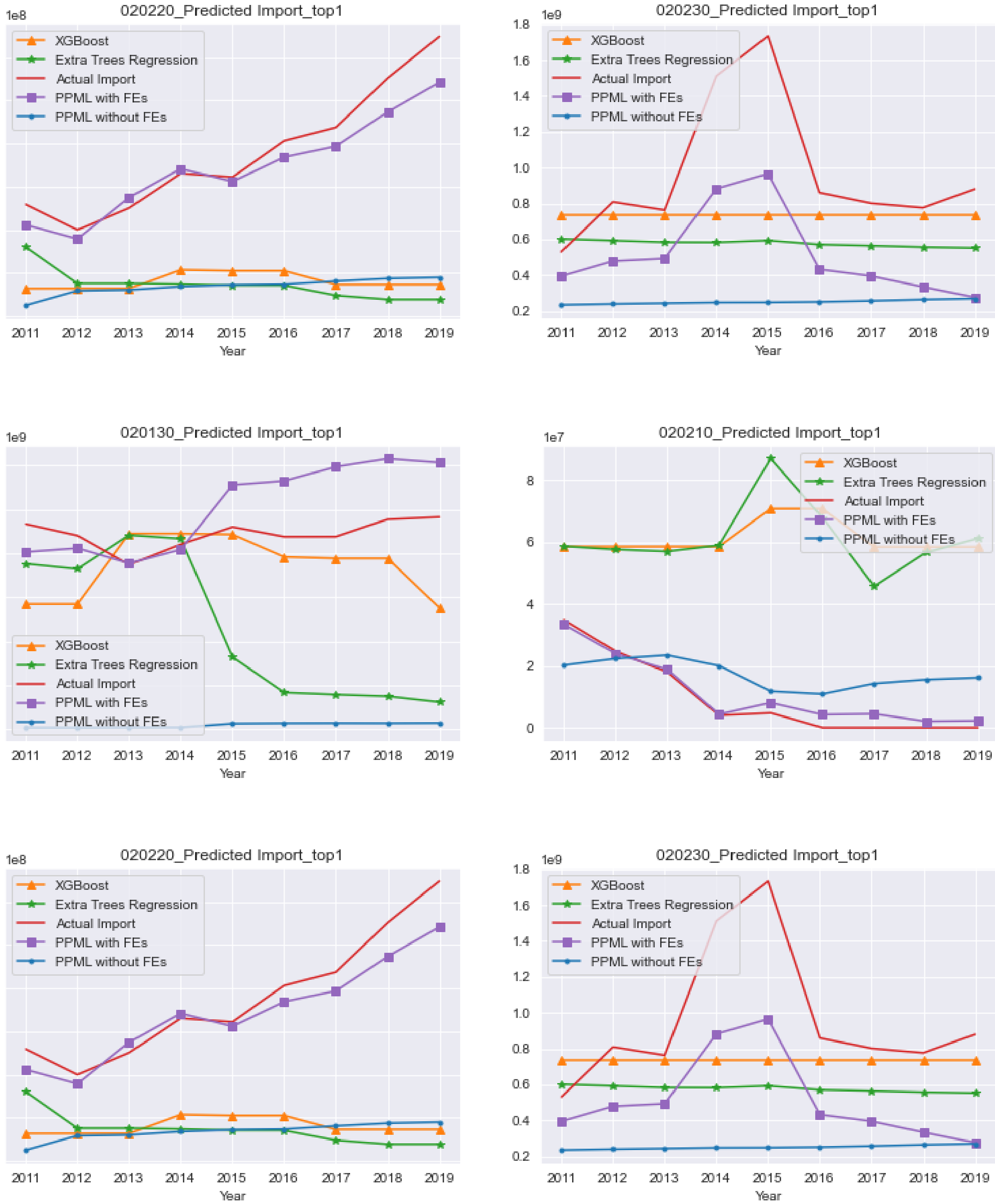


Figure 3.5. Supervised ML Prediction of Top 2 Trade Pair

Figure 3.6. Simulation of the PPML Estimation

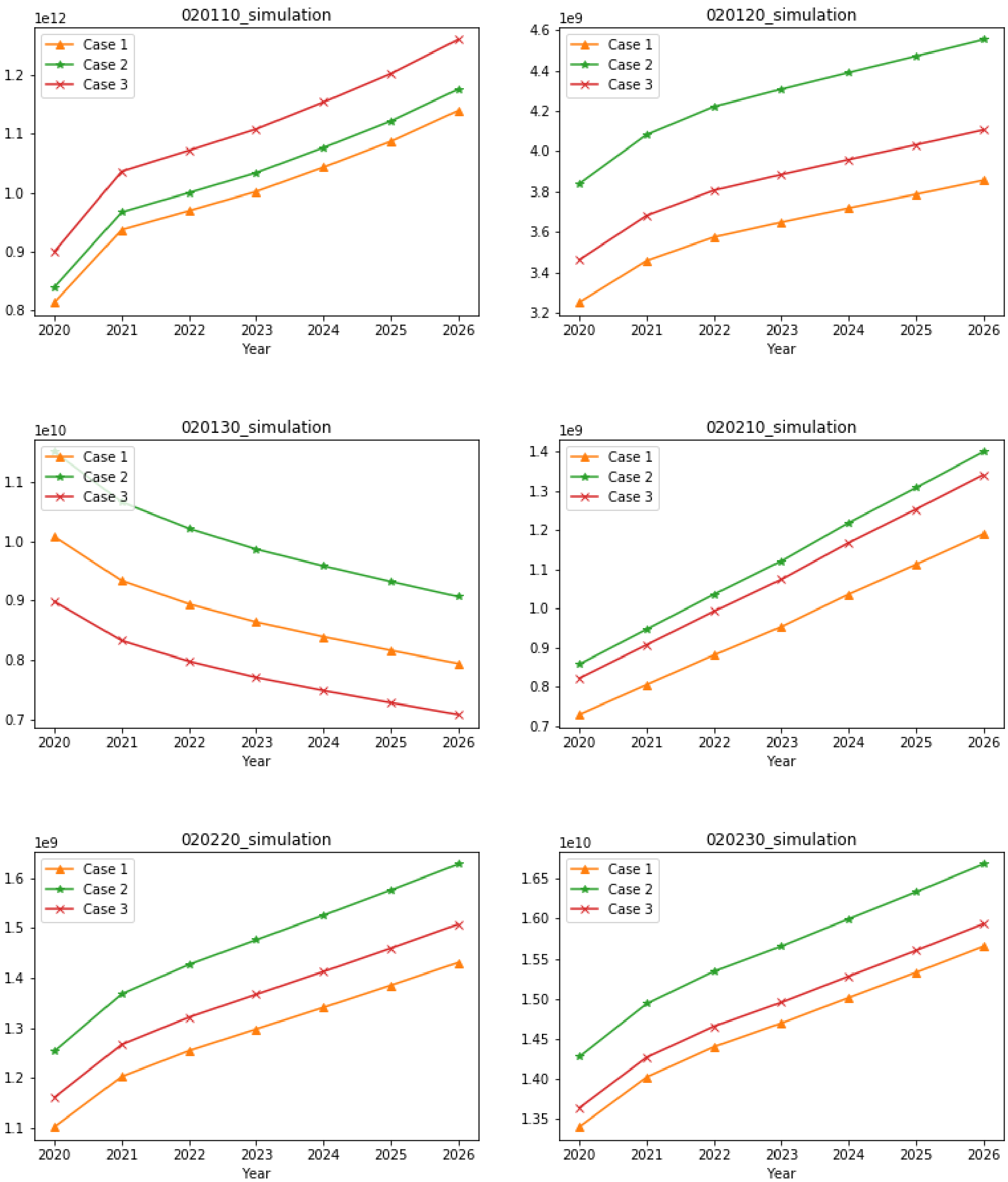
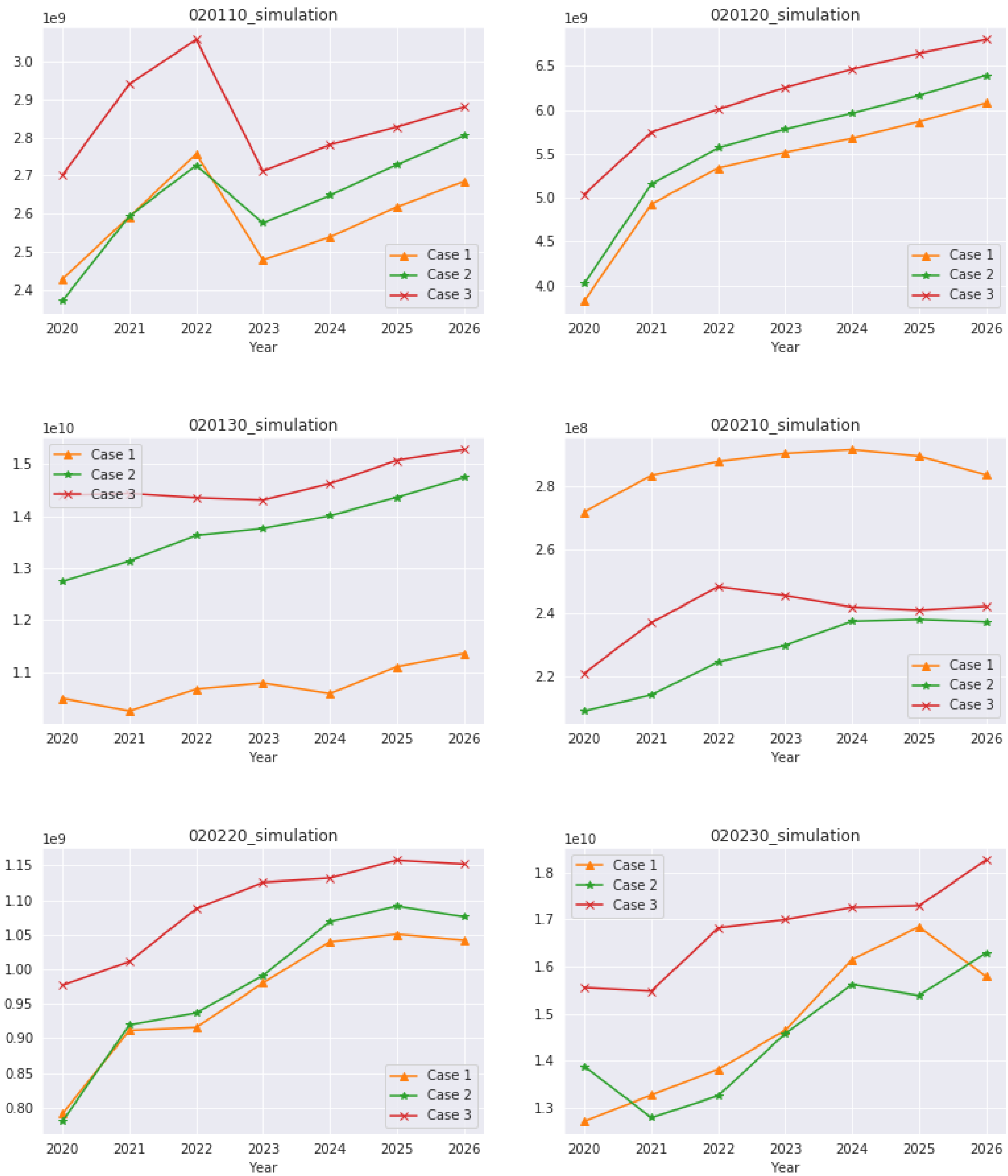


Figure 3.7. Simulation of the ML Models



Appendix B

Gurevich and Herman (2018) provided 70 gravity variables, but not all the variables cover the time period till 2019. In addition, including all the variables into the analysis model might cause several problems including multi collinearity. Gopinath et al. (2021) add 35 predictors in the ML models and provide the feature importance of the variables. Based on the paper, 32 gravity variables were firstly selected to define PPML estimation model after excluding some variables which have missing values after 2015. Appendix Table 1 presents the top 10 information provider among the 32 features.

The following shows the list of the 32 variables: contiguity, distance, tariff, common language, CU, FTA, GDP of exporter, GDP of importer, population of exporter, population of importer, latitude of exporter, latitude of importer, longitude of exporter, longitude of importer, stock of exporter, stock of importer, landlocked dummy of exporter, landlocked dummy of exporter, island dummy of exporter, island dummy of importer, PTA_goods agreement, PTA_services agreement, EIA, PSA, EU dummy of exporter, EU dummy of importer, WTO dummy of exporter, WTO dummy of importer, GATT dummy of exporter, GATT dummy of importer, colonial dummy of exporter, colonial dummy of importer

Appendix table 2 and 3 show the PPML and ML estimation results using annual data, and the predicted values are presented in appendix figure 1 and 2. The estimated results are similar to the analysis using monthly data. The PPML estimation results show better performance compared to the ML models, showing the predicted value of PPML track the actual import values very closely.

An alternative trade data set, which is combined based on countries, was built to see how the estimation results change with aggregated data. The import data of beef is the summation of all the six commodities for two trading partners, and the aggregated tariff rate are derived by averaging all the six tariff rates. Then top 5 beef exporting countries were selected for this analysis. Australia topped the list of beef exports by country, followed by the United States, Brazil, Netherlands, and New Zealand. The same explanatory variables were included in the model, and the cut-off year for the training data was also 2010 as the same as before. The PPML analysis for the top 5 exporter was impossible for insufficient observation numbers, hence only the supervised ML models were conducted.

Appendix B Table 1. Ranking of Feature Importance Scores

Rank	020110	020120	020130	020210	020220	020230
1	GDP_Destination	Population_Origin	Longitude_Destination	GDP_Origin	Stock_Destination	Latitude_Origin
2	Population_Origin	GDP_Origin	Distance	Distance	Longitud_Destination	GDP_Origin
3	GDP_Origin	Population_Destination	Latitude_Origin	Latitude_Destination	GDP_Destination	GDP_Destination
4	Population_Destination	GDP_Destination	GDP_Origin	Common Language	Distance	Population_Origin
5	Longitude_Destination	Common Language	Longitude_Origin	GDP_Destination	Population_Destination	Stock_Origin
6	Distance	Latitude_Destination	Population_Origin	Population_Destination	Latitude_Destination	Population_Destination
7	Longitude_Origin	Distance	Stock_Origin	Island_Origin	GDP_Origin	Distance
8	Latitude_Origin	Latitude_Origin	Tariff rate	Tariff rate	Tariff rate	Longitude_Origin
9	Latitude_Destination	Longitude_Destination	Population_Destination	Landlocked_Origin	Population_Origin	Stock_Destination
10	Common Language	Longitude_Origin	GDP_Destination	Longitude_Origin	Stock_Origin	Island_Origin

Appendix B Table 2. PPML Estimation of the Gravity Model (Annual Data)

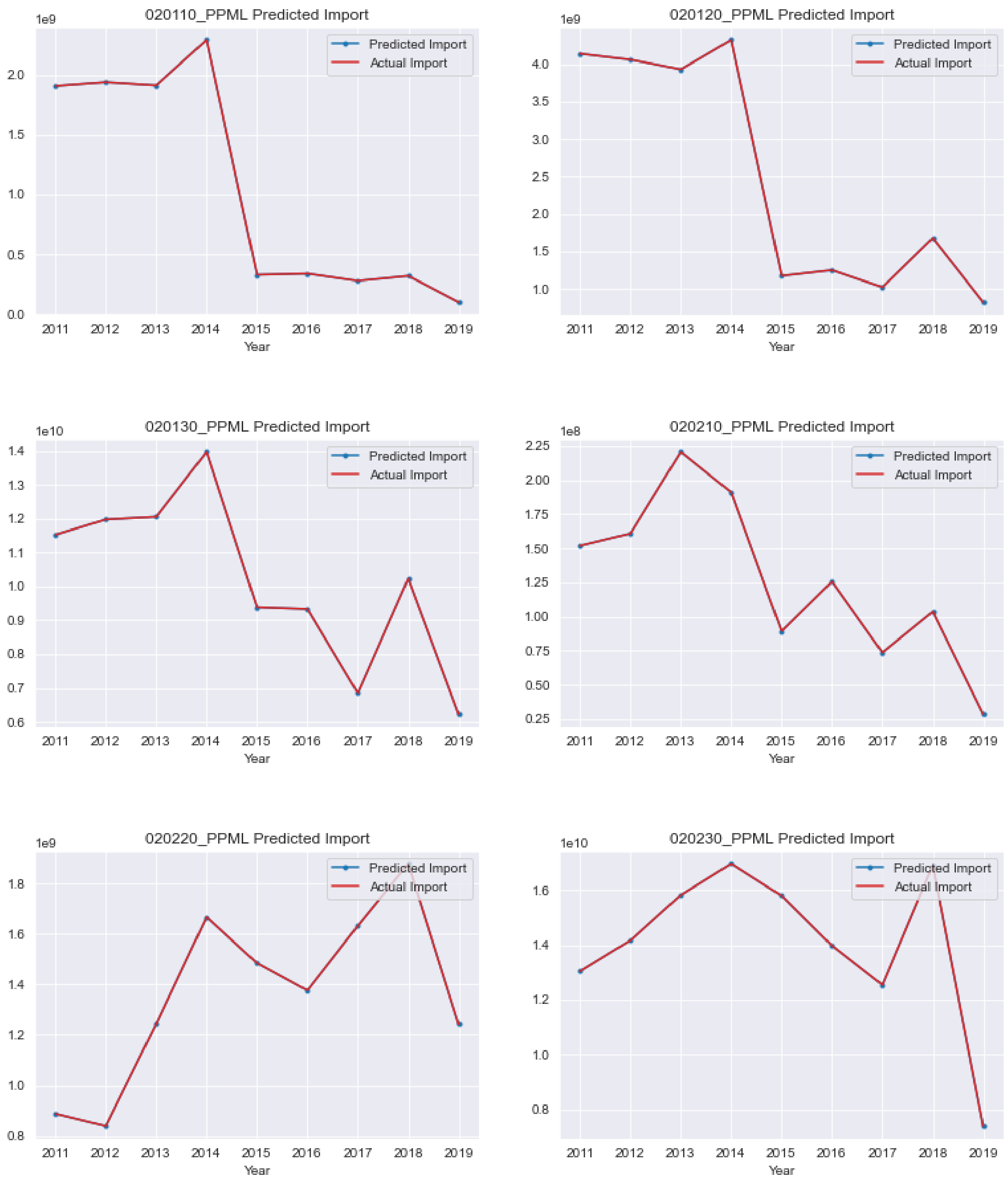
	020110	020120	020130	020210	020220	020230
$\ln(GDP)$	0.86+	0.31	0.59**	0.60*	0.28	0.18
	(0.49)	(0.28)	(0.19)	(0.27)	(0.34)	(0.30)
$\ln(Distance)$	-0.67*	-1.67***	-1.86***	-1.77***	-1.99***	-0.88**
	(0.28)	(0.19)	(0.20)	(0.20)	(0.27)	(0.29)
$\ln\left(1+\frac{Tariff_{ij}}{100}\right)$	-1.98**	0.05	-2.37**	-1.14	-1.82*	-1.71***
	(0.75)	(1.08)	(0.74)	(0.91)	(0.85)	(0.37)
$Contiguity$	0.96***	0.19	0.03	0.28	0.12	-0.24
	(0.29)	(0.20)	(0.36)	(0.32)	(0.29)	(0.33)
NS	-0.04	-0.01	0.02**	-0.01	0.00	0.01+
	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
$Common\ language$	0.66+	0.59**	0.60	2.38***	1.55***	0.25
	(0.36)	(0.21)	(0.42)	(0.29)	(0.32)	(0.26)
CU	5.29***	3.59***	1.66**	4.55***	3.05***	1.76***
	(0.62)	(0.26)	(0.56)	(0.51)	(0.40)	(0.48)
FTA	4.22***	2.03***	1.15***	0.76*	0.42	0.72+
	(0.48)	(0.26)	(0.28)	(0.33)	(0.29)	(0.39)
$Constant$	-10.97	15.67*	49.34***	10.13	25.00*	19.98*
	(15.07)	(7.93)	(5.80)	(7.55)	(9.93)	(9.58)
N	94,533	158,479	227,133	74,927	191,041	313,443
Pseudo R2	0.93	0.94	0.91	0.94	0.90	0.84
MAE	22,654.13	40,222.30	149,822.13	1,763.95	17,267.47	295,630.78
RMSE	781,581.25	1,172,606.6	3,391,660.3	55,808.13	566,773.31	7,190,758.5
RMAE	56.17	45.81	57.99	35.33	63.19	101.40
RRMSE	1,938.06	1,335.41	1,312.88	1,117.68	2,073.99	2,466.43

Standard errors in parentheses
="+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001"

Appendix B Table 3. Performance Scores of the Supervised ML Analysis (Annual Data)

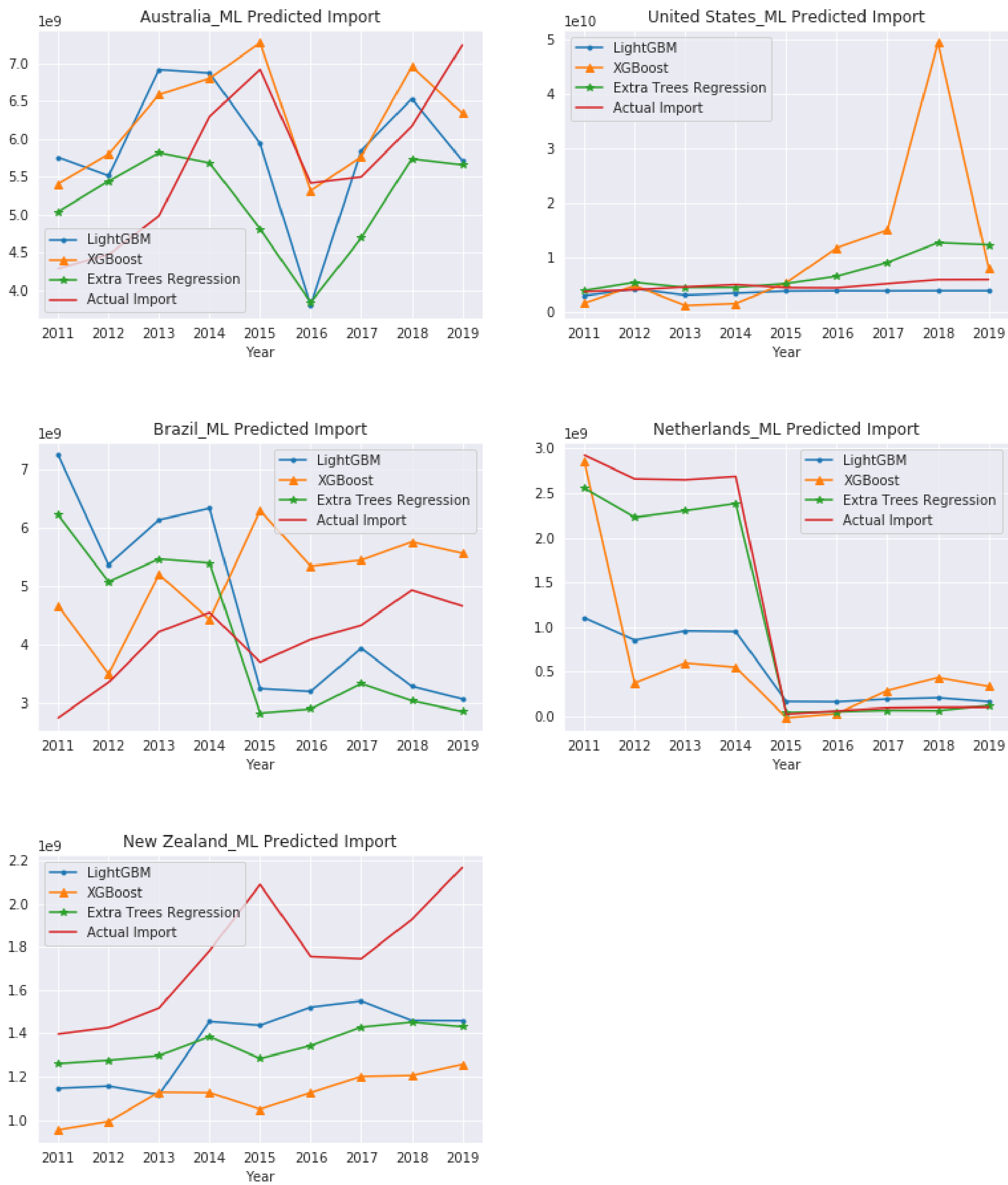
		020110	020120	020130	020210	020220	020230
LightGBM	R^2	0.29	0.39	0.57	0.02	0.18	0.37
	MAE	59,914.38	127,652.54	528,280.10	9,212.32	59,165.13	817,102.51
	RMSE	1,227,414	2,409,321	7,052,643	281,206	2,290,767	10,632,006
	RMAE	203.190	182.03	184.32	257.00	154.31	206.13
	RRMSE	4,162.58	3,435.70	2,460.67	7,845.06	5,974.59	2,682.09
XGBoost	R^2	0.25	0.38	-0.07	-2.28	0.42	0.36
	MAE	49,715.15	111,215.18	561,198.02	10,894.4	55,503.17	578,077.79
	RMSE	1,256,254	2,424,350	11,114,105	514,463	1,928,626	10,731,414
	RMAE	168.60	158.59	195.80	303.93	144.76	145.83
	RRMSE	4,260.39	3,457.13	3,877.71	14,352.44	5,030.09	2,707.17
Extra trees Regression	R^2	0.39	0.47	0.47	-1.79	0.18	0.34
	MAE	35,861.25	75,256.41	379,289.13	8,664.15	53,222.42	561,519.99
	RMSE	1,140,274	2,252,390	7,826,246	474,695	2,295,252	10,914,073
	RMAE	121.62	107.32	132.33	241.71	138.81	141.65
	RRMSE	3,867.06	3,211.91	2,730.58	13,242.97	5,986.29	2,753.25
Observations	Train	526,399	526,399	526,399	526,399	526,399	526,399
	Test	319,469	319,469	319,464	319,463	319,459	319,459

Appendix B Figure 3. PPML Estimation Predictions (Annual Data)



Appendix B Figure 2. Supervised ML Prediction (Annual Data)

Appendix B Figure 3. Supervised ML Prediction of Top 5 Exporters (Monthly Data)



CHAPTER 4

EFFECTS OF DISEASE AND IMPORTING SOURCES ON GLOBAL BEEF DEMAND[10]

[10] Sei Jeong, Chen Zhen, and Gopinath Munisamy. To be submitted to Journal of Agricultural and Resource Economics.

Abstract

The global meat consumption, especially the demand for beef, has increased significantly with rapid economic growth in many countries in recent decades. This study analyzes global beef import demand at the Harmonized System six-digit (HS-6) level through a multi-stage systems approach. The first stage estimates the demand function for total imported beef using the Almost Ideal Demand System (AIDS) approach while including controls for cattle diseases such as foot and mouth disease (FMD) and Bovine spongiform encephalopathy (BSE) as well as free trade agreements. The second stage estimates individual demand functions (by HS-6 commodity and importing country) for separate importing sources, i.e., the top five exporters. Results indicate that domestic outbreaks of FMD increase the total import demand for beef, whereas BSE outbreaks have the opposite effect. The cross-price elasticities from the second stage demonstrate that beef imported from a country is not necessarily a substitute for beef imported from other countries in certain import markets.

Introduction

Predicting future economic conditions is critical to progress in agriculture, as competitive markets and exchange of goods and services encourage investment and promote economic growth. In predicting the future, knowledge of future demand is critical to public and private decision making. Especially, information on demand elasticities are crucial inputs into many ex-ante analyses of consumption, welfare, and various trade policies. Such demand information is also required to assess the impact of technological change on food security in developing countries. A better understanding of export and import demand is critical to decision-making given the growing interdependence among countries and efforts to maximize the benefits of international trade (Kee et al., 2008; Dey et al., 2011; Yin and Hamori, 2011).

Within the meat market, which has grown in tandem with population and income growth, the global beef demand has grown significantly in the last three decades. For instance, rapid economic growth in the Asia Pacific has led to increased beef demand, which has been beneficial to U.S. producers (USDA, Agricultural Outlook 1996). However, animal disease outbreaks, such as foot and mouth disease (FMD) and bovine spongiform encephalopathy (BSE), have affected the international beef trade (Lloyd et al., 2006; Wieck and Holland, 2010; Knight-Jones and Rushton, 2013; Kompas et al., 2015; Webb et al, 2018). In one specific case, in 2003 the state of Washington experienced a BSE outbreak, which resulted in a decrease of US beef exports in the first quarter of 2004, coupled with an increase in Brazil's beef exports that year. Brazil has become a major exporter of beef, especially during 2001-2005 when its exports increased to over 1 million metric tons (Coffey et al., 2005; Steiger, 2006). Yang et al. (2013) also show that FMD outbreaks negatively impact international pork trade. In an exporting country, a FMD break will likely reduce exports, while that in an importing country will increase demand for imports, but these effects depend on the presence of a vaccination or a slaughter policy. Furthermore, the trade impact also depends on policy alignment between the importing country and the principles specified by the World Organization for Animal Health (OIE). For FMD, the OIE Terrestrial Animal Health Code requires the geographical separation of production zones from areas where FMD is present. Permitted policy responses include trade restrictions that either ban imports from areas where no separation has been established, or allow imports only from particular zones or compartments of the exporting country that is recognized as applying acceptable biosecurity standards (Shanafelt and Perrings, 2017).

This study derives the import demand elasticities of a broad group of countries at a disaggregated level (Harmonized System, HS, 6-digits) of beef commodities. Given that meat plays an important role in the diets of many countries, numerous studies have estimated meat demand. However, most studies estimate import demand elasticities at the aggregate level, when combined with trade policy evaluations at specific tariff lines, resulting in a mismatch aggregation bias (Kee et al., 2008). With HS 6-digit data, our analysis is expected to reveal the differences in import demands across country groups at the individual product line. Consider, for instance, frozen bovine cuts (boneless), i.e., HS 020230, where lower U.S. exports in November 2019 appear to be offset by an increase in Australian exports of the same cut of beef (UN Comtrade database, accessed 13 September, 2021). This implies that beef imported from the U.S. and Australia may be substitutes on the global market. However, data from Japan and South Korea, the major beef importers of the U.S. and Australia, do not show a clear substitution relationship between the two countries' beef. For the same product line, HS 020230, the import values of Korea from the U.S. are the highest in January and the lowest in December, 2019?. Likewise, those from Australia are also the highest in January and the lowest in December. The import values of Japan from the two countries are the highest in April. These seasonal patterns suggest that imported beefs from U.S. and Australia may not be a direct substitute for each other in some countries at the disaggregated level.

Methodology and Data

This section provides details on the multi-stage budgeting approach yielding the Almost Ideal Demand System (AIDS) for econometric analysis (Dey et al., 2011; Guta, 2012; Ramírez, 2013; Sha et al., 2015; Liu et al., 2016).

The multi-stage budgeting approach used for this research begins with an estimate for the demand function for imported beef on an aggregate level, i.e. world imports level. From there, the next stage involves estimating individual demand functions for separate importing sources, i.e., top five sources in the context of this study. Given the multitude of consumption goods purchased by households, an empirical estimation of demand for all goods creates a dimensionality problem. It is for this reason that empirical estimation of demand models requires numerous equations (Blundell et al, 1993; Fan et al, 1995). The multi-stage budgeting framework extends the idea of an exhaustive expenditure system to different levels of stages. These stages separate goods belonging to broad food categories, and then allocating expenditures over individual commodities (Fan et al, 1995). This approach allows for reducing the dimensionality of the estimated system, and thus creating a more reliable study (Dey et al, 2011).

The data are described alongside the model, but unlike previous studies monthly international import data obtained from Global Trade Atlas (GAT) are used in estimation. Moreover, as detailed below, data on Consumer Price Index (CPI) of foods and beverages, GDP Index, trade barriers, including those related to disease incidence such as foot and mouth disease

(FMD) and bovine spongiform encephalopathy (BSE), and free trade agreements are compiled for estimating the AIDS system.

First Stage: Import Demand for Beef Commodities

We specify the first-stage import demand for each of six beef commodities to be a function of its own price, price of all other foods, animal disease outbreaks, and other demand shifters as follows :

$$\ln Q_{it} = a_0 + a_1 \ln P_{it} + a_2 lnP_{f\&b,it} + a_3 \ln Y_{it} + \theta_1 FTA_{it} + \theta_2 BSE_{it} + \theta_3 FMD_{it} + \left(\sum_{\mathrm{r}} b_r m_{r,it}\right) + FE_i + \varepsilon_{it}. \quad (1)$$

Here, Q_{it} denotes the quantity of one of six beef commodities imported by country i from all sources in month t. We obtain import data from Global Trade Atlas (GTA). The six beef commodities include three HS codes for fresh or chilled bovine meat: 020110 (bovine carcasses and half carcasses), 020120 (bovine cuts bone in), and 020130 (bovine cuts boneless); and three HS codes for frozen bovine meat: 020210 (bovine carcasses and half carcasses), 020220 (bovine cuts bone in), and 020230 (bovine cuts boneless). We estimate Eq. (1) separately for the six HS codes. Eq. (1) is a reduced-form specification of demand in that it is not derived directly from utility maximization or its dual. Reduced-form first-stage demand is common in the import demand literature (e.g., Sha et al., 2015) owing to its tractability.

P_{it} is the unit value of imports plus tariff. We obtain bilateral tariff rate data by country and commodity from the TRAINS (up to 2014) and the Market Access Map (after 2014). Missing tariff data are replaced by data from the preceding period following the approach of Jayasinghe et al. (2010). In a few cases where there are no reported preceding-period data, the world average tariff rate replace the missing values. $P_{f\&b,it}$ is the consumer price index of food and beverage and Y_{it} is the GDP index, both obtained from OECD Stat.

The trade policy literature emphasizes the importance of a variety of trade costs besides import tariffs (Hoekman and Nicita, 2011). To capture the effect of trade facilitation in general, FTA_{it} is specified to be an indicator for any free trade agreement of country i with the worldwide top five importing sources i.e. top five exporting countries.[11] We rely on the NSF-Kellogg Institute Data Base on Economic Integration Agreements[12] to create FTA_{it} up to 2017 and country reports[13] for post-2017 FTA_{it} values.

To measure trade distortion caused by foot and mouth disease (FMD) and bovine spongiform encephalopathy (BSE), the respective diseases' dummies are included in Eq. (1). As mentioned earlier, the two cattle diseases have had great impact on trade volume between trading partners and consumption of imported beef in some countries. Kawashima and Puspito Sari (2010) demonstrate that country of origin bias grows in importance while the elasticity of substitution becomes less important after a BSE outbreak in an exporting country. OIE has historical data of animal diseases from 1996. For FMD, monthly data are accessible from 1996 to 2004, then bi-annual data after 2004. We track national reports on the OIE website and the World Reference Laboratory to populate the monthly FMD_{it} variable. However, for BSE it is challenging to track outbreaks at monthly frequency for European countries because BSE was widespread in Europe for a long period of time. When we cannot determine the month of outbreak for a country through internet search and the country's government websites, we

[11] Initially, we created five different FTA dummies, representing each country i's free trade agreements with each of the top five exporting countries. However, the top beef exporting countries are most often European Union (EU) members, leading to multicollinearity among the FTA dummies.
[12] https://kellogg.nd.edu/nsf-kellogg-institute-data-base-economic-integration-agreements
[13] For example, those of the Office of the United States Trade Representative, and Korean Ministry of Trade, Industry and Energy.

construct this country's BSE_{it} based on the bi-annual OIE data.[14] Finally, in equation (1), monthly dummy variables (m_r) capture seasonality in beef import, and FE_i represents country fixed effects.

The unit value P_{it} may be endogenous because of the conventional demand-supply simultaneity, unit value bias (Deaton 1988), or measurement error (Kee et al., 2008). To address these problems, we drop unit value outliers identified based on the inter-quartile range method. Before excluding the outliers, the calculated unit price ($/kg) had high variation, ranging from zero to $10,000, but this range was significantly reduced after the exclusion of the outliers (See table 1 for the maximum value of import price). In addition, we instrument P_{it} by its lagged value (z_1), inverse-distance weighted averaged of the unit value (z_2), and trade-weighted distance (z_3). Kee et al., (2008) propose the latter two instrumental variables to address endogeneity in import demand. z_2 for country i is constructed as

$$z_{2,it} = \sum_{k \neq i} \left(\frac{1}{distance_{ki}} \right) P_{it} \tag{2}$$

where $distance_{ki}$ is the distance between country k and i, obtained from Gurevich and Herman (2018); and z_3 is built as

$$z_{3,it} = \ln \left(\sum_{k} w_{kt} distance_{ki} \right), \tag{3}$$

where w_{kt} is the share of country k in world exports of beef in year t. For both instruments, the identification assumption is that transportation cost increases with the distance between trading partners and geographical distance is uncorrelated with demand shocks in the importing country.

[14] The bi-annual data were mostly used for European countries. There were not many BSE outbreaks in other countries.

Table 4.1 and 4.2 present the summary statistics on data used to estimate Eq. (1). From the tables, n refers to the number of importing countries in the data for each beef commodity. The number of importing countries ranges from 34 to 36 based on the HS6 level. The monthly GDP index and price index for food and beverage are only available on OECD Statistics. Due to lack of information regarding non-OECD member countries, the final data include some countries that were members of the OECD during 1996 through 2021.

Second Stage: Import Demand Shares by Country of Origin

In the second stage. we examine import demand by source for each HS code conditional on total import expenditure allocated to the HS code in the first stage. We use the LA/AIDS functional form to characterize this second stage demand. For importing country i, the second-stage demand for beef of a HS 6-digit code imported from source j is specified as:

$$W_{jt} = \alpha_j + \sum_k \gamma_{jk} \ln P_{kt} + \beta_j \ln\left(\frac{M_t}{P_t^*}\right) + \epsilon_{jt},\ \ j, k = 1,2, \dots ,6, \tag{4}$$

where W_{jt} is the budget share from source j, and P_{kt} refers to the unit value of imports from source k. M_t is the total expenditure allocated to this HS code in the first stage, P_t^* is the Stone price index for imports of the HS code of interest, and ϵ_{jt} is the residual. Specifically, the Stone's price index is calculated as:

$$\ln P_t^* = \sum_{k=1}^{n} \bar{W}_{kt} \ln P_{kt} \tag{5}$$

$\bar{W}_{kt}$ is the mean expenditure share of source k in total imports of a beef commodity in a country. The total number of sources of imported beef (j, k) is six including the top five beef exporters and all other countries. The top five beef exporters differ by beef commodity and over time. For instance, the United States topped the list of fresh/chilled beef (0201 at four-digit HS level) exporters for the year 2020, followed by Australia and Netherlands, while Brazil was the largest exporter of frozen beef (0202 at four-digit HS level) according to USDA (2020). The top five beef exporters in this study are selected by summing the import values of all importing countries in the final sample from 1996 to 2021. The following is the list of top five exporters by HS code:

1) 020110 (fresh or chilled bovine carcasses and half carcasses): Germany, Netherlands, Spain, Poland, France.
2) 020120 (fresh or chilled bovine cuts bone in): Germany, France, Netherlands, Poland, Canada.
3) 020130 (fresh or chilled bovine cuts boneless): United States, Canada, Australia, Ireland, Netherlands.
4) 020210 (frozen bovine carcasses and half carcasses): Ukraine, Belarus, Germany, Belgium, Australia.
5) 020220 (frozen bovine cuts bone in): United States, Australia, Uruguay, New Zealand, Germany.
6) 020230 (frozen bovine cuts boneless): Australia, Brazil, New Zealand, United States, Uruguay.

Table 4.2 presents the summary statistics of import unit values from the top exporters. Note that the top exporters have similar average unit values within each product line.

Including monthly dummy variables as specified in equation (1) in the first stage, the second-stage estimation model is specified as:[15]

[15] The dummy variables such as FTA, BSE, FMD were considered in the second-stage estimation. When estimating equation (6) for each importing country, the lack of variation in these dummy variables created collinearity problems. However, whenever included, the measured elasticities with the did not change significantly from the specification without the FTA, BSE, and FMD dummies. Additional explanation and logistics are provided in the next section.

$$W_{jt} = \alpha_j + \left(\sum_{\mathrm{r}} b_{rj} m_{rt}\right) + \sum_{k=1}^{n} \gamma_{jk} \ln P_{kt} + \beta_j \ln\left(\frac{M_t}{P_t^*}\right) + \epsilon_{jt}, \tag{6}$$

where $j, k = 1, \cdots, n$; and $n = 6$, including top 5 exporters and all others. An initial attempt was made to analyze panel data with country fixed effects as done in the first stage, but the above budget share was estimated for each importer country in the second stage. When using disaggregated level data such as HS6 level, it is highly likely that there is no import flow of a certain beef commodity from some countries. Broadly three factors could derive the zero consumption or zero trade flow: 1) variations in preference across samples; 2) infrequent purchasing; and 3) misreporting (Keen, 1986). Dealing with zero observations in the sample was challenging, and sometimes it was impossible to analyze the data due to insufficient observations. Adding specific value or any other replacement could lead to bias. In the end, the models were estimated for all the importers in the sample, then weighted average elasticities were measured.

Estimation Results

Stage-one Demand Results

Table 4.3 reports the first-stage import demand estimates for the six HS 6-digit beef commodities. The price elasticity of beef import demand ranges from -0.25 to -0.80, and all are statistically significant at the 1% level. Usually, frozen beef – HS 020210, HS 020220, and HS 020230 – are less elastic compared to fresh beef – 020110, 020120, and 020130 – with the exception of 020110. This pattern suggests that perishability might matter in the determination of price elasticity. Fresh or chilled meat has a shorter storage period than frozen meat, causing it to be more sensitive to price changes. As expected, the response to the food and beverage price index are significantly positive for all beef commodities and clearly show that there is a substitutive relationship between imported beef and other foods. The coefficient on the GDP index is positive and significant in all product lines, except 020210, indicating a positive income effect on beef imports.

The positive and statistically significant coefficients on FTA in the import demands for all except HS 020120 show that free trade agreements significantly increase beef imports. The coefficient on FTA in the import demand for HS 020120 is negative but not statistically significant. These results are consistent with the literature that finds trade facilitation is key to promoting international trade flows (Moïsé and Sorescu, 2013). While the event of domestic BES outbreaks has a negative effect on all beef imports, domestic cases of FMD have a positive and significant effect on three of the six beef commodities, two of which are frozen.

The opposing effects of BSE and FMD on import demand are not surprising. FMD is transmitted not only between cattle but also between other livestock. In fact, a significant number of FMD outbreaks occurred simultaneously in other livestock, such as pigs in most countries. This has made it difficult to supply substitutes for domestic beef. As mentioned earlier, Yang et al. (2012) also show that FMD-infected importers may increase pork imports. For instance, importers with a slaughter policy tend to increase pork imports due to the shortage of domestic supply. The results of Dinku and Matsuda’s study (2018) show that the BSE outbreaks in Japan might be the primary factor driving Japanese import demand away from beef to pork. Their study shows that demand for pork and chicken increased after BSE outbreaks in Japan, but the demand for pork fell following a FMD outbreak. The decline in beef demand caused by BSE outbreaks was also shown in other studies (Henson and Mazzocchi, 2002; Pennings et al., 2002; Paarlberg et al., 2005).

The difference in food safety risk associated with BSE and FMD may be another factor explaining the difference in the impact of these two diseases on import demand. Even though FMD is transmissible to humans in theory, it crosses the species barrier with much difficulty and with little effect. FMD does not have huge human health risks while BSE can be fatal (Prempeh et al., 2001). Therefore, the outbreaks of BSE could lead to much more significant declines in beef demand. Paarlberg et al. (2005) also report that consumers understand the distinctive nature of the two diseases. Consumers respond strongly to BSE while only moderately to FMD.

The coefficients on seasonal dummies indicate strong seasonality in demand, with higher imports in May through December for HS 020130 (bovine cuts boneless) and May through October for HS 020230 (bovine cuts boneless), while demand for HS 020110 (fresh or chilled carcasses and half-carcasses) does not show any seasonal patterns. The United States, Japan and Korea are the top importers for frozen beef, and beef barbecue is popular during summer in those countries, especially Japan and Korea. In December, beef consumption is the highest likely due to end-of-the-year parties as well as celebrations of holidays including Christmas.

Stage-two Demand Results

Recall that we estimate demand for each HS code imported from the top five exporting countries. Because country i does not always import from all five countries, there are a total of 31 combinations of importing sources, ranging from one to all five countries. We estimate the second-stage LA/AID model for each HS code and for each combination separately. Afterwards, the estimation results corresponding to the larger number combinations were selected to measure weighted average elasticities. In some cases, for the same importing country, the model was estimated for the same number of chosen countries but with different country variations. For example, for country i, the model estimation is available for two different combinations: USA and Brazil, and Spain and Uruguay. In this case, the model with a larger number of observations is selected to measure the weighted average.

Table 4.4 and 4.5 report the measured weighted average of own- and cross-price elasticities as well as expenditure elasticities for the six beef commodities. First of all, all the expenditure elasticities are positive and significant, while most own-price elasticities are negative as expected. Most expenditure elasticities are relatively close to 1, but for 020210 in the case of Ukraine, the elasticity is estimated as 2.027. This high value of elasticity is related to the limitations of the data. As seen in the table, despite analyzing the demand of each importing country for each combination, the analysis for the 020210 still shows the limitations of the analysis. There are not many importing countries that could be analyzed for this commodity as well as limited observations. Additionally, only the demand system model of Russia was used to estimate own- and cross- price elasticity between Ukraine and Belarus. Therefore, the relatively elastic expenditure elasticity of Ukraine may be biased. The number of observations for the commodity is less than that of other commodities. Furthermore, it is also difficult to estimate the expenditure share when only considering the top 5 importing sources. Cross-price elasticity could only be estimated in two sets: Ukraine and Belarus, and Germany and Belgium. However, only their own-price elasticities were significant, deeming cross-price elasticities insignificant.

The own price elasticity for 020110 is higher in France and Germany with a value of -1.2, while Netherlands' own price elasticity is relatively low compared to other exporters with a value of -0.669. For the case of 020130, the Unites States, top exporter of the commodity, has relatively higher own price elasticity compared to other countries. On the contrary, Ireland has lower elasticity. Uruguay's own-price elasticity for 020230 is measured as -2.142, which is the highest value among the exporting countries and given commodities.

The measured cross-price elasticities in the table show that geographically close countries serve as a substitute for each other: Germany-Spain, Poland-France, USA-Uruguay, Brazil-USA. Although the distance between the two countries is not close, it was measured that there are also mutual substitutions between France and Canada, and between Australia and Uruguay.

For 020120, the cross-price elasticity between Canada and Poland is extremely high with -32.405, implying that the import price from Poland has way larger effects on the import expenditure share on Canada in OECD countries. However, it is better to understand this figure as the impact of change within each country rather than the bilateral relationship. Prior to the discovery of BSE in Canada, cattle and beef production activities had been expanding because of growing demand among foreign consumers. In particular, the expenditure share on Canada in the US beef import market has exceeded 90%. Meanwhile, in the data used for the analysis, no Polish beef was imported into the United States during the same period. Most European countries have imported Polish beef, and Canadian beef exports to these countries were very small. According to the data used, the expenditure share on Canadian beef imports in European countries increased in 2010, but soon declined significantly and again remains at a marginal level. Polish beef prices continued to rise during this period, eventually resulting in very large elasticity, as shown in the table.

Conclusion

This article analyzes the world's beef import demand through a multi-stage approach with an aim to estimate the effects of domestic and international contexts on beef imports. In specific, free trade agreement and beef-related diseases such as FMD and BSE are expected to have a significant impact on beef imports in global markets, and these were included in the demand system of this study. In addition, price and expenditure elasticities by import sources were measured based on global scale flows. Results indicate that the domestic outbreaks of FMD increase the import demand for beef, whereas BSE outbreaks reduce the demand for beef imports. The results verify that there is a difference in perception of the two diseases among consumers. It is believed that concerns over BSE may have driven the decline in consumption of beef itself, which includes import and domestic demand.

In the second stage, the analysis of global import demand was conducted considering the top five importing sources, then own- and cross-price elasticities and expenditure price elasticities were measured. As some previous studies have pointed out, there have been several problems, including missing data, especially when estimating the model at the HS6 digit level in the second step. Using the Stone's price index, weighted average elasticities were measured by adopting the basic LA/AIDS methodology for each importing country. The measured elasticities imply that geographical locations close to each other have a great influence on the mutual substitution between the countries.

There have been many studies that analyze the demand system within a country using time series data, but this study analyzed the beef import demand considering the global market. Therefore, variables that can facilitate trade flows, such as free trade agreements, were considered and the import demand system of OECD countries was analyzed using panel data. The elasticities measured in the second stage demonstrate that beef imported from a country cannot necessarily be substituted for beef imported from other countries in the international market. Mutual substitutability was strong between geographically close import sources. New Zealand or Australian beef, which are recognized as relatively high-quality meat, seems to have relatively low substitutability compared to other exporting countries. Through these results, it is possible to evaluate the perception of the quality of each major beef exporting country in the world market, and it is expected that these countries will be able to suggest development plans to increase exports based on this evaluation.

Table 4.2. Summary Statistics

Variable		Mean	SD	Min	Max	Observation
Quantity (t)	020110	1,590	2,541	1 (kg)	15,042	
	020120	2,128	3,672	1 (kg)	25,978	
	020130	4,578	6,817	18 (kg)	49,497	
	020210	145	475	1 (kg)	6,769	
	020220	636	2,541	1.7 (kg)	41,746	020110:
	020230	5,349	13,778	10 (kg)	194,923	N=6,380
Price ($/kg)	020110	5.03	2.12	0.13	12.44	n=35
	020120	6.16	4.10	0.06	25.70	$\bar{T}$=182.3
	020130	8.82	5.41	0.13	33.85	
	020210	4.85	2.84	0.27	16.60	
	020220	5.35	3.17	0.06	20.86	020120:
	020230	5.90	3.34	0.33	20.26	N=8,774
Price index of food and beverage	020110	90.35	16.35	23.20	192.69	n=36
	020120	89.12	17.95	10.39	173.16	$\bar{T}$=243.72
	020130	87.71	19.79	23.20	241.84	
	020210	90.86	15.64	27.12	141.43	
	020220	88.22	20.38	23.20	241.84	020130:
	020230	86.99	20.81	23.20	241.84	N=9,324
GDP index	020110	99.90	1.83	78.89	105.37	n=36
	020120	99.91	1.71	78.89	105.37	$\bar{T}$=259
	020130	99.92	1.70	78.89	105.37	
	020210	99.89	1.84	82.65	105.37	
	020220	99.92	1.74	78.89	105.37	020210:
	020230	99.92	1.70	78.89	105.37	N=3,972
FTA	020110	0.88	0.33	0	1	n=33
	020120	0.84	0.37	0	1	$\bar{T}$=120.36
	020130	0.86	0.35	0	1	
	020210	0.95	0.21	0	1	
	020220	0.85	0.36	0	1	020220:
	020230	0.19	0.39	0	1	N=8,616
FMD	020110	0.02	0.13	0	1	n=36
	020120	0.02	0.15	0	1	$\bar{T}$=239.33
	020130	0.03	0.17	0	1	
	020210	0.02	0.14	0	1	
	020220	0.03	0.18	0	1	020230:
	020230	0.03	0.17	0	1	N=9,730
BSE	020110	0.32	0.47	0	1	n=36
	020120	0.26	0.44	0	1	$\bar{T}$=270.28
	020130	0.25	0.43	0	1	
	020210	0.38	0.48	0	1	
	020220	0.24	0.43	0	1	
	020230	0.24	0.43	0	1	

Table 4.2. Summary Statistics of Price Based on Importing Sources

	Variable	Mean	SD	Min	Max	Observation (N)	Country (n)
020110	Price_Germany	5.02	2.45	0.24	12.43	3,464	22
	Price_Netherlands	5.90	2.00	0.03	12.49	3,713	25
	Price_Spain	4.22	1.78	0.37	11.74	1,835	19
	Price_Poland	5.00	1.83	0.56	12.51	2,509	21
	Price_France	5.73	2.43	0.10	12.49	2,769	22
	Price_Others	5.03	2.09	0.13	12.44	5,876	35
020120	Price_Germany	5.89	3.79	0.16	25.70	4,669	26
	Price_France	7.61	5.15	0.15	25.68	3,747	26
	Price_Netherlands	7.35	4.32	0.82	25.76	5,026	30
	Price_Poland	5.42	2.84	0.67	25.66	3,663	23
	Price_Canada	9.47	6.46	0.92	25.83	858	20
	Price_Others	6.16	4.02	0.06	25.70	8,435	36
020130	Price_USA	15.52	8.86	0.68	33.84	3,174	27
	Price_Canada	13.39	9.06	0.35	33.75	1,949	24
	Price_Australia	13.74	7.55	0.15	33.81	3,253	27
	Price_Ireland	9.11	5.54	0.52	33.79	4,486	28
	Price_Netherlands	9.80	4.99	1.29	33.85	5,578	31
	Price_Others	8.94	5.46	0.13	33.85	8,910	36
020210	Price_Ukraine	3.42	0.78	1.61	4.84	131	1
	Price_Belarus	3.36	0.60	1.99	5.40	100	1
	Price_Germany	5.86	3.67	0.08	16.66	1,108	19
	Price_Belgium	5.38	3.03	0.00	16.34	594	13
	Price_Australia	3.63	2.53	0.63	14.96	133	10
	Price_Others	4.85	2.80	0.27	16.60	3,565	33
020220	Price_USA	7.09	4.22	0.21	20.80	1,673	24
	Price_Australia	5.53	4.03	0.16	20.88	1,807	22
	Price_Uruguay	4.95	4.64	0.53	20.47	772	21
	Price_New Zealand	7.42	5.59	0.27	20.82	1,637	24
	Price_Germany	6.16	3.98	0.23	20.80	3,491	27
	Price_Others	5.39	3.18	0.06	20.86	7,441	36
020230	Price_Australia	5.97	4.09	0.29	20.26	3,006	32
	Price_Brazil	9.57	4.60	0.07	20.27	3,792	29
	Price_New Zealand	8.26	5.50	0.04	20.26	3,504	30
	Price_USA	6.66	4.25	0.02	20.27	2,223	28
	Price_Uruguay	8.69	4.85	2.28	20.26	4,767	33
	Price_Others	5.55	3.20	0.33	20.21	9,203	36

Table 4.3. Estimated Results of First Stage: Import Demands

Variable	020110	020120	020130	020210	020220	020230
$Constant$	8.38*** (0.76)	8.28*** (0.68)	6.40*** (0.54)	8.49*** (1.80)	5.62*** (0.79)	10.82*** (0.49)
$\ln Price$	-0.27*** (0.08)	-0.80*** (0.05)	-0.78*** (0.04)	-0.36*** (0.11)	-0.62*** (0.05)	-0.25*** (0.04)
$P_{f\&b}$	0.02*** (0.00)	0.02*** (0.00)	0.04*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.01*** (0.00)
Y: GDP index	0.02*** (0.01)	0.04*** (0.01)	0.06*** (0.01)	-0.01 (0.02)	0.04*** (0.01)	0.02*** (0.00)
FTA	0.94*** (0.23)	-0.08 (0.11)	0.26*** (0.08)	1.26*** (0.34)	0.77*** (0.09)	0.84*** (0.05)
FMD	-0.13 (0.33)	-0.07 (0.18)	0.20* (0.10)	0.73 (1.11)	0.32** (0.14)	0.42*** (0.06)
BSE	-0.21*** (0.04)	-0.28*** (0.03)	-0.36*** (0.03)	-0.06 (0.09)	-0.38*** (0.04)	-0.08*** (0.02)
Feb	-0.09 (0.07)	-0.03 (0.06)	-0.08* (0.04)	0.23 (0.16)	0.06 (0.07)	-0.03 (0.04)
Mar	0.06 (0.07)	0.04 (0.06)	0.05 (0.04)	0.23 (0.16)	0.1 (0.07)	0.11*** (0.04)
Apr	-0.02 (0.07)	0.03 (0.06)	0.05 (0.04)	-0.00 (0.16)	-0.02 (0.07)	0.06 (0.04)
May	0.00 (0.07)	0.08 (0.06)	0.12*** (0.04)	0.28* (0.16)	0.08 (0.07)	0.09** (0.04)
Jun	-0.05 (0.07)	0.10* (0.06)	0.14*** (0.04)	0.00 (0.16)	0.08 (0.07)	0.10*** (0.04)
Jul	-0.09 (0.07)	0.11** (0.06)	0.13*** (0.04)	-0.04 (0.16)	0.04 (0.07)	0.10*** (0.04)
Aug	-0.01 (0.07)	0.07 (0.06)	0.11*** (0.04)	-0.19 (0.16)	0.13* (0.07)	0.09** (0.04)
Sep	0.10 (0.07)	0.06 (0.06)	0.11** (0.04)	0.01 (0.16)	0.05 (0.07)	0.10** (0.04)
Oct	0.05 (0.07)	0.09* (0.06)	0.16*** (0.04)	0.01 (0.16)	0.21*** (0.07)	0.08** (0.04)
Nov	0.05 (0.07)	-0.01 (0.06)	0.10** (0.04)	-0.00 (0.16)	0.09 (0.07)	0.02 (0.04)
Dec	0.05 (0.07)	0.03 (0.06)	0.20*** (0.04)	0.22 (0.16)	0.05 (0.07)	0.04 (0.04)
$Observations$	4,801	6,590	7,723	2,287	6,449	8,442

Notes: Standard errors in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 4.4. Estimated Price and Expenditure Elasticities

020110	Germany	Netherlands	Spain	Poland	France
Germany	-1.203***	0.129**	0.213**	-0.452***	-0.052
	(0.183)	(0.058)	(0.095)	(0.148)	(0.089)
Netherlands	0.286	-0.669***	0.063	-0.551***	-0.287***
	(0.185)	(0.110)	(0.091)	(0.174)	(0.106)
Spain	0.219**	-0.008	-0.975***	-0.261**	0.05
	(0.102)	(0.057)	(0.145)	(0.108)	(0.093)
Poland	-0.652***	-0.315***	-0.295**	0.15	0.175
	(0.169)	(0.077)	(0.114)	(0.276)	(0.112)
France	-0.013	-0.237***	0.145	0.222**	-1.239***
	(0.100)	(0.089)	(0.111)	(0.111)	(0.186)
Expenditure	0.966***	0.975***	0.83***	0.704***	1.193***
	(0.107)	(0.066)	(0.099)	(0.120)	(0.092)

020120	Germany	France	Netherlands	Poland	Canada
Germany	-0.970***	-0.082	-0.065	-0.12	2.339
	(0.125)	(0.056)	(0.073)	(0.161)	(4.268)
France	-0.274***	-1.101***	-0.066	0.568***	0.113***
	(0.097)	(0.225)	(0.118)	(0.179)	(0.066)
Netherlands	-0.218***	-0.068	-0.787***	-0.155	-0.059
	(0.082)	(0.074)	(0.121)	(0.163)	(0.063)
Poland	-0.084	0.21***	-0.017	-0.328	-32.405***
	(0.073)	(0.072)	(0.065)	(0.400)	(9.673)
Canada	0.028	0.057**	-0.037	-0.122***	-0.447
	(0.023)	(0.033)	(0.036)	(0.038)	(0.375)
Expenditure	1.341***	1.187***	0.752***	0.994***	0.856***
	(0.076)	(0.103)	(0.082)	(0.142)	(0.340)

020130	USA	Canada	Australia	Ireland	Netherlands
USA	-1.194***	0.639	0.34**	-0.228**	-0.077
	(0.153)	(1.010)	(0.150)	(0.094)	(0.050)
Canada	0.054	-0.819**	-0.046	-0.008	0.057
	(0.062)	(0.331)	(0.091)	(0.040)	(0.043)
Australia	0.081	-0.15	-1.169***	-0.115	0.017
	(0.076)	(0.182)	(0.182)	(0.071)	(0.051)
Ireland	-0.548***	0.012	-0.197*	-0.56***	0.061
	(0.187)	(0.038)	(0.112)	(0.170)	(0.074)
Netherlands	-0.091**	0.044	0.008	-0.007	-0.616***
	(0.040)	(0.038)	(0.040)	(0.143)	(0.135)
Expenditure	1.217***	0.862***	0.825***	1.041***	0.83***
	(0.086)	(0.187)	(0.133)	(0.120)	(0.073)

Table 4.5. Estimated Price and Expenditure Elasticities (*continued*)

020120	**Ukraine**	**Belarus**	**Germany**	**Belgium**	**Australia**
Ukraine	-1.650**	0.065	-	-	-
	(0.702)	(0.081)			
Belarus	-0.498	-0.977***	-	-	-
	(0.488)	(0.096)			
Germany	-	-	-1.023***	0.048	-
			(0.215)	(0.055)	
Belgium	-	-	-0.075	-1.082***	-
			(0.246)	(0.138)	
Australia	-	-	-	-	-
Expenditure	2.027***	0.911***	1.186***	0.728***	-
	(0.224)	(0.033)	(0.126)	(0.040)	
020220	**USA**	**Australia**	**Uruguay**	**New Zealand**	**Germany**
USA	-1.17***	0.321**	0.144***	0.096	-0.741***
	(0.072)	(0.128)	(0.055)	(0.154)	(0.219)
Australia	-0.025	-0.928***	0.224*	0.048	-
	(0.062)	(0.161)	(0.132)	(0.190)	
Uruguay	1.695*	0.576*	0.028	0.023	0.045
	(0.957)	(0.338)	(0.294)	(0.325)	(0.111)
New Zealand	-0.083**	0.006	0.034	-0.300	0.000
	(0.039)	(0.100)	(0.139)	(0.216)	(0.013)
Germany	-1.094**	0.994***	-0.209	-0.039	-1.312***
	(0.541)	(0.331)	(0.378)	(0.160)	(0.130)
Expenditure	1.263***	0.528***	0.568***	0.347***	1.106***
	(0.041)	(0.087)	(0.085)	(0.102)	(0.061)
020230	**Australia**	**Brazil**	**New Zealand**	**USA**	**Uruguay**
Australia	-1.717***	0.258	0.599**	-0.711*	0.709
	(0.258)	(0.211)	(0.258)	(0.408)	(0.436)
Brazil	1.167*	-1.684***	3.287***	3.356**	0.531
	(0.676)	(0.478)	(1.108)	(1.656)	(0.918)
New Zealand	0.258	0.353*	-1.440***	-0.123	0.494
	(0.162)	(0.202)	(0.321)	(0.175)	(0.396)
USA	-0.508**	0.274*	-0.153	-0.368	-1.373
	(0.213)	(0.140)	(0.295)	(0.386)	(1.019)
Uruguay	0.152	0.046	0.234	-0.160	-2.142***
	(0.115)	(0.314)	(0.203)	(0.211)	(0.661)
Expenditure	1.154***	1.320***	0.801***	0.999***	0.623***
	(0.080)	(0.092)	(0.072)	(0.189)	(0.162)

CHAPTER 5

CONCLUSION

This dissertation research features three separate studies that deal with international trade of agricultural commodities and the application of machine learning methods. Predictions of future economic conditions is critical to progress in agriculture, as it supports the livelihood of millions of farmers. Therefore, stakeholders have sought information on changing patterns of economic activity including trade, price and demand patterns and the underlying economic factors. Meanwhile, conventional statistical and econometric techniques have been challenged in modeling economic relationships due to the emerging volume of finer data and complexities in such relationships. The application of advanced methods including machine learning in this dissertation allows for alternative and robust specifications of complex economic relationships with a greater accuracy than traditional approaches. They are also flexible to identify which variables among the set of features provide more information to understanding patterns of economic activity.

The first study examines the effect of international market information – especially price volatility and inventories – on domestic market price dynamics in the case of agricultural commodities. A structural model is employed to uncover relationships among commodity price, price volatility, inventories, and convenience yield. The resulting empirical model includes equations for convenience yield and price formation as functions of inventories (domestic and international), volatility (domestic and international), trade openness, and other controls. LASSO, a regularization and variable-selection method, is used in the first stage to identify the best instrument set for each endogenous variable given large instrument set. The main advantage of the use of LASSO is that it provides good approximations to the optimal instruments when the number of available instruments is large. Another advantage of the use of LASSO as a variable selection technique is that any prior knowledge about the identity of the most relevant instruments is not necessary. Then, a nonlinear approach is applied to estimate the second stage of the model. The results of this study show that international market information plays a critical role in domestic market price dynamics. International price volatility has a stronger effect on domestic prices than that of international inventories. Especially, the results show the potential to utilization of international market information including WASDE (World Agricultural Supply and Demand Estimates) on domestic price dynamics models in a globalized economy.

The second study investigates the bilateral trade of beef using traditional Poisson Pseudo Maximum Likelihood (PPML) methods and new ML analysis and makes a modest attempt to compare the advantages and disadvantages of them. The PPML was developed recently to estimate the gravity model to overcome the potential bias caused by the large proportion of zero observations in bilateral trade data. However, PPML techniques have been challenged in modeling the gravity relationship, especially the need for a large number of pair-wise fixed effects. The main advantage of the PPML is its strong predictive power as it has been used in many studies for a long time period. However, the powerful prediction power mainly comes from the inclusion of a large number of fixed effects. ML models have the advantage in selecting explanatory variables when a long list of explanatory variables is to be considered. The ML models show another advantage in doing out-of-sample prediction, while PPML method does not do out-of-sample prediction. Although the ML models have some limitations such as the lack of a theoretical explanation, they have sufficient potential for development that can be used in various analysis in the future.

The last study analyzes the world's beef import demand at a disaggregated level of beef commodities through a multi-stage approach. The multi-stage budgeting approach begins with an estimation of the demand function for imported beef including cattle diseases such as Foot and Mouth Disease (FMD) and Bovine Spongiform Encephalopathy (BSE) as well as free trade agreements into the demand system. The second stage estimates individual demand functions for separate importing sources, the top five exporters to each importing country. Using the Stone's price index, weighted average elasticities were measured by adopting the basic LA/AIDS methodology for each importing country. Results indicate that the domestic outbreaks of FMD increase the import demand for beef, whereas BSE outbreaks reduce the demand for beef imports. The elasticities measured in the second stage demonstrate that beef imported from a country cannot necessarily be substituted for beef imported from other countries in the international market. All three studies document the emerging advantages of advanced techniques and finer data in quantifying economic relationships.

REFERENCES

Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of cleaner production*, *203*, 810-821.

Akanni, L. O. (2020). Returns and volatility spillover between food prices and exchange rate in Nigeria. Journal of Agribusiness in Developing and Emerging Economies, 10(3), 307-325.

Allen, T., Arkolakis, C., & Takahashi, Y. (2020). Universal gravity. Journal of Political Economy, 128(2), 393-433.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. The American economic review, 69(1), 106-116.

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. American economic review, 93(1), 170-192.

Anderson, E. (2014). Time differences, communication and trade: Longitude matters II. Review of World Economics, 150(2), 337-369.

Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica, 77(4), 1229-1279.

Baier, S. L., & Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade?. Journal of international Economics, 71(1), 72-95.

Baldwin, R., & Taglioni, D. (2006). Gravity for dummies and dummies for gravity equations (No. w12516). National bureau of economic research.

Baldwin, R., & Taglioni, D. (2007). Trade effects of the euro: A comparison of estimators. Journal of Economic Integration, 780-818.

Beck, S. (2001). Autoregressive conditional heteroscedasticity in commodity spot prices. Journal of Applied Econometrics, 16(2), 115-132.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, 80(6), 2369-2429.

Belson, W. A. (1959). Matching and prediction on the principle of biological classification. Journal of the Royal Statistical Society: Series C (Applied Statistics), 8(2), 65-75.

Bergstrand, J. H. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. The review of economics and statistics, 474-481.

Blundell, R., Pashardes, P., & Weber, G. (1993). What do we learn about consumer demand patterns from micro data?. The American Economic Review, 570-597.

Cafiero, C., & Wright, B. D. (2011). The empirical relevance of the competitive storage model. Journal of Econometrics, 162(1), 44-54.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. The Econometrics Journal, 20(3), S1-S13.

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, 61-69.

Chopra, A., & Bessler, D. A. (2005, April). Impact of BSE and FMD on beef industry in UK. In NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, St. Louis, Missouri, April (pp. 18-19).

Coffey, B., Mintert, J., Fox, J. A., Schroeder, T. C., & Valentin, L. (2005). The economic impact of BSE on the US beef industry: Product value losses, regulatory costs, and consumer reactions.

Conconi, P., Magerman, G., & Plaku, A. (2020). The gravity of intermediate goods. Review of Industrial Organization, 57(2), 223-243.

De Sousa, J. (2012). The currency union effect on trade is decreasing over time. Economics Letters, 117(3), 917-920.

Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. The American economic review, 70(3), 312-326.

Deaton, A., & Laroque, G. (1992). On the behaviour of commodity prices. The review of economic studies, 59(1), 1-23.

Dey, M. M., Alam, M. F., & Paraguas, F. J. (2011). A multistage budgeting approach to the analysis of demand for fish: An application to inland areas of Bangladesh. Marine Resource Economics, 26(1), 35-58.

DINKU, S. Y., & MATSUDA, T. (2018). Evaluating the Impact of the BSE and FMD Outbreaks on Meat Demand An Engel Curve Analysis of Japanese Daily Data. Journal of the Japanese Society of Agricultural Technology Management, 25(1), 1-12.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. Econometrica, 70(5), 1741-1779.

Fan, S., Wailes, E. J., & Cramer, G. L. (1995). Household demand in rural China: A two-stage LES-AIDS model. American Journal of Agricultural Economics, 77(1), 54-62.

FAO (2020), “Agricultural Market Information System”, available at: http://www.amis-outlook.org/indicators/prices/en/ (accessed 2 November 2020).

Feenstra, R. Advanced International Trade: theory and evidence, 2004. Princeton and Oxford: Princeton University Press, pp. xii, Advanced, 484.

Felipe, J., & Kumar, U. (2012). The role of trade facilitation in Central Asia: A gravity model. Eastern european economics, 50(4), 5-20.

Fernández-Val, I., & Vella, F. (2011). Bias corrections for two-step fixed effects panel data estimators. Journal of Econometrics, 163(2), 144-162.

Feyrer, J. (2009). Distance, trade, and income-the 1967 to 1975 closing of the suez canal as a natural experiment (No. w15557). National Bureau of Economic Research.

Figueiredo, E., Lima, L. R., & Orefice, G. (2020). Migration, trade and spillover effects. Journal of Comparative Economics, 48(2), 405-421.

Fletcher, M. E. (1958). The Suez Canal and world shipping, 1869-1914. The Journal of Economic History, 18(4), 556-573.

Fontagné, L., Martin, P., & Orefice, G. (2018). The international elasticity puzzle is worse than you think. Journal of International Economics, 115, 115-129.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

Ghatak, S., Silaghi, M. I. P., & Daly, V. (2009). Trade and migration flows between some CEE countries and the UK. The Journal of International Trade & Economic Development, 18(1), 61-78.

Gillson, I, and Fouad, A. (Eds). (2014), Trade policy and food security: improving access to food in developing countries in the wake of high world prices, World Bank Publications, Washington, DC

Glick, R., & Rose, A. K. (2016). Currency unions and trade: A post-EMU reassessment. European Economic Review, 87, 78-91.

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

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. Review of Finance, 17(1), 35-105.

Gouel, C. (2012). Agricultural price instability: a survey of competing explanations and remedies. Journal of economic surveys, 26(1), 129-156.

Gouel, C. (2016). Trade policy coordination and food price volatility. American Journal of Agricultural Economics, 98(4), 1018-1037.

Gouel, C., Gautam, M., & Martin, W. J. (2016). Managing food price volatility in a large open country: the case of wheat in India. Oxford Economic Papers, 68(3), 811-835.

Grant, J. H., Ning, X., & Peterson, E. B. (2018). Trade elasticities and trade disputes: New evidence from tariffs and relative preference margins.

Greene, W. H. (2012), Econometric analysis, 7th ed, Boston, MA:.

Gurevich, T., & Herman, P. (2018). The dynamic gravity dataset: 1948-2016. US International Trade Commission, Office of Economics Working Paper.

Gurevich, T., & Herman, P. (2018). The dynamic gravity dataset: 1948–2016. US International Trade Commission, Office of Economics Working Paper.

Guta, D. D. (2012). Application of an almost ideal demand system (AIDS) to Ethiopian rural residential energy use: Panel data evidence. Energy policy, 50, 528-539.

Hansen, C., Hausman, J., & Newey, W. (2008). Estimation with many instrumental variables. Journal of Business & Economic Statistics, 26(4), 398-422.

Henson, S., & Mazzocchi, M. (2002). Impact of bovine spongiform encephalopathy on agribusiness in the United Kingdom: Results of an event study of equity prices. American Journal of Agricultural Economics, 84(2), 370-386.

Hoekman, B., & Nicita, A. (2011). Trade policy, trade costs, and developing country trade. World Development, 39(12), 2069-2079.

Jayasinghe, S., Beghin, J. C., & Moschini, G. (2010). Determinants of world demand for US corn seeds: the role of trade costs. American Journal of Agricultural Economics, 92(4), 999-1010.

John, V., Liu, Z., Guo, C., Mita, S., & Kidono, K. (2015, November). Real-time lane estimation using deep features and extra trees regression. In Image and Video Technology (pp. 721-733). Springer, Cham.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Kawashima, S., & Puspito Sari, D. A. (2010). Time-varying Armington elasticity and country-of-origin bias: from the dynamic perspective of the Japanese demand for beef imports. Australian Journal of Agricultural and Resource Economics, 54(1), 27-41.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, 3146-3154.

Kee, H. L., Nicita, A., & Olarreaga, M. (2008). Import demand elasticities and trade distortions. The Review of Economics and Statistics, 90(4), 666-682.

Keen, M. (1986). Zero expenditures and the estimation of Engel curves. Journal of Applied Econometrics, 1(3), 277-286.

Kepaptsoglou, K., Karlaftis, M. G., & Tsamboulas, D. (2010). The gravity model specification for modeling international trade flows and free trade agreement effects: a 10-year review of empirical studies. The open economics journal, 3(1).

Khandelwal, P. (2017). Which algorithm takes the crown: Light GBM vs XGBOOST?. Analytics Vidhya, 12.

Knight-Jones, T. J., & Rushton, J. (2013). The economic impacts of foot and mouth disease–What are they, how big are they and where do they occur?. Preventive veterinary medicine, 112(3-4), 161-173.

Knoll, S., Padula, A. D., Barcellos, J. O. J., Pumi, G., Zhou, S., & Zhong, F. (2018). The role of distance on international commerce: the impact on Sino-Brazilian beef trade. Journal of Agribusiness in Developing and Emerging Economies.

Kompas, T., Nguyen, H. T. M., & Ha, P. V. (2015). Food and biosecurity: livestock production and towards a world free of foot-and-mouth disease. Food Security, 7(2), 291-302.

Kuik, O., Branger, F., & Quirion, P. (2019). Competitive advantage in the renewable energy industry: Evidence from a gravity model. Renewable energy, 131, 472-481.

Larch, M., Wanner, J., Yotov, Y. V., & Zylkin, T. (2019). Currency unions and trade: A PPML re-assessment with high-dimensional fixed effects. Oxford Bulletin of Economics and Statistics, 81(3), 487-510.

Lence, S. H., & Hayes, D. J. (2002). US farm policy and the volatility of commodity prices and farm revenues. American Journal of Agricultural Economics, 84(2), 335-351.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

Liu, P., Lien, K., & Asche, F. (2016). The impact of media coverage and demographics on the demand for Norwegian salmon. Aquaculture Economics & Management, 20(4), 342-356.

Lloyd, T. A., McCorriston, S., Morgan, C. W., & Rayner, A. J. (2006). Food scares, market power and price transmission: the UK BSE crisis. European Review of Agricultural Economics, 33(2), 119-147.

Lopez, A. O. (2020). Protectionism and Trade in Renewable Energy Infrastructure.

Martin, W. (2020). Making Gravity Great Again. Working Paper, International Food Policy Research Institute.

Matthews, A. (2014). Trade rules, food security and the multilateral trade negotiations. European Review of Agricultural Economics, 41(3), 511-535.

Melitz, J. (2007). North, South and distance in the gravity model. European Economic Review, 51(4), 971-991.

Miranda, M. J., & Glauber, J. W. (1993). Estimation of dynamic nonlinear rational expectations models of primary commodity markets with private and government stockholding. The review of economics and statistics, 463-470.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning. MIT press.

Moïsé, E., & Sorescu, S. (2013). Trade facilitation indicators: The potential impact of trade facilitation on developing countries' trade.

Moïsé, E., & Sorescu, S. (2013). Trade facilitation indicators: The potential impact of trade facilitation on developing countries' trade.

Muth, J. F. (1961). Rational expectations and the theory of price movements. Econometrica: Journal of the Econometric Society, 315-335.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica: Journal of the Econometric Society, 1-32.

Osborne, T. (2004). Market news in commodity price theory: Application to the Ethiopian grain market. The Review of Economic Studies, 71(1), 133-164.

Paarlberg, P. L., Lee, J. G., & Seitzinger, A. H. (2005). Economic modeling of livestock disease outbreaks. International Food and Agribusiness Management Review, 8(1030-2016-82469), 62-77.

Pennings, J. M., Wansink, B., & Meulenberg, M. T. (2002). A note on modeling consumer reactions to a crisis: The case of the mad cow disease. International Journal of research in marketing, 19(1), 91-100.

Pietola, K., Liu, X., and Robles, M. (2010), “Price, inventories, and volatility in the global wheat market”, IFPRI Discussion Paper 00996, International Food Policy Research Institute (IFPRI),Washington, DC, June.

Pindyck, R. S. (2004). Volatility and commodity price dynamics. Journal of Futures Markets: Futures, Options, and Other Derivative Products, 24(11), 1029-1047.

Prempeh, H., Smith, R., & Müller, B. (2001). Foot and mouth disease: the human consequences: The health consequences are slight, the economic ones huge. Bmj, 322(7286), 565-566.

Ramírez, A. (2013). A multi-stage almost ideal demand system: The case of beef demand in Colombia. Revista Colombiana de Estadística, 36(1), 23-42.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804-818.

Schierhorn, F., Meyfroidt, P., Kastner, T., Kuemmerle, T., Prishchepov, A. V., & Müller, D. (2016). The dynamics of beef trade between Brazil and Russia and their environmental implications. Global Food Security, 11, 84-92.

Sévi, B. (2015). Explaining the convenience yield in the WTI crude oil market using realized volatility and jumps. Economic Modelling, 44, 243-251.

Sha, S., Santos, J. I., Roheim, C. A., & Asche, F. (2015). Media coverage of PCB contamination of farmed salmon: The response of US import demand. Aquaculture economics & management, 19(3), 336-352.

Shanafelt, D. W., & Perrings, C. A. (2017). Foot and mouth disease: the risks of the international trade in live animals. Rev. Sci. Tech. l'OIE, 36, 839-865.

Sharma, R., Kamble, S. S., Gunasekaran, A., Kumar, V., & Kumar, A. (2020). A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Computers & Operations Research, 119, 104926.

Shepherd, B. (2013). Gravity model of international trade: A user guide.

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

Steiger, C. (2006). Modern beef production in Brazil and Argentina. Choices, 21(316-2016-6394), 105-110.

Tinbergen, J. (1962). Shaping the world economy; suggestions for an international economic policy.

Traore, B. B., Kamsu-Foguem, B., & Tangara, F. (2017). Data mining techniques on satellite images for discovery of risk areas. Expert Systems with Applications, 72, 443-456.

USDA, Economic Research Service (2010), “Retail food prices are less volatile than farm prices”, available at: https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58356 (accessed 16 Feburary 2021).

USDA, Economic Research Service (2020): https://www.ers.usda.gov/data-products/agricultural-trade-multipliers/2018-data-overview/ accessed August 05, 2021.

USDA, Foreign Agricultural Service (2019), “Livestock and Poultry: World Markets and Trade”, available at: https://usda.library.cornell.edu/concern/publications/73666448-x?locale=en (accessed 2 November 2020).

Varian, H. R. (2014). Big data: New tricks for econometrics. Journal of Economic Perspectives, 28(2), 3-28.

Webb, M., Gibson, J., & Strutt, A. (2018). The impact of diseases on international beef trade: Market switching and persistent effects. Food Policy, 75, 93-108.

Weidner, M., & Zylkin, T. (2021). Bias and consistency in three-way gravity models. Journal of International Economics, 132, 103513.

Wieck, C., & Holland, D. W. (2010). The economic effect of the Canadian BSE outbreak on the US economy. Applied Economics, 42(8), 935-946.

Yang, S. H., Reed, M. R., & Saghaian, S. H. (2012). International pork trade and foot-and-mouth disease (No. 323-2016-11526).

Yin, F., & Hamori, S. (2011). Estimating the import demand function in the autoregressive distributed lag framework: the case of China. Economics bulletin, 31(2), 1576-1591.

Yoshida, N. (2017). Local institutions and global value chains: Development and challenges of shrimp aquaculture export industry in Vietnam. Journal of Agribusiness in Developing and Emerging Economies.

Yotov, Y. V., Piermartini, R., Monteiro, J. A., & Larch, M. (2016). An advanced guide to trade policy analysis: The structural gravity model. Geneva: World Trade Organization.