ABSTRACT

ZHANG, YIFEI. Three Essays on Agricultural Risk and Price Analysis. (Under the direction of Dr. Barry K. Goodwin).

The three essays in this dissertation investigate actuarial risks in agricultural production, the substitutability between softwood lumber imported from Canada and domestically produced lumber in the United States (U.S.), and the effect of the COVID-19 pandemic on the U.S. manufactured wood industry.

In Chapter 1, we propose to use Copula models to determine the pricing of agricultural insurance policies. The profitability of agricultural production is typically realized at the end of the production season. This implies that producers are making investment decisions under many unexpected risks. A correct assessment of agricultural risks has discernible effects on the actuarial aspect of risk management and the determination of pricing and insurance premiums. However, quantifying these risks becomes challenging when they are driven by factors that are spatially correlated. Copula models provide a flexible framework for modeling the dependence structure between different variables. This approach is especially valuable in the context of agricultural insurance research as it allows us to characterize the underlying dependence relationships and provide a more accurate representation of the risks. We explore the implications of multivariate copula models in two empirical studies for both crop and livestock insurance, demonstrating alternative approaches to model portfolio risks under different scenarios. Our results show that the standard Gaussian assumption currently adopted by USDA-RMA may underprice crop insurance portfolio risks, especially with high price volatility.

Chapter 2 assesses the substitutability between Canadian softwood lumber imports and domestically produced softwood lumber in the U.S. The U.S.-Canada softwood lumber dispute stands as the most protracted trade conflict between the two countries, which begins when U.S. lumber producers claim Canadian lumber imports are unfairly subsidized by the government, thereby, harming the domestic lumber industry. In response to the petition, the U.S. Department of Commerce has imposed countervailing and antidumping duties on Canadian softwood lumber imports to protect the U.S. market. An important aspect concerning this trade dispute is the substitutability of lumber products between the two countries. Canadian lumber imports are detrimental to the U.S. industry only when the products are considered substitutes. In such a scenario, the imposition of trade restrictions would have a significant influence on the domestic market. Motivated by this

consideration, we investigate the dynamic relationships between prices for Spruce-Pine-Fir (SPF), a softwood species primarily imported from Canada, and Southern Yellow Pine (SYP), one of the most popular softwood species domestically produced in the US, using threshold models. We examine the nonlinear relationship by utilizing the logged difference between the two lumber prices as the forcing variable. Our empirical findings indicate that the degree of substitutability between the two lumber species becomes more pronounced as the price difference increases.

In Chapter 3, we estimate the impact of the COVID-19 pandemic on the U.S. manufactured wood markets. The COVID-19 pandemic introduces unprecedented volatility into the U.S. manufactured wood product industry. This paper assesses the large price swings in manufactured wood products, focusing on SPF, SYP, and Oriented Strand Board (OSB) sheathing. We argue that disruptions in the labor market and housing sector contribute significantly to changes in price volatility. To comprehend the pandemic's impact on the manufactured wood product industry, we conduct an event study to create a world without the presence of the pandemic to compare to the actual observed data. We utilize vector autoregressive (VAR) models and vector error correction (VEC) models including various macroeconomic exogenous variables. Furthermore, we find that SPF is considered a substitute for SYP but not vice versa, and OSB is a complement to both softwood lumber products.

Three Essays on Agricultural Risk and Price Analysis

by
Yifei Zhang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina
2024

APPROVED BY:

Dr. Junjie Wu	Dr. Jeffrey P. Prestemon
Dr. Ayse Kabukcuoglu Dur	Dr. Barry K. Goodwin Chair of Advisory Committee

DEDICATION

This dissertation is dedicated to my mother, for always trusting me, encouraging me, and loving me unconditionally.

BIOGRAPHY

Yifei Zhang was born in Luoyang, China. She started studying abroad in the United States at the age of 15 and attended Hernando Christian Academy in Harrisonburg, FL. Upon graduating from high school, she received her bachelor's degree in Economics from James Madison University in 2017 and earned a master's degree in Economics from Miami University, Oxford in 2018. Later that year, she joined North Carolina State University to pursue a Ph.D. in Economics under the guidance of Dr. Barry K. Goodwin.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Barry K. Goodwin, for his continued support and advice during my Ph.D. journey. His patience, understanding, and encouragement were invaluable. Thank you for instilling confidence in me and contributing to my growth as a researcher. The completion of this degree could not have been possible without Barry's expertise and guidance.

I extend my appreciation to my dissertation committee members, Dr. Jeffrey P. Prestemon, Dr. Junjie Wu, and Dr. Ayse Kabukcuoglu Dur, for their dedicated service on my dissertation committee. I am grateful for the insightful comments and valuable suggestions they provided that significantly improved my research.

I would like to thank my host family, who made the U.S. feel like home. Grandma and Janessa, thank you for always welcoming me back as part of the family. Meeting you during high school has been a truly fortunate and meaningful experience.

Finally, I want to express my heartfelt gratitude to my family for their unwavering support. I feel very blessed to have been raised in an environment filled with love and encouragement. Thank you for always having faith in me and being proud of me. A special thank you to all my friends for taking care of me. My time at NC State would not have been the same without your care and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

CHARACTERIZING SYSTEMIC RISKS USING COPULAS: EVIDENCE FROM THE U.S. AGRICULTURAL INSURANCE MARKET

1.1 Introduction

Risk management plays an important role in agricultural business decision-making. Due to the nature of agricultural production, farmers and ranchers must make ex-ante decisions and realize profits at the end of the production process, making it inherently risky. Uncertainties arise from various random shocks that agricultural products face, such as extreme weather events, diseases, and price fluctuations. Therefore, crop and livestock insurance programs have become major fixtures as mechanisms intended to manage such risks and stabilize producers' income.

The United States (U.S.) Federal Crop Insurance Program (FCIP), overseen by the U.S.

Department of Agriculture's Risk Management Agency (USDA, RMA), takes a substantial role as a major policy instrument for managing these risks. Simultaneously, it provides significant financial support to the U.S. farm sector. According to the Federal Crop Insurance Corporation (FCIC) Summary of Business Report by USDA RMA (2024), federal crop insurance participation has steadily increased over the past decades. In 2011, the net acres insured was around 265 million acres. This number rose to 290 million acres in 2016 and exceeded 444 million acres in 2021, covering more than 80% of eligible planted acres for commodity crops[1] (Tsiboe and Turner 2023). Large crop insurance participation implies a large amount of liability. In 2010, RMA managed over $114 billion worth of insurance liability, and this figure increased to approximately $137 billion in 2021 (USDA RMA 2024). Increases in participation and liability describe the importance of analyzing the risk profile associated with agricultural insurance programs for the federal government to determine the budgetary costs.

An interesting characteristic of the U.S. agricultural insurance program is the collaborative relationship between the federal government, specifically the FCIC, and private insurance companies. Insurance products are sold through a public-private partnership to administer and deliver coverage for farmers and ranchers. The FCIC provides financial assistance, including subsidies and reinsurance to private insurance companies. This collaborative insurance practice is carried out through the Standard Reinsurance Agreements (SRA) and the Livestock Price Reinsurance Agreements (LPRA). Under these reinsurance agreements, private insurance companies, often known as Approved Insurance Providers (AIPs), have the flexibility to select their insurance pool by retaining certain policies and ceding others to an "assigned risk pool", which places the majority of risk in the hands of the federal government (Ker and McGowan 2000). While reinsurance agreements underline the government's financial commitment to the insurance program, they create a complex system of risk sharing as they only cover a portion of the underwriting risks that the AIPs may bear, making risk selection a major focus in both the industry and academic research.

The RMA is responsible for pricing insurance policies sold by AIPs, and at the same time, the AIPs cannot discriminate among eligible producers. To ensure the participation of the private sector, insurance companies receive a subsidy for a portion of their administrative and operating expenses from the federal government (Ker and McGowan 2000). Furthermore, insurance companies can strategically transfer the majority of the liability of any undesirable policies by placing each policy into one of the three funds: assigned

[1]Commodity crops are barley, corn, cotton, flaxseed, oat, peanut, potato, rice, rye, sorghum, soybean, sugarbeet, sugarcane, sunflower, sweet potato, tobacco, and wheat (Tsiboe and Turner 2023).

risk, developmental, or commercial, for which the extent of risk sharing varies significantly across these funds (Ker and McGowan 2000). This implies that AIPs have the advantage of asymmetric information and can engage in adverse selection to retain more underwriting gains than losses[2]. In addition, the AIPs are allowed to seek reinsurance through arrangements with international reinsurance companies[3] for the remaining portion of the liability that has not been ceded to the FCIC.

Agricultural insurance programs, especially the crop insurance program, account for a significant amount of treasury transfers from taxpayers. For example, the average annual cost of the insurance program for 2011-2020 is approximate $6.1 billion (Tsiboe and Turner 2023). The cost of crop insurance programs mainly arises from two aspects: the government subsidies provided to cover farmers' insurance premiums, exceeding 60% of the total premium in recent years, and the compensation to AIPs (GAO 2023). Some literature has criticized the effectiveness of subsidized crop insurance (Smith 2011; Holly Wang and Zhang 2003; Goodwin and Vado 2007). In particular, the argument is that the substantial subsidies offered to agricultural producers and private insurance corporations could lead to inefficiencies, and the government's involvement in agricultural insurance programs could crowd out the private market. Subsidized crop insurance can distort the market by transferring a significant amount of wealth from taxpayers to farmers and private insurance companies (Goodwin and Smith 2013). In addition, the U.S. Government Accountability Office (GAO) has suggested Congress reduce subsidies to high-income participants and adjust compensation to AIPs to reflect market conditions (GAO 2023). On the contrary, researchers argue that government support is necessary since the private market cannot provide efficient risk management mechanisms due to the significant systemic risk associated with extreme weather events Coble et al. (2003).

The study of systemic risks in agricultural insurance is particularly relevant since the 2018 Farm Bill expired at the end of FY2023, and Congress is considering a new farm bill (Monke 2023). Quantifying the degree of systemic risk is an important aspect of projecting future government expenditure. Since the underlying pricing of insurance policies depends on the accurate modeling of random variables that determine losses (Goodwin et al. 2018), recent research has focused on the appropriate measurement of insurance portfolio risks with models capable of capturing tail dependence, asymmetric risks, and other characteristics of joint likelihood functions. Such approaches are capable of representing the

[2]For a more detailed analysis of the reinsurance programs, please refer to Miranda and Glauber (1997).

[3]Major private reinsurers that assume a significant share of the risk associated with the U.S. crop insurance portfolio include, among others, MunichRe, SwissRe, and PartnerRe.

dependence structure in portfolios of agricultural insurance products and have become a specific area of interest (Frees and Valdez 1998; Goodwin and Mahul 2004; Ramsey and Goodwin 2019; Ramsey et al. 2019; Sun and Wang 2022). Most of these models are expressed in terms of copulas, providing a straightforward and flexible representation of joint likelihood functions for a portfolio of random variables with known or estimated marginal distributions. The application of copula-based models to the representation of multivariate risks has become a significant focus of applied research.

The objective of the USDA-RMA actuarial procedure is not profit-driven, as it would be for private insurance companies in the competitive market. As described by provisions in the Federal Crop Insurance Act §1508 *(i)*, the RMA is targeting insurance premium rates that assure the actuarial soundness of insurance operations while maximizing participation and ensuring equity for agricultural producers. From an insurance provider's perspective, actuarial fairness is achieved when the total premium collected equals the expected claim payments. Since Congress requires an expected program-wide loss ratio of no more than one[4], this emphasizes the importance of conducting and maintaining actuarially fair rates. Linearly correlated risks and constant dependence structure are commonly assumed in an actuarial framework (Embrechts et al. 2002). On a similar note, Ramsey et al. (2019) have pointed out that the current RMA's pricing of crop revenue insurance relies on the normality assumption and crop insurance policies are priced without the consideration of yield dependency. This implies that the determination of actuarially sound premium rates is in question.

The USDA-RMA offers a variety of crop insurance programs, with Crop Revenue Coverage (CRC) being the most widely available revenue protection policy. At the county level, the RMA has released Area Risk Protection Insurance (ARPI)[5] to provide coverage based on the yield or revenue of an entire area (generally a county). Modeling a county-level portfolio is challenging since one must consider the underlying dependence relationships across county-level yields, as yield risks are spatially correlated. As noted by Goodwin (2001), the systemic risks become more significant under extreme weather events, making the analysis of dependence structure particularly relevant in determining the resilience of county-level yields.

[4]Loss ratio is the ratio between the total claim payments and the total premium earned. A loss ratio of less than one indicates fewer program expenditures and, therefore, less taxpayer contribution. The USDA RMA described this program-wide loss ratio requirement on their "Frequently Asked Question" page for Corn and Soybeans should be no more than one. There is no target set for Livestocks.

[5]ARPI is an insurance plan that replaces the Group Risk Plan (GRP) and the Group Risk Income Protection Plan (GRIP). The goal is to provide an efficient insurance policy for all area plans with uniform pricing.

It is important to note that livestock insurance, in comparison to various crop insurance programs, is considerably more complex and offers limited policy options to livestock producers. Unlike the long history of federally regulated crop insurance programs dating back to the 1930s, the first livestock pilot program, the Livestock Risk Protection (LRP) program, entered the market in 2002 (Glauber 2022). There was also the Livestock Gross Margin (LGM) program available in the same year that is often purchased by swine producers. The LRP protects livestock producers from declines in the price of the animals, while the LGM covers a decline in the expected margin (margin being the difference in animal price and the cost of feed). Furthermore, since the first authorization of livestock disaster programs in the 2008 Farm Bill, the Livestock Indemnity Program (LIP) and the Livestock Disaster Program (LFP) were reauthorized in the 2014 Farm Bill after their expiration in 2011, and a permanent baseline funding was provided (USDA FSA 2024). These programs offer financial assistance to livestock producers who suffer abnormally high mortality due to natural disasters.

In many cases, cattle producers choose to purchase options contracts to hedge against risks instead of paying a premium for an insurance plan. One reason for this preference is the low indemnity payments and subsidies received from insurance programs. Historically, RMA was limited to a $20 million budget per fiscal year on livestock insurance policies under the Agricultural Risk Protection Act of 2000 (Glauber 2022). This restriction was eliminated by the 2018 Bipartisan Budget Act. The removal of the expenditure cap on livestock insurance has led to a dramatic increase in participation, profoundly influencing the liability insured by livestock policies, which has surged from $500 million in 2018 to over $21 billion in 2022 (Fischer and Outlaw 2023). From a budgetary perspective, the rapid growth in livestock raises concerns among taxpayers that the cost of the program may exceed the Congressional Budget Office's (CBO) forecasted expenditure (Glauber 2022). This again emphasizes the importance of pricing insurance policies to ensure actuarial soundness.

This paper proposes the use of copula models to measure dependence structure in agricultural insurance policies in a multivariate setting. Copula models offer a flexible way to represent the complex dependence relationships often found in a high-dimensional setting. In particular, we apply the multivariate Gaussian copula, multivariate Student-t copula, and three types of Vine copula to demonstrate how the approaches adopted for assessing risks can have substantial effects on the pricing of insurance policies.

We include two empirical analyses to investigate systemic risk in both crop and livestock insurance policies. The first empirical application focuses on revenue insurance plans for

corn, the largest field crop in the U.S. We select the six largest corn-producing counties in Iowa to estimate loss probabilities and actuarially fair premium rates given the characterized dependence structures. We pay attention to how the pricing of portfolios varies depending on the copula model and coverage levels. Different coverage levels are considered since the degree of spatial correlations in corn yields is larger when extreme yield events happen (Goodwin 2001). Additionally, we analyze the effect of price volatility on insurance prices and find that the standard normality assumption underestimates the associated risks when market volatility is high, leading to a lower premium.

In the second empirical analysis, we turn our attention to livestock insurance plans for cattle producers. As mentioned previously, there are fewer policies designed for livestock insurance in the U.S. Despite the federal government's provision of comprehensive legislation to cover various losses and protect farmers and ranchers, there is currently no cattle insurance program that accounts for quality risk. In recent decades, there has been an increased demand for higher-quality beef in the market, which has prompted the adoption of a grid pricing system. Traditionally, fed cattle were priced either on a live or dressed weight basis. In contrast, the grid pricing system determines the value of an individual animal based on its carcass quality and yield grade associating to a premium or discount. Belasco et al. (2010) find significant trade-offs between yield and quality grades, bringing up one of the complications in addressing quality risk for insurance policies.

In this empirical study, we account for correlations between various yield and quality grades with copula models using a dataset comprising more than 8,000 individual feedlots in Kansas and Nebraska. Similar to the first empirical study, we compare the probability of loss and the actuarially fair premium rates for various livestock insurance portfolios at different coverage levels. In particular, we calculate the probability of losses and premium rates with and without the consideration of quality risk. Our results demonstrate the inclusion of quality risk significantly increases the loss probabilities and premium rates. In addition, the Gaussian copula proves effective in accurately characterizing the risks associated with cattle quality.

The rest of the paper is structured as follows. Section 1.2 discusses the theoretical framework and various copula models applied in the study. Sections 1.3 and 1.4 describe the two empirical studies, including descriptions of data sources and model-building procedures, and report the empirical findings of each study. We estimate portfolio risk by considering marginal distributions for individual components of a portfolio and the joint distribution defining the overall portfolio using Gaussian, Student-t, and Vine copulas. Lastly, Section 1.6 concludes the paper by exploring the effectiveness of copula models for

capturing dependent risk and calculating premiums in agriculture insurance schemes.

1.2 Theoretical Framework

The concept of copulas was first introduced by Sklar in 1959 in the context of probabilistic metric spaces. Let F be a n-dimensional multivariate joint distribution function with marginal distributions $F_i(x_i)$, the Sklar's Theorem (Sklar 1973) states that there exists a copula function $C : [0,1]^d \rightarrow [0,1]$ such that

$$F(x_1, x_2, ..., x_n) = C\{F_1(x_1), ..., F_n(x_n)\} \tag{1.1}$$

where $x_1, \ldots, x_n \in \mathbb{R}$ are random variables. Given the marginal distributions are continuous, the copula function C is unique and contains all dependence information in the joint distribution. Taking the inverse of Equation 1.1, we can rewrite the copula function as

$$C(u_1,\ u_2,\ ,...,\ u_n) = F(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)) \tag{1.2}$$

where $F_i^{-1}(u_i)$'s are the inverse marginal distribution functions, and $u_1, \ldots u_n \in [0,1]$. Intuitively, any multivariate joint distribution can be written in terms of its univariate marginal distribution functions and a copula function, where the dependence structure between random variables is described by the copula function.

For some n-variate copula density c, the joint density function f is

$$f(x_1, x_2, ..., x_n) = c(F_1(x_1), F_2(x_2), ..., F_n(x_n)) \cdot f_1(x_1) \cdots f_n(x_n) \tag{1.3}$$

The applications of copula-based models have attracted a lot of attention in recent decades, including their use in actuarial estimation of joint densities and portfolio risk (Goodwin and Hungerford 2014; Ahmed and Serra 2014; Feng and Hayes 2016). Copula-based models can estimate marginal distributions independently of modeling the dependence structure, providing a way to couple or link different marginal distributions of random variables together and specify a joint probability distribution function, providing greater flexibility (Nagler et al. 2017).

Bivariate copula models are well-investigated in the literature due to their simplicity and functionality (Trivedi and Zimmer 2006; Genest and Rivest 1989; Nelsen 2013). The models take on many forms, including parametric and nonparametric. In general, for the

bivariate case, the previously defined joint density function f as well as the copula density c can be simplified to

$$f(x_1, x_2) = c_{12}\{(F_1(x_1), F_2(x_2))\} f_1(x_1) f_2(x_2) \tag{1.4}$$

$$c(u_1, u_2) = f(F_1^{-1}(u_1), F_2^{-1}(u_2)). \tag{1.5}$$

Popular copula families are elliptical copula and Archimedean copula. Define Φ_ρ as the standard normal cumulative distribution function (CDF) with the correlation parameter $\rho \in (-1, 1)$, the bivariate Gaussian copula is defined as

$$c(u_1, u_2) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \tag{1.6}$$

Similarly, define $t_{\rho,\nu}$ as the student-t distribution with the correlation parameter ρ and $\nu > 2$ degrees of freedom, the bivariate student-t copula can be defined as

$$c_t(u_1, u_2) = t_{\rho,\nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) \tag{1.7}$$

Elliptical copulas such as Gaussian and Student-t are attractive in application because of the ease of implementation. They are heavily restrictive, however. Gaussian copulas cannot account for tail dependence and student-t copulas are not able to account for asymmetry independence. Clayton and Gumbel copulas are in the Archimedean copula family, which can model asymmetric dependence relations with heavy restrictions.

While multiple copula families are defined to accommodate various dependence structures between two variables, the estimation of multivariate density is still in the earlier stage of work. High-dimensional copula models are expected to provide more flexible and informative dependence relationships that are favored in real-life applications such as in the finance and actuarial industry. However, the transition from two dimensions to three dimensions (and even higher dimensions) is a huge leap forward. For example, nonparametric density estimators are favored in many applications, such as in the estimation of VaR, due to their flexibility. These estimators face the problem of the curse of dimensionality in higher dimensions: they converge more slowly to the true density as the dimension increases. In other words, accurate estimates require an excessively large number of observations as the dimension increases.

Standard copula models for higher dimension data such as the multivariate Gaussian and Student-t are easy to implement yet lack flexibility when trying to address problems

such as tail dependence and asymmetries. Nevertheless, several models are introduced to properly estimate high-dimensional copulas, including hierarchical Archimedean copula (HAC), vine copula, and factor copula. This paper focuses on the vine copula for its flexibility.

The vine copula was first proposed by Joe (1996) and further developed by Bedford and Cooke (2002) and Kurowicka and Cooke (2006). Recall Equation 1.3, the idea behind vine copulas is that any complicated multi-dimensional copula density c can be decomposed into the product of a series of bivariate copulas. A general d-dimensional copula can be decomposed into $d(d-1)/2$ bivariate copulas (Bedford and Cooke 2002). This idea was extended by Aas et al. (2009) to model multivariate dependence relationships using pair copulas. This procedure is also known as the pair-copula constructions (PCCs). The decomposition can then be organized as a graphical structure, which is what we refer to as the vine copula structure.

Since the building blocks are pair copulas, we are able to adopt the rich bivariate families in the construction of the complex multivariate dependence models. There are many different ways to factorize the multivariate density. Bedford and Cooke (2002) introduced the regular vine *(R-vine)* to help organize various constructions. Kurowicka and Cooke (2006) derived two special cases of regular vines: the canonical vine *(C-vine)* and the drawable vine *(D-vine)*, each provides a specific way of decomposing the density. The density of a four-dimensional R-vine structure can be expressed as

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = {} & f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4) \\ & \cdot c_{12}\{F_1(x_1),F_2(x_2)\}\cdot c_{23}\{F_1(x_2),F_3(x_3)\}\cdot c_{24}\{F_1(x_2),F_4(x_4)\} \\ & \cdot c_{13|2}\{F(x_1|x_2),F(x_3|x_2)\}\cdot c_{34|2}\{F(x_3|x_2),F(x_4|x_2)\} \\ & \cdot c_{14|23}\{F(x_1|x_2,x_3),F(x_4|x_2,x_3)\}, \end{aligned}$$

the C-vine structure

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = {} & f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4) \\ & \cdot c_{12}\{F_1(x_1),F_2(x_2)\}\cdot c_{13}\{F_1(x_1),F_3(x_3)\}\cdot c_{14}\{F_1(x_1),F_4(x_4)\} \\ & \cdot c_{23|1}\{F(x_2|x_1),F(x_3|x_1)\}\cdot c_{24|1}\{F(x_2|x_1),F(x_4|x_1)\} \\ & \cdot c_{34|12}\{F(x_3|x_1,x_2),F(x_4|x_1,x_2)\}, \end{aligned}$$

and the D-vine structure

$$
\begin{aligned}
f(x_1, x_2, x_3, x_4) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\
&\cdot c_{12}\{F_1(x_1), F_2(x_2)\} \cdot c_{23}\{F_2(x_2), F_3(x_3)\} \cdot c_{34}\{F_3(x_3), F_4(x_4)\} \\
&\cdot c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot c_{24|3}\{F(x_2|x_3), F(x_4|x_3)\} \\
&\cdot c_{14|23}\{F(x_1|x_2, x_3), F(x_4|x_2, x_3)\}.
\end{aligned}
$$

where $c_{24|3}$ is the joint density of the conditional vector $(F_{2|3}(X_2|X_3), F_{4|3}(X_4|X_3))|X_2$.

All three vine structures represent a high level of flexibility for high-dimensional modeling, each with its own characteristics and properties. A C-vine uses a canonical parameterization, where the parameters of the copulas are directly interpretable as Kendall's tau coefficients. This makes it easier to interpret the strength and direction of dependencies. However, since C-vine enforces a specific canonical order in the construction process, the ordering of variables can significantly impact the specification of the model. D-vine employs a parameterization based on the concept of dissimilarity, measuring the difference between observed and modeled copula values. This can be useful in capturing non-linear dependencies. On the other hand, a R-vine uses a more general parameterization, and the order in which vines are constructed may be chosen based on data characteristics or other considerations, providing more flexibility but sacrificing some interpretability. In empirical applications, we consider all three vine structures and determine the most appropriate one using information criteria.

Once the dependence structure is characterized using different copula models, we generate Monte Carlo samples from the joint distribution and estimate the probabilities of loss as well as actuarially fair premium rates. Another issue to address is the measures of marginal densities. It is an important step in the insurance rate-making procedure as marginal densities underline risks for a specific variable. For example, a precise measure of marginal density for yield describes yields that trigger losses. For both empirical studies, we select the optimal marginal distribution using maximum likelihood from a list of parametric densities including Weibull, gamma, and log-normal[6]. We consider only parametric specifications as they offer a wide range of shapes to provide flexibility and provide an explicit functional representation of the marginals.

In summary, the probability of loss and actuarially fair premium rates are estimated with procedures outlined as follows,

[6]Density plots and goodness-of-fit results for Weibull, gamma, and log-normal are included. However, our selection includes more candidates such as beta, inverse Gaussian, and inverse Weibull.

Step 1: Specify the individual marginal distributions of variables.

Step 2: Rank transform to normalize variables to a uniform distribution, then fit the data points to copula models.

Step 3: Simulate 10,000 draws from the fitted joint distributions for each variable.

Step 4: Transform the generated data using the inverse margins estimated in Step 1.

Step 5: Calculate probabilities of loss and actuarially fair premium rate. The probability of loss is defined as the number of times that indemnities are positive out of the 10,000 simulated data points, and the actuarially fair premium rate is defined as the ratio between expected loss and liability.

1.3 Empirical Application I: Crop Revenue Protection Insurance

Revenue Protection (RP) is currently the most popular insurance product on the market. These policies protect against the loss of revenue due to yield loss, changes in price, or a combination of both. In this empirical study, we analyze several synthetic revenue protection policies for six adjacent counties in Northwest Iowa: Cherokee, Ida, Plymouth, Sac, Sioux, and Woodbury. These counties were selected because they were among the most productive in Iowa for corn in 2019. As the largest corn-producing state in the U.S., Iowa's corn production was estimated at 2.58 billion bushels in 2019, according to the USDA National Agricultural Statistics Service (NASS). Therefore, analyzing such a state can provide representative results that can be generalized to other states and crops in the Midwest.

Considering the spatial correlation between county yields, and recognizing the natural correlation between prices and yields, the joint distribution between county yields and prices becomes a key factor in pricing a revenue insurance policy and determining liability. Furthermore, as we examine corn revenue protection plans for adjacent counties, we present a comparable representation of the Area Risk Protection Insurance (ARPI) contracts offered by the RMA.

Annual production and acres planted data at the county level from 1960 to 2019 are retrieved from the USDA-NASS. The yield data, measured in bushels produced per acre, is calculated as the ratio between total production and acres planted. To model the price risk embedded in revenue insurance, we follow Goodwin and Hungerford (2014) and Goodwin et al. (2018) and use the logarithm of the ratio of the October and February corn futures prices, which can be viewed as the return over the growing season. The monthly averages

of the daily settlement prices in February and October for the December corn futures contract. The price data are obtained from the Chicago Board of Trade (CBOT) from 1960 to 2019. February and October prices are selected since these are the months when insurance policies are sold and when harvest prices are realized.

For crop revenue insurance policies, a loss can be described as

$$Loss_{revenue} = max(0, (\lambda \bar{y} p_p - y^* p_r)) \tag{1.8}$$

where $\lambda \in (0,\ 1)$ is the coverage level, $\bar{y}$ is the expected yield determined using the average of the actual production history (APH) at the time when purchasing the crop insurance policy, P_p is the projected price that is predetermined, y^* is the realized yield, and P_r is the realized price at the time of harvest. Therefore, an indemnity payment occurs when the realized revenue, $y^* p_r$, is less than the revenue guarantee, $\lambda \bar{y} p_p$. As shown by the defined loss in Equation 1.8, revenue insurance considers two sources of risk as both the harvested price and realized yield are unknown until the time of the harvest, making them behave as random variables.

The coverage level is another important component when pricing insurance policies. The current revenue protection plans offered by USDA-RMA have eight coverage options ranging from 50% to 85% in 5% increments. Producers can also choose to include the Enhanced Coverage Option (ECO) in addition to the revenue protection, which offers a choice of 90 or 95% coverage level. For our simulation studies, we consider four coverage levels – 65%, 75%, 85%, and 95%. Moreover, individual counties' exposures to risk, or the levels of acreage, may affect the pricing of insurance policies under a group risk insurance plan. Therefore, in addition to the analysis of each county, we include two synthetic portfolios adjusting the weights of each county. The first portfolio weights each county equally, and the second portfolio weights each county based on the 2019 planted acreage.

The goal of this study is to compare the loss probabilities and the actuarially fair premium rates for policies simulated under different dependence structures characterized by various copula models, given the coverage level. For an insurance policy to be actuarially fair, its premium should be set to a rate that is equal to the expected loss. Mathematically, it should be represented as,

$$\begin{aligned} Premium &= Expected\ Losses \\ &= P(Loss)E(Loss|Loss > 0) \end{aligned} \tag{1.9}$$

where the loss is defined in Equation 1.8, and the probability of loss ($P(Loss)$) is the number of times a loss occurs out of the total simulations. We calculate the actuarially fair rate by taking the ratio between the expected losses and liability (the guaranteed revenue). This calculation provides information on how much premium should be collected per dollar of liability provided.

1.3.1 Empirical Results

There is an upward trend expected in corn yields over time due to technological improvement and more advanced farming practices (Xu et al. 2013). Therefore, the first step to model yield involves detrending the data. The trend of each county's corn yield is accounted for by fitting a nonparametric locally weighted scatter plot smoothing (LOESS) line to each county's yields. Since LOESS does not require any assumptions about the form of the functional relationship, it provides a more flexible fit compared to linear regression (Goodwin et al. 2018). Specifically, we fit the yield data to the following regression,

$$y_t = g(t) + e_t \tag{1.10}$$

where $g(t)$ is generated using nonparametric methods. We then centered on the 2019 prediction for each county by adding back their residuals to generate the detrended yield[7].

$$\tilde{y} = \hat{y}_{2019} + e_t \tag{1.11}$$

where $\hat{y}_{2019}$ is the fitted value for 2019, e_t is the residual at time t, and $\tilde{y}$ is the recentered and normalized detrended yield.

The fitted local regression line is presented in Figure 1.1. We overlay the LOESS fit on a scatter plot of yield data for six counties, clearly showing that the LOESS fit captures a positive trend in all counties' corn yields. Additionally, Figure 1.2 presents the empirical distributions for detrended yields and the price in the diagonal boxes. The scatterplots between the paired yields and the price, along with the corresponding Kendall's tau correlations, are also included. We observe a high positive correlation between the detrended yields. Given that the selected counties are located near each other, this confirms the geographical correlations of crop yields. Regarding the price, we consider the logarithm of the ratio of the October and February prices for the December corn futures contract. As

[7]This is a simple and naive yet common way to normalize yields. This method assumes the residuals to be homoskedastic. The residuals, however, are more likely to be heteroskedastic in real life.

expected, the price is negatively correlated with county yields.

To derive the probability of loss and premium rates from the simulated random yields and prices from estimated copula models, we need to specify the individual marginal distributions of yields and prices. It should be noted that the selection of the marginal distribution of each variable is not related to the fitting of copula models. Marginal distribution becomes particularly useful when transforming the simulated data back to their empirical distribution, as the fitting of a copula requires all variables to be transformed into a uniform distribution. The probability densities and cumulative densities are plotted in Figure 1.3, and the goodness-of-fit results are presented in Table 1.1[8]. Weibull is the optimal density for county yields according to the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and loglikelihood. The selection of Weibull is common in the literature as it accommodates the negative skewness often observed in crop yield (Goodwin and Hungerford 2014). For the price, we adopt the common assumption that prices are log-normal; therefore, a normal distribution is assumed for the log of price ratios.

After specifying the densities for detrended yield and price, the next step is to estimate the copula models. We consider Guassian, Student-t, and Vine copulas. As discussed, Vine copula models take advantage of the well-developed bivariate copula families and provide a flexible way to model high-dimensional dependence. We consider three Vine copula structures, the R-vine, C-vine, and D-vine, and select the best-performing model using AIC, BIC, and loglikelihood. The goodness-of-fit statistics for the fitted copula models are presented in Table 1.2. We observe that the R-vine structure has the lowest AIC and BIC, although its loglikelihood is slightly lower than the C-vine structure. We choose to use R-vine to compare with the Gaussian and Student-t copulas in the insurance pricing analysis. It should also be noted that all three vine structures provide a better fit than the Gaussian and Student-t copula models, proving in general, that a more flexible model is preferred over restricted models.

Copula estimations for the Gaussian, Student-t, and R-vine models are presented in Table 1.3 and Table 1.4. We observe the parameter estimates of the Gaussian and Student-t copula are very similar. In addition, the estimated degrees of freedom of the Student-t copula is 4.439, which reflects the presence of tail dependence. These estimates reflect the expected correlations between county yields and prices, with positive estimates between two county yields (for example, ρ_{12}) and negative estimates between yields and prices (for example, ρ_{17}). Table 1.4 presents the coefficient estimates for the R-vine copula, where the

[8]The results and plots only present for Weibull, gamma, and log-normal, but the selection was searched over a set of marginal distributions including normal, exponential, and student-t.

negative correlation between yields and prices is evident.

With knowledge of marginal distributions, we simulate 10,000 data points from the joint distributions fitted through Gaussian, Student-t, and R-vine copula models. Simulated variables are then transformed back to their empirical marginal distributions based on the Weibull estimates presented in Table 1.1 to calculate loss probabilities and premium rates. Annually, the RMA Price Discovery Tool announces the projected price by averaging February contracts and determines the harvest price based on the average of October contracts. Since the February price is known at the time of purchasing the insurance policies, we use the 2020 RMA's projected February corn futures price for Iowa of $3.88 per bushel corn in the simulation. Because we use the log of the price ratio in the estimation of dependence relationships, it is necessary to invert the simulated price ratios to solve for the October prices. As mentioned, we make the common assumption that price follows a log-normal distribution, therefore, the log of the price ratio is assumed to follow a normal distribution with a volatility of 0.18[9]. Specifically, we use the following expression to calculate the October prices based on the assumed February P_{feb} and the volatility σ,

$$P_{oct} = P_{feb} \times e^{-\frac{\sigma^2}{2} + \sigma P_{sim}} \tag{1.12}$$

where $P_{sim} \sim N(0,1)$ is the simulated price ratio.

The loss probabilities and corresponding premium rates are presented in Table 1.5. As expected, the probabilities of loss and premiums increase as the coverage level increases, although the increase is not linear. When comparing the premium rates of individual counties to the synthetic portfolios, we observe that most of the time, portfolios that include all counties have lower premium rates than individual policies. This holds across all coverage levels and all copula models. This suggests that the risk associated with an insurance contract is lower when the risks are aggregated. Moreover, when comparing the two portfolios, we do not observe significant differences in loss probabilities and premium rates. This is expected since all six counties are among the most productive counties in Iowa, and the differences in their levels of acreage do not have a significant impact on the pricing of insurance policies.

Possibly a more interesting finding is the differences in rates at the 65% coverage level. Compared to premium rates implied by the Gaussian copula, both R-vine and Student-t copulas imply higher rates for all counties, with the Student-t copula's implied rates being higher than those implied by the R-vine copula. The rate differences become smaller as the

[9]This volatility is reported by RMA's Price Discovery Tool for 2020 December Corn futures contracts.

coverage level increases. As both R-vine and Student-t copula structures account for tail dependence, this result suggests that the Gaussian copula model can underprice insurance policies in the presence of extreme events. Considering the difference between the R-vine and the Student-t copula, this may suggest that the Student-t copula can potentially overprice insurance policies.

Price volatility serves as a metric for the market expectation of price uncertainty, making it a crucial factor to consider when the RMA sets premiums. To access how price volatility influences the pricing of revenue insurance contracts, we conduct the same analysis with double the previous price volatility (0.36), and the results are presented in Table 1.6. Similar to what we observed in the case of lower price volatility, premium rates are lower when risks are aggregated in groups, as shown by the two synthetic portfolios.

Doubling the price volatility leads to a significant increase in loss probabilities for insurance at all coverage levels, although the magnitudes vary. For instance, at the 65% coverage level, the Gaussian copula implies a loss probability of 0.12% for Cherokee County. This probability increases by 74 times to 8.91% in the high price volatility case. At the 95% coverage level, the loss probability rises from 38.74% to 48.53%, a 1.25 times increase. This implies that the escalation in price volatility impacts policies with lower coverage levels more than those with higher coverage levels. It is worth noting that while the rate difference is smaller at the 65% coverage level between copula models, the R-vine copula consistently implies higher premium rates than the Gaussian and Student-t copulas at higher coverage levels. This suggests that during periods of increased market turbulence, both Gaussian and Student-t copulas have the potential to underprice premiums.

Overall, in situations of relatively low price volatility, the probabilities of loss and premium rates suggested by the Gaussian copula are close to those proposed by the R-vine copula. However, the Gaussian copula's assumption of zero tail dependence fails to capture the state-dependent risk observed in agricultural yields during extreme events, such as floods or droughts (Goodwin and Hungerford 2014). When facing higher price volatility, both Gaussian and Student-t copulas indicate lower premium rates compared to the R-vine copula. These findings suggest that the accuracy of the Gaussian copula may be influenced by market conditions, emphasizing the preference for models that account for tail dependence in scenarios of increased price uncertainties.

1.4 Empirical Application II: Livestock Risk Protection Plan

The application of copula models can have great impacts in the context of livestock insurance in the U.S. In this empirical study, we focus on the market of fed cattle, although the applied procedure can easily be generalized to other livestock such as swine. Commercial cattle production comprises three phases: the production phase of weaned feeder calves, the backgrounding phase of production of feeder-lot-ready calves, and the finishing phase, during which cattle are fattened for slaughter. We concentrate on the finishing phase, where cattle are sent to a dry lot on full feed with grains for fattening. We adopt a similar analysis procedure as in the previous application, using individual lot data instead of aggregated data for a county. Uncertainties may arise during the feeding period, which typically lasts three to five months.

An important source of risk in the Fed cattle industry is the pricing mechanism. There are three common methods for pricing fed cattle in the market: by live weight, by dressed weight, and based on a grid price. With live weight and dressed weight pricing, typically, one average price is established for the entire lot. The advantage of having an average price is its flexibility, as prices are negotiated after transaction costs are established. The downside, however, is that these methods do not provide appropriate price signals to cattle producers (Hogan et al. 2009). On the other hand, the grid pricing system does not have this concern, as the price is determined based on a single animal.

The idea of using grid pricing for the fed cattle market was first introduced in the 1990s. The Value-Based Marketing Task Force in 1990 (VBMTF) highlighted inconsistent quality issues in the beef market, leading to an innovation in the pricing mechanism. As noted in the 2016 National Beef Quality Audit (NBQA) conducted by Beef Quality Assurance, consumer satisfaction ranked as the second priority for all marketing sectors except for packers (Boykin et al. 2017). A greater share of companies expressed a willingness to pay a premium for guaranteed quality attributes. The audit also mentioned that, according to many survey respondents, a consistent offering in size is more important than an increase in size. A large carcass makes it harder to meet consumer-specific thickness and weights. Consequently, about 66 percent of further processors are willing to pay a premium in exchange for guaranteed weight and size (Boykin et al. 2017).

A typical grid consists of five yield grades and five quality grades, with each grade linked to a premium or a discount added to a base price. The grid also includes other discount indicators such as carcass weights and dark cutter[10]. While specific grids can vary by firm

[10]Dark cutting beef is a result of pre-harvest stress. When animals undergo long-term stress before slaugh-

and packing locations, they share common features Feuz (1999). Table 1.7 demonstrates an example grid pricing mechanism, which is calculated from weekly forecasted prices for the associated premiums and discounts. This example grid consists of five quality grades and five yield grades to reflect the animal's carcass attributes. Quality grade choice and yield grade 3 indicate the base price, reflecting the value of a standardized carcass. As prices are established based on the quality and yield grades of the beef, with higher-quality beef rewarded with a premium and lower-quality beef penalized with a discount, grid pricing places a greater emphasis on the quality and consistency of beef. This, in turn, encourages producers to improve the quality and produce more market-desirable types of cattle.

The adoption of grid pricing can significantly impact livestock insurance for cattle producers. In a grid pricing system, premiums and discounts are linked to the characteristics of the beef, directly affecting its market value. If the pricing grid consistently results in higher prices for certain quality grades of beef, cattle producers may need to adjust their insurance coverage levels to ensure adequate protection. In addition, premiums and discounts change weekly in response to the supply of fed cattle and consumer demand. The unrealized premiums and discounts introduce greater variability in underwriting. Fausti et al. (2014) compare the risk of a producer's decision to sell cattle on a grid relative to selling cattle at a pen-average price. They find evidence that the risk premium associated with grid pricing shocks can influence producers' market decisions.

A federal safety net that considers the quality risk embedded in the grid pricing system can encourage greater adoption of this pricing mechanism. However, the current cattle insurance programs provided by the RMA are not designed to incorporate the grid pricing system, where the potential risks associated with such a pricing system can significantly impact pricing strategies. In the U.S., one of the most popular livestock insurance programs provided by the RMA is the Livestock Risk Protection (LRP) plan, which solely protects producers from price declines without incorporating the quality aspect of beef. Motivated by this concern, the goal of this empirical study is to assess how the adoption of a grid pricing mechanism would affect the risk portfolio of cattle LRP insurance. To our knowledge, this is the first study that incorporates the quality risk perspective into livestock insurance pricing.

The motivation behind employing copula models is the correlation among the five

ter—whether due to extended transportation, changes in temperature, or rough handling—their muscle glycogen reserves are depleted prior to slaughter. This depletion results in beef with a deep red or purple color compared to the bright red hue that consumers expect. It is firm and dry in texture, which is unacceptable to consumers

yield grades, five quality grades, and three discount factors (as shown in Table 1.7). Copula models offer a suitable approach to address the high degree of dimensionality in this complex scenario. Similar to the first empirical study, we calculate the loss probabilities and actuarially fair premium rates, comparing them across various copula models, coverage levels, and risk aspects.

1.4.1 Data and Empirical Methods

We utilize the data obtained from five feedlots in Kansas and Nebraska that use grid pricing as their main marketing strategy. This dataset comprises a total of 8,569 observations, with an average of 129.2 heads per pen. When feeder cattle enter the lot, characteristics such as in-weight and sex are recorded. At the end of the feeding period, cattle are slaughtered and categorized into five yield grades (y1, y2, y3, y4, y5), five quality grades (Prime, CAB[11], Choice, Select, Standard), and three discount factors (dark cutters, heavy, light). The quality and yield grades and dark cutters are measured as a percentage of the total carcass weight. The heavy and light discounts are measured as the percentage of the total number of carcasses in a pen that exceeds 950 lbs. or is below 550 lbs., respectively. Carcass beef grades and standards are determined by the USDA Agricultural Marketing Service (AMS). Quality grades, as described by AMS, consider the marbling and tenderness of beef, while yield grades are determined by the thickness of the layer of external fat. YG1 refers to carcasses with only a thin layer of fat and, therefore, yields the most edible lean meat (USDA AMS 2024).

The descriptive statistics of each grid and a set of exogenous pen characteristics are reported in Table 1.8. Notice that most of the grid pricing variables are left censored and few of them are almost fully censored. For example, the degree of censoring of yield grade five is more than 80 percent; in addition, all of the discount factors are censored for more than half of the observations. Figure 1.4 contains histograms for yield and quality grades and the discount factors. We can visually see Prime, CAB, Standard, YG1, YG4, YG5, and all of the discount factors are left censored at zero.

The quality and yield of each pen can be characterized by a set of pen characteristics.

[11] CAB is not a formal USDA beef quality grade, but it requires similar marbling requirements as the upper two-third choice defined by the USDA.

Following Belasco et al. (2010), we define the following system of equations,

$$\begin{bmatrix} Q_i \\ Y_i \\ D_i \end{bmatrix} = x_{1i} * \beta + \epsilon_i, \tag{1.13}$$

where Q_i is the five quality grades: Prime, CAB, Choice, and Select; and Y_i denotes the five yields (YG1-YG5). Lastly, D_i represents the three discount factors: dark cutters, heavy, and light. Each dependent variable is conditioned on the same set of independent variables, x_{1i}, which includes the log of average placement weight, the sex of each pen (steer, heifer, or mixed), the location of each pen (Kansas or Nebraska), and the time of placement (Spring, Summer, Fall, or Winter). For dependent variables that are censored, we propose to use a heteroscedastic Tobit model to account for censoring, with underlying latent variable following Student-t distribution. The heteroscedastic Tobit model no longer assumes a constant variance, but rather is conditioned on the set of explanatory variables.

Belasco et al. (2009) found that the residuals in the above equations are likely to be correlated. These correlations typically arise from unobserved variables such as weather and genetics. While they proposed a multivariate Tobit model to address cross-equation correlation, this method is restricted to having only one censored dependent variable. Instead, we fit the residuals to various copulas models to control for possible correlations. The steps of fitting copula models are similar to those in the first empirical study. We first rank-transform each residual vector into uniform distributions, and then the copula models are fitted. After defining the joint distribution, we simulate data points from each copula structure to preserve rank correlation. Lastly, we inversely transform the simulated data back to their empirical marginal distributions using parameter estimates for residual marginals presented in Table 1.9. We then calculate the probability of loss and actuarially fair premium rates, accounting for different risk portfolios.

To assess the correlation across errors in each equation, we calculate Kendall's rank correlation of the residuals. The correlation estimates are presented in Table 1.11 and Table 1.12. These results confirm cross-equation correlations, validating the use of copula models to capture dependence relationships. We observe a negative correlation between larger carcass yield grades and higher beef quality grades. For example, the correlation between YG1 and Prime, CAB, and Choice grades is negative, while it becomes positive with Select and Standard grades. This trade-off between larger carcass yields and higher beef quality holds in almost all cases. These findings are consistent with the ones described

in Belasco et al. (2010).

With the confirmation of correlations among residuals, we then fit the copula models to characterize the dependence relationships. The goodness-of-fit statistics of each copula model are presented in Table 1.10. The R-vine copula has the lowest AIC and BIC, and the highest log-likelihood among the three vine copula structures; therefore, it is selected for comparison with Gaussian and Student-t copula in the pricing analysis. Additionally, all the vine copula models suggest a superior fit than the Gaussian and Student-t copula models.

1.4.2 Empirical Results

The calculation of LRP producer premium and indemnity is straightforward using the following formulas, which are provided by USDA AMS. Note that there is a subsidy factor included in the calculation. USDA provides subsidies to motivate the purchase of insurance. The subsidy factor is usually set at 20 percent.

Producer Premium Calculation

Step 1: Insured Value = Number of Head × *Target Weight* × *Coverage Price*

Step 2: Total Premium = Insured Value × *Rate*

Step 3: Producer Premium = Total Premium - (Total Premium × *Subsidy Factor)*

Producer Indemnity Calculation

Indemnity = Number of Head × *Target Weight* × *(Coverage Price - Actual Ending Price)*

Carcass weight only affects the monetary amount of liability guaranteed and indemnity payments when a loss occurs; it does not impact the loss probability and actuarially fair premium rates, which are the focus of this study. Therefore, we proceed with the simulation study without considering the target weight. We simulated 10,000 data points for each copula structure corresponding to each quality and yield grade and the three discount factors. The simulated variables are then inversely transformed back to their empirical marginal distributions. The simulated censored variables are truncated at zero to preserve the censoring aspects. The price associated with the LRP insurance is discovered from the cattle futures market. Therefore, the cattle price used in the simulation study is based on the daily live cattle futures from 1/1/2009 to 12/31/2018 reported by the Chicago Mercantile Exchange (CME).

To incorporate quality risk into the pricing of LRP insurance, we obtained the National Weekly Direct Slaughter Cattle Premiums and Discounts report (LM-CT155) from 2001 to 2021 retrieved from USDA AMS. This dataset provides information on how premiums and

discounts change over time due to market conditions. Generally, premiums and discounts associated with yield grades and the three discount factors are more static and predictable than those associated with quality grades. We estimate the probabilistic characteristics of these premiums and discounts by fitting each one to an ARIMA model. Each premium and discount price series is first deseasonalized before fitting the model. We then forecast data for 20 weeks to obtain the expected premiums and discounts for each grade. The mean forecasted premiums and discounts are presented in Table 1.13. These estimates are used to calculate the example grid shown in Table 1.7.

We then calculate the probability of loss and the actuarially fair premium rates for a hypothetical mixed pen in Kansas. The probability of loss is defined as the number of times, out of the 10,000 simulated data points, that the actual ending price is less than the coverage price. The guaranteed coverage price is calculated by multiplying the expected end value by the coverage level and the number of head. The number of head per pen is set to 129, the average head per pen from the dataset. We set the expected end value to $122, which is the expected end value for the 17-week endorsement contract in Kansas reported by USDA AMS. When purchasing LRP policies, cattle producers may choose a coverage plan ranging from 70 to 100 percent. Therefore, we consider four coverage levels in our calculation: 75%, 85%, 95%, and 100%. Lastly, we define the actuarially fair premium rates as the ratio of expected loss and coverage price.

Recall that the quality and yield grades, as well as dark cutters, represent percentages of carcass weight falling into each grade category. Therefore, we can obtain the expected premiums or discounts associated with each simulated data point by multiplying them by the expected value of each premium/discount variable obtained from the forecast. These premiums and discounts are then added to the simulated live cattle prices. In addition, 'heavy' and 'light' refers to the percentage of cattle in each pen that exceeds 965 lbs. or is below 550 lbs. These two discounts are incorporated back into the realized pen-level revenue after determining the number of heads that would incur losses.

The calculated loss probabilities and premium rates are presented in Table 1.14. The benchmark model only compares the simulated live cattle price to the expected end value, which is the current practice of the LRP insurance program. Compared to the benchmark model, we observe a substantial increase in the probability of loss and actuarially fair premium rates when including quality risks. This is expected since losses can be incurred by multiple sources of risks under this insurance contract. A more meaningful result is the magnitude of the increase. Specifically, the average increase in probabilities of loss across four coverage levels is 76%, while the increase in premium rates is 126%. To put

this into perspective, the total premium collected in Kansas in 2023 for the LRP plans is $27,840,426 (RMA, Summary of Business Portal). An average increase in the premium rate of 126% implies a potential difference of $35 million per year. Lastly, we find no difference when comparing the calculation of loss probabilities and premium rates across different copula models. This suggests that no significant tail dependence and skewness were found among the dependence relations of the grid pricing variables, indicating that the standard Gaussian assumption provides a good representation in the case of cattle quality risks.

1.5 Conclusion

Given the significant government involvement in agricultural insurance programs and the related public policy issues regarding the amount of subsidy, an essential aspect is to quantify the systemic risks associated with the insurance policies. This, in turn, requires a quantitative representation of the risks associated with portfolios of insurance policies that differ by insurance plan, crop, and location.

In this study, we aim to characterize the systemic risks implied in agricultural insurance using Gaussian, Student-t, and three Vine copulas. The Gaussian assumption is currently applied in the existing insurance rating practice that does not account for tail dependence, while the Student-t copula allows for symmetric tail dependence. Vine copula models utilize pair-copula construction techniques to model different sources of risks. These models offer a great approach to characterizing the dependence structure for multivariate variables. We consider two empirical analyses, each involving the calculation of loss probabilities and actuarially fair premium rates based on different copula models.

The first analysis assesses corn revenue insurance for six counties in Iowa. We observe that Gaussian copulas tend to underprice insurance premiums, especially when price volatility is high. The second analysis is driven by a market shift from pen-level transactions to sales based on individual animal quality characteristics in the cattle industry. The adoption of the grid pricing system introduces more risk factors into the production process, leading to increased variability in producer revenues. Given the correlation among grid pricing variables, accurate mathematical representations of these risks can benefit both cattle producers and private insurance providers. Our results show that the inclusion of quality risk significantly raises the premium rates of a standard LRP plan. Moreover, we find that the Gaussian copula effectively captures cattle quality risks, with no significant differences observed across copula models.

These analyses emphasize the importance of determining dependence structures within agricultural insurance portfolios, which provide important implications for the federal government, private insurance providers, and agricultural producers. Although the analysis of corn revenue insurance contracts focuses on county-level data, it can easily be generalized to farm-level insurance contracts. The federal government offers the Whole-Farm Revenue Protection (WFRP) insurance plan, which serves as a safety net for all commodities on the farm under a single insurance policy, covering both crops and livestock. Given the high correlation between the yield and implied revenue of these commodities, precise estimation of the dependence structure among random variables holds crucial implications for revenue insurance pricing and serves as a guiding analysis for insurers and reinsurers.

1.6 Tables and Figures

Table 1.1: Goodness-of-Fit for County Yield Marginal Distributions

County	Weibull			Gamma			Log-Normal			Optimal Density	Weibull Estimates	
	AIC	BIC	Lglk	AIC	BIC	Lglk	AIC	BIC	Lglk		Shape	Scale
Cherokee	489.18	493.37	-242.59	501.70	505.89	-248.85	503.35	507.54	-249.67	Weibull	17.89	213.31
											(1.85)	(1.62)
Ida	501.75	505.94	-248.87	516.87	521.06	-256.43	518.93	523.11	-257.46	Weibull	16.28	215.84
											(1.66)	(1.79)
Plymouth	519.10	523.29	-257.55	532.87	537.06	-264.44	535.81	540.00	-265.91	Weibull	12.87	199.52
											(1.33)	(2.10)
Sac	508.46	512.64	-252.23	525.90	530.08	-260.95	528.43	532.62	-262.22	Weibull	15.08	208.54
											(1.59)	(1.87)
Sioux	497.21	501.40	-246.61	509.69	513.88	-252.85	511.55	515.74	-253.78	Weibull	13.76	174.46
											(1.47)	(1.72)
Woodbury	504.57	508.76	-250.29	509.99	514.18	-253.00	510.99	515.18	-253.50	Weibull	14.89	207.21
											(1.52)	(1.89)

Distributions are fitted to detrended yields of each county, centered around the yield of 2019.
Parentheses contain standard errors.

Table 1.2: Empirical Study 1 – Goodness-of-Fit for Copula Models

	Gaussian	Student-t	R-Vine	C-Vine	D-Vine
AIC	-384.05	-410.75	-449.04	-448.92	-428.94
BIC	-340.07	-364.68	-396.69	-394.47	-376.58
Lglk	213.02	227.38	249.52	250.46	239.47

Table 1.3: Gaussian and Student-t Copula Estimations

	Gaussian Copula		Student-t Copula	
Parameters	Estimates	Standard Errors	Estimates	Standard Errors
ρ_{12}	0.914	0.036	0.905	0.041
ρ_{13}	0.839	0.066	0.831	0.080
ρ_{14}	0.841	0.061	0.827	0.072
ρ_{15}	0.792	0.091	0.790	0.099
ρ_{16}	0.844	0.065	0.824	0.089
ρ_{17}	-0.363	0.130	-0.375	0.134
ρ_{23}	0.840	0.040	0.862	0.037
ρ_{24}	0.816	0.079	0.820	0.079
ρ_{25}	0.772	0.074	0.755	0.078
ρ_{26}	0.875	0.046	0.894	0.042
ρ_{27}	-0.361	0.161	-0.368	0.162
ρ_{34}	0.736	0.071	0.722	0.088
ρ_{35}	0.790	0.084	0.808	0.110
ρ_{36}	0.871	0.047	0.878	0.044
ρ_{37}	-0.375	0.150	-0.376	0.153
ρ_{45}	0.720	0.140	0.717	0.141
ρ_{46}	0.801	0.068	0.803	0.068
ρ_{47}	-0.270	0.258	-0.272	0.268
ρ_{56}	0.745	0.112	0.745	0.144
ρ_{57}	-0.315	0.203	-0.321	0.250
ρ_{67}	-0.413	0.137	-0.371	0.181

The degrees of freedom for the Student-t copula model is 4.439.
1 = Cherokee, 2 = Ida, 3 = Plymouth, 4 = Sac, 5 = Sioux, 6 = Woodbury, 7 = Price.

Table 1.4: R-vine Copula Estimation Results[1]

	Factorization	Copula Family	Parameter 1	Parameter 2	τ
Tree 1	1, 4	R_{180} Gumbel	2.82 (0.30)	-	0.65
	1, 7	R_{270} Gumbel	-1.31 (0.14)	-	-0.24
	2, 1	R_{180} BB7	3.61 (0.60)	3.02 (0.76)	0.71
	2, 5	Frank	7.34 (1.12)	-	0.58
	6, 2	Frank	11.98 (1.63)	-	0.71
	6, 3	R_{180} Gumbel	3.07 (0.33)	-	0.67
Tree 2	2, 4 | 1	Gaussian	0.24 (0.12)	-	0.16
	2, 7 | 1	R_{270} Tawn Type2	-4.14 (1.45)	0.09 (0.02)	-0.08
	5, 1 | 2	Joe	1.43 (0.18)	-	0.2
	6, 5 | 2	R_{180} Gumbel	1.14 (0.11)	-	0.12
	3, 2 | 6	Student t	0.37 (0.14)	2.17 (0.80)	0.24
Tree 3	5, 4 | 2, 1	Independent	-	-	0.00
	5, 7 | 2, 1	Independent	-	-	0.00
	6, 1 | 5, 2	Independent	-	-	0.00
	3, 5 | 6, 2	Student t	0.25 (0.15)	2.16 (1.14)	0.16
Tree 4	6, 4 | 5, 2, 1	R_{180} Clayton	0.34 (0.19)	-	0.15
	6, 7 | 5, 2, 1	R_{90} Tawn Type1	-20.00 (13.14)	0.01 (0.001)	-0.01
	3, 1 | 6, 5, 2	Clayton	0.28 (0.19)	-	0.12
Tree 5	3, 4 | 6, 5, 2, 1	R_{90} Clayton	-0.37 (0.23)	-	-0.16
	3, 7 | 6, 5, 2, 1	R_{270} Tawn Type2	-5.03 (3.56)	0.03 (0.02)	-0.03
Tree 6	7, 4 | 3, 6, 5, 2, 1	R_{90} Tawn Type2	-8.85 (3.93)	0.08 (0.01)	0.08

τ: Empirical Kendall's tau.
Parentheses contain standard errors.
1 = Cherokee, 2 = Ida, 3 = Plymouth, 4 = Sac, 5 = Sioux, 6 = Woodbury, 7 = Price.

Table 1.5: Revenue Insurance (Price Vol = 0.18): Probability of Loss and Premium Rates

	Gaussian Copula		Student-t Copula		R-Vine Copula	
	Pr(Loss)	Rate	Pr(Loss)	Rate	Pr(Loss)	Rate
65% Coverage Level						
Cherokee	0.0012	0.00005	0.0042	0.00029	0.0019	0.00010
Ida	0.0021	0.00007	0.0047	0.00036	0.0014	0.00008
Plymouth	0.0022	0.00010	0.0065	0.00056	0.0029	0.00018
Sac	0.0031	0.00015	0.006	0.00053	0.0031	0.00020
Sioux	0.0029	0.00016	0.0074	0.00062	0.0043	0.00030
Woodbury	0.0012	0.00004	0.0049	0.00045	0.0011	0.00005
Portfolio1	0.0016	0.00005	0.0044	0.00032	0.0007	0.00006
Portfolio2	0.0016	0.00005	0.0046	0.00033	0.0006	0.00006
75% Coverage Level						
Cherokee	0.0248	0.0012	0.0282	0.0019	0.0239	0.0013
Ida	0.0257	0.0013	0.0305	0.0022	0.0196	0.0009
Plymouth	0.0314	0.0017	0.0359	0.0028	0.0267	0.0016
Sac	0.0362	0.0021	0.0369	0.0027	0.0282	0.0017
Sioux	0.0326	0.0019	0.0378	0.0030	0.0364	0.0022
Woodbury	0.0247	0.0012	0.0300	0.0022	0.0231	0.0010
Portfolio1	0.0237	0.0012	0.0291	0.0020	0.0191	0.0009
Portfolio2	0.0245	0.0012	0.0291	0.0020	0.0195	0.0009
85% Coverage Level						
Cherokee	0.1404	0.0094	0.1402	0.0105	0.1382	0.0093
Ida	0.1454	0.0098	0.1415	0.0109	0.1461	0.0086
Plymouth	0.1555	0.0111	0.1489	0.0121	0.1606	0.0107
Sac	0.1578	0.0120	0.1518	0.0124	0.1454	0.0103
Sioux	0.1545	0.0116	0.1556	0.0127	0.1596	0.0122
Woodbury	0.1368	0.0093	0.1452	0.0111	0.1573	0.0098
Portfolio1	0.1422	0.0095	0.1434	0.0107	0.1397	0.0086
Portfolio2	0.1430	0.0096	0.1445	0.0108	0.1411	0.0087
95% Coverage Level						
Cherokee	0.3874	0.0354	0.3881	0.0364	0.3911	0.0352
Ida	0.3892	0.0359	0.3887	0.0370	0.3992	0.0357
Plymouth	0.3911	0.0377	0.3948	0.0386	0.4133	0.0392
Sac	0.3872	0.0389	0.3964	0.0393	0.3847	0.0360
Sioux	0.3875	0.0382	0.3974	0.0396	0.3948	0.0395
Woodbury	0.3896	0.0351	0.3965	0.0377	0.4057	0.0378
Portfolio1	0.3900	0.0357	0.3896	0.0371	0.3976	0.0352
Portfolio2	0.3905	0.0357	0.3899	0.0372	0.3976	0.0354

Table 1.6: Revenue Insurance (Price Vol = 0.36): Probability of Loss and Premium Rates

	Gaussian Copula		Student-t Copula		R-Vine Copula	
	Pr(Loss)	Rate	Pr(Loss)	Rate	Pr(Loss)	Rate
65% Coverage Level						
Cherokee	0.0891	0.0102	0.0865	0.0107	0.0904	0.0105
Ida	0.0871	0.0102	0.0871	0.0108	0.0872	0.0093
Plymouth	0.0873	0.0099	0.0847	0.0106	0.0907	0.0095
Sac	0.0900	0.0108	0.0915	0.0118	0.0882	0.0102
Sioux	0.0905	0.0105	0.0883	0.0113	0.0937	0.0110
Woodbury	0.0841	0.0096	0.0872	0.0108	0.0907	0.0097
Portfolio1	0.0875	0.0100	0.0864	0.0108	0.0887	0.0097
Portfolio2	0.0876	0.0100	0.0863	0.0108	0.0882	0.0097
75% Coverage Level						
Cherokee	0.1968	0.0273	0.1954	0.0276	0.2014	0.0282
Ida	0.1952	0.0271	0.1962	0.0270	0.2047	0.0272
Plymouth	0.1941	0.0270	0.1926	0.0273	0.2054	0.0277
Sac	0.2027	0.0283	0.1990	0.0292	0.1978	0.0277
Sioux	0.1999	0.0279	0.1958	0.0282	0.2080	0.0292
Woodbury	0.1927	0.0263	0.1910	0.0276	0.2070	0.0279
Portfolio1	0.1948	0.0269	0.1951	0.0277	0.2021	0.0274
Portfolio2	0.1950	0.0269	0.1951	0.0277	0.2019	0.0274
85% Coverage Level						
Cherokee	0.3368	0.0551	0.3387	0.0554	0.3443	0.0569
Ida	0.3383	0.0549	0.3387	0.0555	0.3487	0.0563
Plymouth	0.3392	0.0549	0.3374	0.0550	0.3550	0.0574
Sac	0.3431	0.0567	0.3430	0.0573	0.3456	0.0560
Sioux	0.3407	0.0561	0.3386	0.0562	0.3523	0.0585
Woodbury	0.3383	0.0541	0.3405	0.0554	0.3541	0.0575
Portfolio1	0.3390	0.0548	0.3397	0.0556	0.3489	0.0564
Portfolio2	0.3392	0.0548	0.3395	0.0556	0.3488	0.0565
95% Coverage Level						
Cherokee	0.4853	0.0927	0.4917	0.0932	0.4946	0.0950
Ida	0.4854	0.0926	0.4900	0.0933	0.5002	0.0951
Plymouth	0.4881	0.0928	0.4918	0.0930	0.5050	0.0966
Sac	0.4889	0.0943	0.4896	0.0949	0.4930	0.0941
Sioux	0.4870	0.0937	0.4958	0.0942	0.4983	0.0971
Woodbury	0.4895	0.0921	0.4893	0.0933	0.5007	0.0965
Portfolio1	0.4858	0.0926	0.4917	0.0936	0.5002	0.0951
Portfolio2	0.4857	0.0926	0.4924	0.0936	0.5013	0.0952

Table 1.7: Example of Grid Premiums and Discounts ($/cwt carcass)

		YG1	YG2	YG3	YG4	YG5
Prime		16.201	14.070	12.481	1.257	-4.203
CAB		8.875	6.744	5.155	-6.069	-11.529
Choice		3.720	1.589	Base	-11.224	-16.684
Select		-13.413	-15.544	-17.133	-28.357	-33.817
Standard		-26.372	-28.503	-30.092	-41.316	-46.776
Dark	-34.324					
Heavy (Carcass >900 lbs)	-8.446					
Light (Carcass <550 lbs)	-19.802					

Table 1.8: Empirical Study 2 – Fed Cattle Descriptive Statistics (No. Observation: 8569)

Varibale	Mean	Std. Dev.	Minimum	Maximum	Censoring Degree (%)
Quality Grades (%)					
Prime	0.863	1.628	0.000	22.600	56.459
CAB	4.354	5.388	0.000	53.500	28.300
Choice	45.671	17.962	1.700	100.000	0.000
Select	43.282	14.887	0.000	87.800	0.187
Standard	1.955	5.691	0.000	73.200	71.152
Yield Grades (%)					
YG1	21.197	12.829	0.000	95.800	0.887
YG2	43.809	10.339	0.000	90.300	0.012
YG3	31.947	13.752	0.000	100.000	0.082
YG4	2.859	3.471	0.000	38.800	24.577
YG5	0.189	0.594	0.000	8.600	84.817
Discount Factors (%)					
Dark Cutters	1.179	2.725	0.000	37.300	59.680
Heavy	1.664	3.624	0.000	50.000	57.580
Light	1.371	2.956	0.000	61.700	54.674
Independent Variables					
Avg. Place Weight (.lb)	758.830	97.822	371.000	1162.000	
Steers	0.523				
Heifers	0.339				
Mixed	0.138				
Kansas	0.894				
Nebraska	0.106				
Spring	0.251				
Summer	0.239				
Fall	0.257				
Winter	0.252				

Table 1.9: Empirical Study 2 – Parameter Estimates for Residual Marginals

	Optimal	Mean/Location[1]	Stand Dev./Scale[2]
Prime	*Logistic*	0.878 (0.012)	0.681 (0.006)
CAB	*Logistic*	0.775 (0.048)	2.582 (0.024)
Choice	*Normal*	0.000 (0.188)	17.41 (0.133)
Select	*Normal*	0.009 (0.157)	14.578 (0.111)
Standard	*Logistic*	4.007 (0.036)	2.063 (0.020)
YG1	*Logistic*	-0.261 (0.132)	7.021 (0.063)
YG2	*Logistic*	-0.014 (0.108)	5.750 (0.052)
YG3	*Normal*	0.004 (0.147)	13.573 (0.104)
YG4	*Logistic*	0.399 (0.031)	1.657 (0.015)
YG5	*Log-Normal*	0.845 (0.002)	0.198 (0.002)
Dark	*Logistic*	1.302 (0.017)	0.964 (0.009)
Heavy	*Logistic*	1.712 (0.025)	1.387 (0.013)
Light	*Logistic*	1.122 (0.019)	1.095 (0.010)

[1] This column displays “Mean" for variables with a normal distribution and “Location" for variables with a logistic distribution.
[2] This column displays “Standard Deviation" for variables with a normal distribution and “Scale" for variables with a logistic distribution.
Parenthesis contains standard errors

Table 1.10: Empirical Study 2 – Goodness-of-fit for Copula Models

	Gaussian	Student-t	R-Vine	C-Vine	D-Vine
AIC	-53200.58	-56386.38	-74278.07	-72577.47	-64631.03
BIC	-52650.22	-55828.96	-73374.91	-71603.76	-63918.39
Lglk	26678.29	28272.19	37267.04	36426.74	32416.52

Table 1.11: Knedall's Tau Correlation Matrix for Yield and Quality Grades and Discount Factors

	YG1	YG2	YG3	YG4	YG5	Dark	Heavy	Light
Prime	-0.185	-0.077	0.184	0.180	0.098	-0.076	-0.093	0.157
CAB	-0.307	0.013	0.218	0.236	0.072	-0.069	-0.012	-0.043
Choice	-0.347	-0.084	0.331	0.205	0.051	-0.080	-0.074	0.052
Select	0.238	0.127	-0.260	-0.175	-0.054	-0.006	0.056	-0.090
Standard	0.157	-0.079	-0.071	-0.108	-0.023	-0.077	-0.095	0.156
Dark	0.081	-0.009	-0.056	-0.109	0.037	–	–	–
Heavy	-0.024	0.007	0.008	0.047	0.103	0.058	–	–
Light	0.047	-0.005	-0.023	-0.008	0.014	-0.064	-0.189	–

Table 1.12: Knedall's Tau Correlation Matrix within Yield Grades and Quality Grades

Within Quality Grades		Prime	CAB	Choice	Select
	CAB	0.213	–	–	–
	Choice	0.286	0.328	–	–
	Select	-0.297	-0.226	-0.698	–
	Standard	0.065	-0.243	-0.065	-0.010

Within Yield Grades		YG1	YG2	YG3	YG4
	YG2	0.001	–	–	–
	YG3	-0.578	-0.395	–	–
	YG4	-0.358	-0.276	0.388	–
	YG5	-0.092	-0.091	0.093	0.184

Table 1.13: ARIMA Forecasts for Premiums and Discount Factors

	Forecast (20 Weeks)	
Variables	Mean	Std. Error
Prime	12.481	0.062
CAB	5.155	0.101
Select	-17.133	3.617
Standard	-30.092	1.178
YG1	3.720	0.008
YG2	1.589	0.004
YG4	-11.224	0.124
YG5	-16.684	0.050
Dark	-34.324	0.212
Heavy	-8.446	0.002
Light	-19.802	0.034

Table 1.14: Cattle Revenue Insurance – Probability of Loss and Premium Rates

	Benchmark		Gaussian Copula		Student-t Copula		R-Vine Copula	
	Pr(Loss)	Premium Rate	Pr(Loss)	Premium Rate	Pr(Loss)	Premium Rate	Pr(Loss)	Premium Rate
75% Coverage Level	0.0452	0.0027	0.1196	0.0093	0.1202	0.0094	0.1201	0.0093
85% Coverage Level	0.1852	0.0148	0.3305	0.0335	0.3330	0.0336	0.3396	0.0336
95% Coverage Level	0.4354	0.0452	0.5928	0.0786	0.5912	0.0788	0.5911	0.0787
100% Coverage Level	0.5667	0.0680	0.7065	0.1073	0.7082	0.1075	0.7082	0.1074

Figure 1.1: Local Quadratic Regression of Iowa Counties

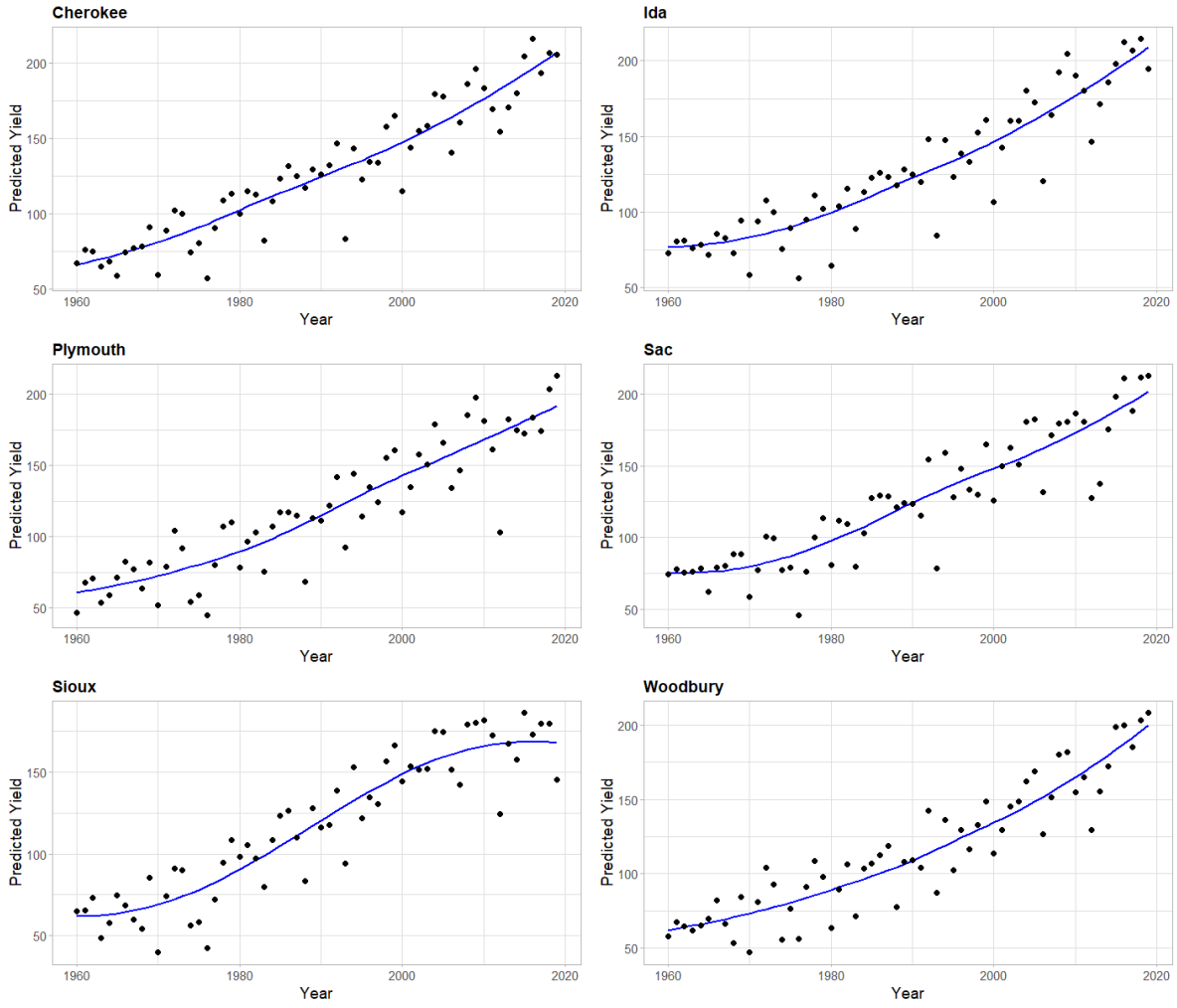
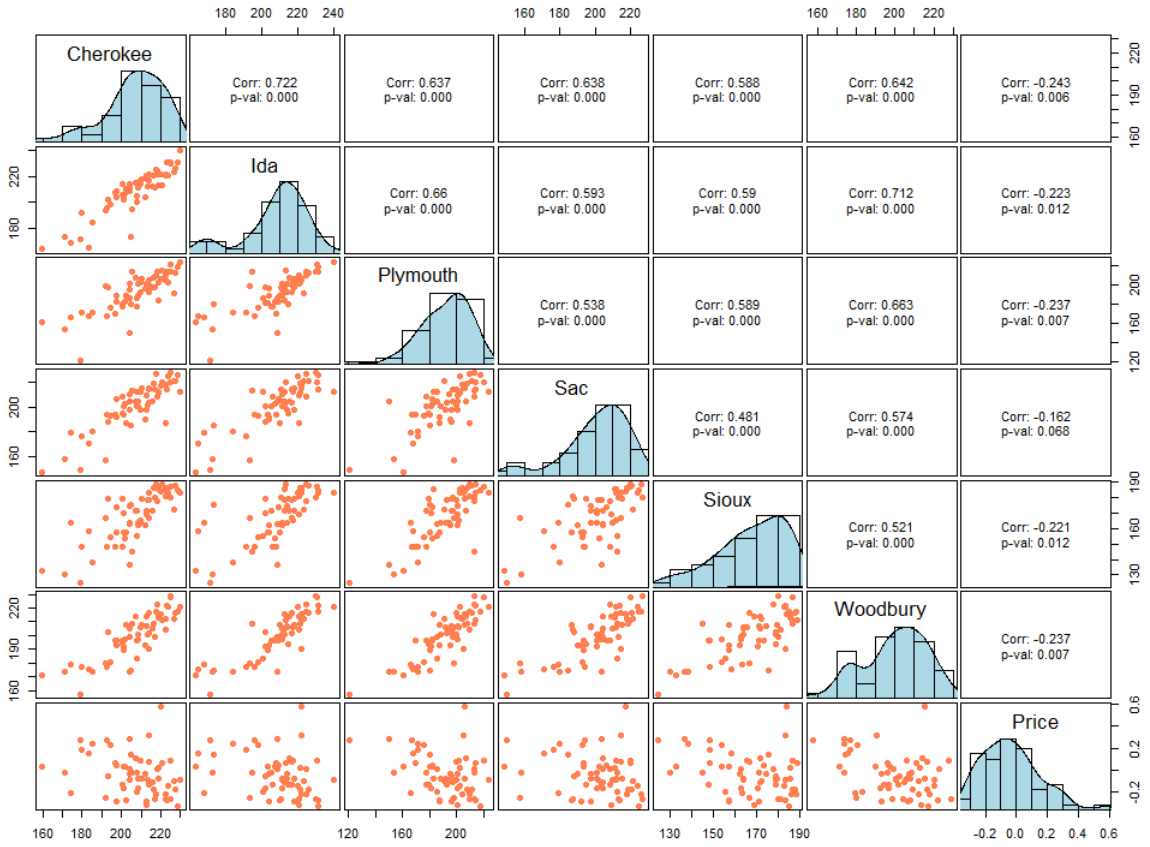


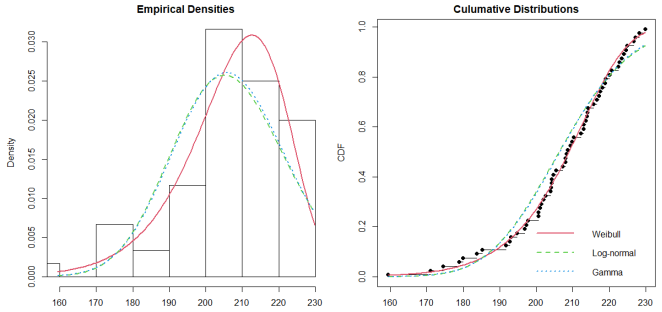
Figure 1.2: Correlations and Empirical Distribution for Detrended County Yield and Price



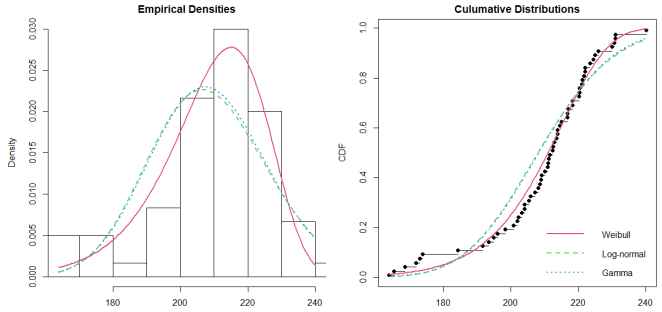
Notes: The diagonal panels plot the histogram and density for each variable; the upper panels illustrate Kendall's tau correlation values with corresponding p-values, and the lower panels visualize the scatterplots of each data pair.

Figure 1.3: PDF and CDF for Selected Counties Yield

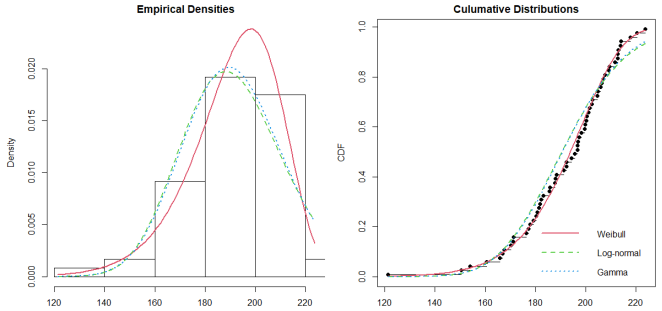
(a) Cherokee



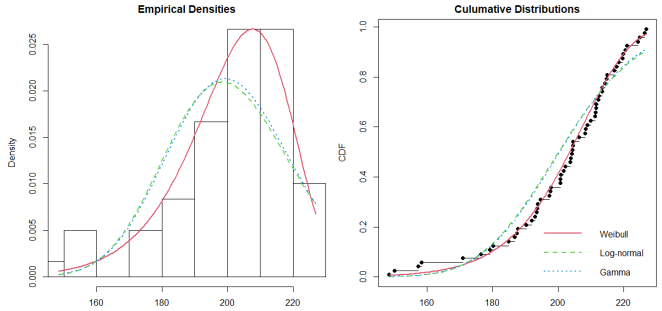
(b) Ida



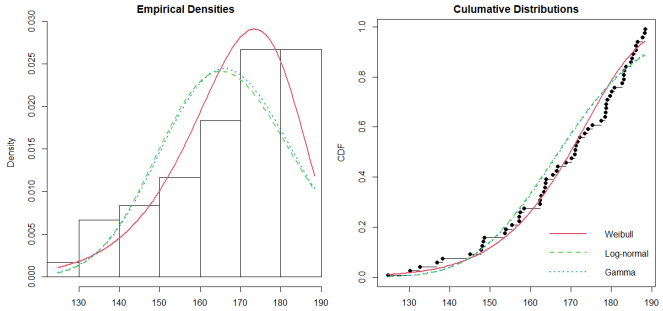
(c) Plymouth



(d) Sac



(e) Sioux



(f) Woodbury

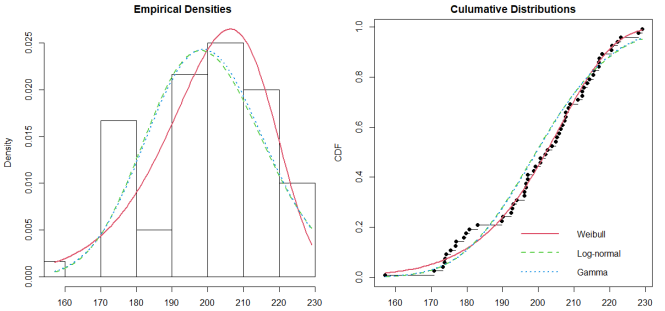
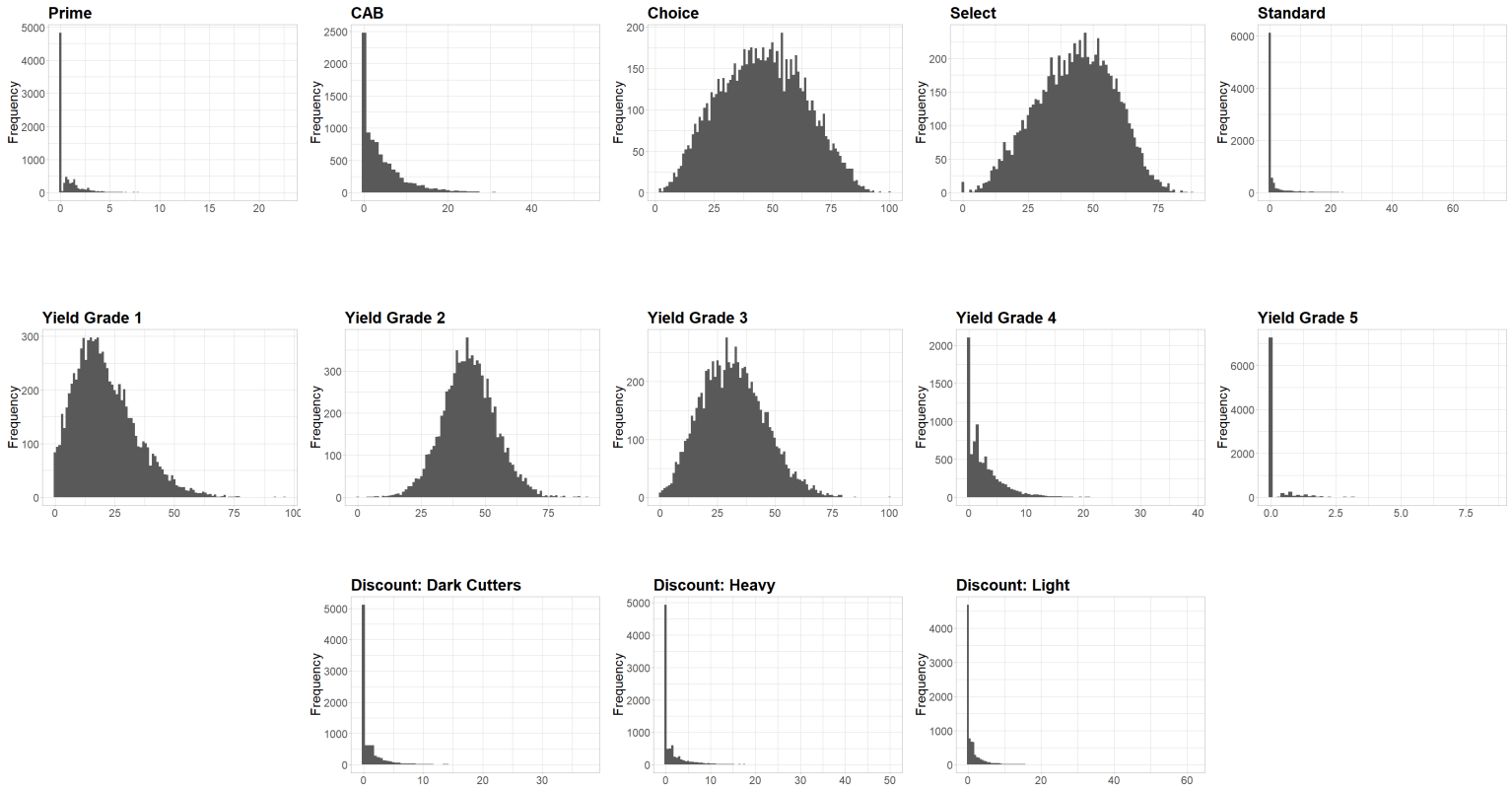


Figure 1.4: Histograms of Percentage Weight of Quality and Yield Grades, and Discount Factors



Notes: contains histograms for five quality grades (first row), five yield grades (second row), and three discount factors (third row). A carcass is discounted as“"heavy" if it weighs greater than 900 lbs, and is discounted as “light" if it weighs less than 550 lbs.

CHAPTER

2

EVALUATION OF SUBSTITUTABILITY BETWEEN THE U.S. AND CANADIAN SOFTWOOD LUMBER THROUGH A THRESHOLD MODEL

2.1 Introduction

The United States (U.S.) holds a prominent position as one of the world's largest producers and consumers of softwood lumber. According to data compiled by the U.S. International Trade Commission (USITC) during an investigation into softwood lumber products from Canada, the apparent softwood lumber consumption in 2022 reached nearly 53 billion board-feet, a 12.8% increase from the 2016 consumption of 47 billion board-feet (USITC 2023). A fair share of the softwood lumber consumed in the U.S. is sourced from Canada. The USITC reports that in 2022, U.S. producers have a market share of 68.7%, while Canada's share is approximately 24.1%, leaving the remaining 7.2% for other exporting markets.

Additionally, in 2015, approximately 41% of U.S. lumber production and 95% of Canadian production came from species common to both countries (USITC 2023). Considering the significant contribution of Canadian exports to U.S. softwood lumber consumption and the overlapping lumber species, the objective of this study is to address the level of substitutability between softwood lumber products in the U.S. and Canada.

Substitutability between imports and domestic-like products is a crucial aspect of international trade analysis, as it determines the efficacy of trade restriction policies on both imports and domestic markets (Gan 2006). If softwood lumber products from Canada are considered substitutes for domestically produced lumber in the U.S., the influx of foreign products with lower prices into the domestic market could harm the U.S. industry. In such a scenario, the imposition of import restrictions would become a desirable protection measure for U.S. producers.

There has been a longstanding trade issue between the U.S. and Canada in the softwood lumber industry, known as the U.S.-Canada Softwood Lumber Dispute. Over the history of the dispute, the U.S. has imposed several trade restrictions on Canadian softwood lumber imports, including antidumping[1] duties (AD) and countervailing duties (CVD). Although various short-term agreements, typically effective for five to nine years, have been implemented to temporarily alleviate the conflict, there was no lasting resolution established, leaving the ongoing friction between the two countries unresolved. The persistence of tension in the bilateral lumber trade can be attributed, in part, to inherent disparities in population levels, and the composition of the lumber market. The U.S. and Canada possess abundant forest resources, but their distinct population sizes and development pressures result in markedly different lumber market scales (Hoover and Fergusson 2018). As the U.S. stands out as the largest producer, importer, and user of lumber, and Canada being its primary exporter, these two countries are particularly susceptible to any changes in trade policies and exogenous shocks from the other market (Devadoss 2009).

This dispute is centered around two issues. The first issue is grounded in the fundamental differences in forestry management practices between the two countries, for which subsidies received by the Canadian lumber industry are a central point of contention (Rahman and Devadoss 2002). Approximately 93.3% of forestland in Canada is publicly owned[2] (Canada Forest Service 2022), compared to 42% of federally and state-owned forest-

[1] The term “dumping" denotes the act of selling lumber to a foreign country at a price below the production cost or selling at the price lower than domestic prices. Allegations of dumping were raised by U.S. lumber producers, contending that the price of Canadian imports to the U.S. market is lower than the production cost in Canada.

[2] The total forestland in Canada is approximately 894 million acres, with about 75.4% of it being provincially

land in the U.S.[3] (Hoover and Riddle 2021). Since the majority of the timber is harvested from provincial crown land in Canada, U.S. lumber producers claim that Canadian stumpage fees, the prices landowners charge for the right to harvest timber, are not determined by a competitive market. This, according to their argument, results in artificially low prices for timber produced in Canada. In contrast, U.S. timber is typically harvested from privately owned lands, where stumpage fees are determined through competitive auctions (Sedjo 2006). Subsidizing an exported commodity violates the World Trade Organization (WTO) Subsidies Agreement, thereby providing U.S. producers with a case against Canadian exporters within the WTO as justification for a CVD against the product category.

The second issue fueling the dispute is the substitutability of imported and domestic lumbers. U.S. producers argue that Canadian lumber imports are substitutes for domestically produced lumber products, thereby unfairly competing with domestic producers and harming the U.S. lumber industry. The issue of substitutability is an important competition factor in the threat of injury analysis (USITC 2019). According to provisions outlined in the U.S. code §1671, if the administering authority concludes that a countervailable subsidy—whether directly or indirectly—has been provided to the production of a commodity imported to the U.S. and such subsidization is causing material injury or threatening material injury to a U.S. industry, a CVD may be imposed on the concerned merchandise. The amount of the CVD is determined to be equal to the net countervailable subsidy (19 U.S.C. § 1671(a)). The imposition of AD requires the establishment of two similar determinations. The administering authority must determine that foreign merchandise is being sold in the U.S. at a rate below its fair value, and a U.S. industry is experiencing material injury or is under the threat of material injury (19 U.S.C. § 1673).

Despite contentions raised by U.S. producers, the Canadian government responds to the allegation by stating that lumber is used in a wide range of industries and, therefore, does not qualify as a subsidy under the U.S. law (Rahman and Devadoss 2002). Moreover, U.S. consumers claim that domestically produced lumber and Canadian lumber are not substitutes because the types of lumber used for housing framings serve different purposes and are not interchangeable (Nagubadi et al. 2004). These perspectives assert that the trade restrictions imposed on Canadian lumber imports are motivated more by protectionist sentiments than by an objective response to perceived unfair practices (Institute 2022).

owned, 13.0% territorial, 2.1% indigenous, and 1.7% federal, leaving 6.7% privately owned land (Canada Forest Service 2022).

[3]The total forestland in the U.S. is approximately 765 million acres. As of 2017, 31% of forests are federally owned, with state and local governments owning 11%. In addition, 38% is owned by private non-corporate households, while private corporations own 20% (Hoover and Riddle 2021).

According to economists at the Montreal Economic Institute (MEI), the placement of tariffs harms the economics of both countries, and such protectionist measures act as barriers to free trade (Rancourt and Giguère 2022). Their analysis indicates that the financial impact of the imposition of tariffs on the U.S. population is 26 times more significant than that on Canadian producers, which translates to a reduction in economic welfare estimated at $1.5 billion for consumers in the U.S. (Rancourt and Giguère 2022). Furthermore, the decrease in welfare for U.S. consumers is projected to exceed the gains experienced by American producers by 7% (Rancourt and Giguère 2022).

Various studies have delved into assessing the welfare impacts arising from trade conflicts within the forestry sector, with scholars finding little support for trade regulatory policies based on statistical estimations. For instance, Buongiorno and Johnston (2018) found that import barriers on foreign lumber increase the welfare of U.S. producers, but the increase was less than the losses incurred by U.S. consumers. On a similar note, van Kooten and Johnston (2014) observed that there would be a very small change in U.S. welfare from the removal of export restrictions on Canadian lumber, estimated at around $16 million, while Canada would gain $91.8 million. Johnston and van Kooten (2017) estimated the effects of the tariff rate quota used in the 2006 Softwood Lumber Agreement (SLA2006), which resulted in an annual deadweight loss of $28 million, borne by U.S. consumers. Johnston and van Kooten (2017) further noted that the absence of SLA2006 would increase the supply share of Alberta while improving aggregate welfare. The formulation of trade policies is complicated and often involves balancing the interests of the executive and legislative branches. Rather than attempting to estimate the welfare changes associated with existing trade duties, the focus of this study shifts towards investigating the substitutability between lumber products from the perspective of Canadian and U.S. lumber price parity.

Earlier studies on substitutability between lumber species yield various results, primarily focusing on calculations of price elasticities of substitution. As noted in USITC's latest report on investigations of Canadian softwood lumber products, the estimated substitution elasticity between U.S.-produced softwood lumber and subject-imported softwood lumber from Canada ranges between 2 and 5 depending on species and purchasers' preferences (USITC 2023). Based on the evidence, the Commission considers the domestic product and subject imports from Canada to have "at least moderate degree of substitutability" (USITC 2023). Nagubadi et al. (2004) investigated the relationship among six lumber species groups in the U.S., including ones imported from Canada, using a restricted translog subcost function. They computed Allen partial elasticities of substitution and identified that the Spruce-Pine-Fir group from Canada is a substitute for untreated Southern Yellow Pine and

structural panels. Gan (2006) calculated the Armington elasticity of substitution between U.S. domestic and imported forest products, finding that elasticity estimates are generally small and stable, although the aggregated level estimates are increasing, which indicates increasing substitutability of imports for domestic forest products over time. Shook et al. (2007) estimated the cross-price elasticity of demand for U.S. softwood lumber species and did not find strong evidence that Canadian softwood species behave as substitutes for U.S. domestic species. Roman et al. (2006) employed a linear approximate almost ideal demand system for U.S. softwood lumber imports from various countries. They found that compared to Canadian lumber imports, the elasticities of substitution from other countries are higher, implying a higher degree of substitutability with U.S. domestic products.

Given the relatively small literature on lumber substitutability, this study contributes to the literature by utilizing threshold models and examining the dynamic price behaviors between the U.S. and Canadian lumber products. Threshold models are commonly applied to assess the degree of market linkage across space, market form, and time. Price transmission, an essential aspect of market integration analysis, is frequently discussed with the notion that external shocks in one market can impact other markets (von Cramon-Taubadel and Goodwin 2021). Market integration for a commodity can be examined both vertically and horizontally. Vertical market integration examines price relationships at different levels of the supply chain, while horizontal integration assesses the extent of market deviation across different locations for identical products (for example, see Goodwin and Holt 1999; Goodwin and Piggott 2001; Bekkerman et al. 2013; Goodwin et al. 2017; Ramsey et al. 2021; Goodwin et al. 2021). In general, arbitrage occurs when the price difference between two markets, either spatially or vertically linked, exceeds the arbitrage costs (for example, transaction costs). Rational agents will then buy in the market where prices are low and sell in the market where prices are high until the price differences are diminished, bringing the market back to its equilibrium.

Although the estimation strategy of this study follows a similar mechanism to traditional market integration analysis, we employ threshold models with a different motivation. In the context of substitutability, when the deviation of price parity exceeds the thresholds, instead of observing arbitrage behaviors that transfer the product from one market to the other, what we observe is the potential shift in U.S. consumers' consumption behaviors from one product to another. Our empirical analysis focuses on Southern Yellow Pine (SYP), one of the leading softwood species produced in the U.S., and Spruce-Pine-Fir (SPF), the predominant softwood species imported from Canada. Both lumber species groups take major shares of U.S. consumption, therefore, their fungibility is always a focal point in

discussions concerning trade policies for Canadian lumber. SPF has a high strength-to-weight ratio, meaning the lumber is lighter but can hold a large amount of weight (AIFP Lumber 2019, 2020). It can also hold nails better and is less susceptible to warping compared to SYP. These characteristics of SPF make it preferred by builders for wall framing. On the other hand, SYP has a much denser fiber composition, making it heavier, and often suited for beams and joists (Nagubadi et al. 2004).

USITC's report defines SPF as a "like product" to SYP as the two share similar physical characteristics and usages and they are sold through overlapping channels of distribution (USITC 2023). In addition, the Commission collected responses to surveys and questionnaires regarding the two lumber products' interchangeability, U.S. producers commonly depict softwood from Canada and domestically produced softwood as "frequently interchangeable", while survey information from the National Association of Home Builders indicates that builders characterized them as "sometimes interchangeable" (USITC 2023). Given the divergence in responses between producers and purchasers, it is reasonable to suggest that consumers are less likely to substitute away from their current softwood product choice in response to a price increase if they value the price increase less than their individual or location preferences. That is, there may exist a threshold in the price differential that describes the degree of substitutability between lumber species. The two lumber species become substitutable only when their price differential exceeds a certain threshold value. Further, the degree of substitutability would become more prominent if the price differential grew larger.

To address the potential presence of threshold effects in the context of lumber substitutability, we employ the Self-Exciting Threshold Autoregressive (SETAR) model taking the log of price differentials as a forcing variable. Our underlying hypothesis is that when the price differentials between the two lumber species fall below the threshold value, the two markets operate independently. Substitution behaviors would only emerge when the price differentials exceed the threshold value. We then consider bivariate Threshold Vector Error Correction (TVEC) models for the paired-price system to jointly evaluate the interactions among prices, providing a more comprehensive analysis of price dynamics. The TVEC model extends the conventional Vector Error Correction (VEC) model by incorporating threshold effects, allowing asymmetric adjustments to exogenous shocks based on the regime of the system. It is worth noting that SYP produced in various regions of the U.S. may exhibit different degrees of substitutability to SPF imported from Canada. The heavy and bulky nature of lumber products often leads to regional preferences for lumber selection because of product accessibility. Moreover, we observe a higher increase in housing

demand in the Southern region of the U.S., potentially contributing to regional variations in the price relationship. To address this issue, we include an additional analysis for SYP produced in three different regions in the U.S. to investigate the regional aspect of lumber production.

With the confirmation of threshold effects, we identify one threshold in lumber price differentials, leading to two regimes referred to as the “low regime" and the “high regime". The low regime occurs when price differentials are below the threshold, the case when the two lumber markets are independent of each other. On the other hand, the high regime occurs when price differentials exceed the threshold, and the two lumber species become substitutes. We then examine the nature of adjustment using both standard and generalized impulse response analysis. We observe that the effect of a shock is significantly more persistent in the high regime than in the low regime. This result is consistent with a situation where substitutability is limited when prices are similar but the products are more likely to be substituted for one another when prices are significantly different. In the regional analysis, the parameter estimates of the regional models show no significant differences among various regions. In addition, SYP prices in all regions respond similarly to shocks from SPF, and vice versa, suggesting no significant difference between the relationships of regional SYP and SPF prices.

The rest of the paper is organized as follows. In Section 2.2, we provide a brief review of the history of the softwood lumber dispute. Section 2.3 outlines the econometric methodology. Section 2.4 describes the data and model specifications. In Section 2.5, we analyze the empirical findings of the analysis. Lastly, Section 2.6 concludes.

2.2 Background: History of the Softwood Lumber Dispute

The modern history of the U.S.-Canada softwood lumber dispute can be divided into five phases: Lumber I (1981-1983), Lumber II (1986-1991), Lumber III (1991-1994), Lumber IV (2001-2006), and Lumber V (2015-present). Among these phases, three short-term settlement agreements were signed to address the dispute, yet a lasting and comprehensive resolution has remained unattainable throughout the historical revolution of the conflict up to the present day.

In 1981, a group of U.S. lumber producers petitioned the U.S. Department of Commerce (USDoC) and the U.S. International Trade Commission (USITC) for an investigation into Canadian lumber imports to the U.S. for a possible countervailing duty (CVD) (Gorte and

Grimmett 2006). This first investigation is known as Lumber I. These producers claimed that Canadian softwood lumber products were unfairly subsidized by the government, allowing Canadian lumber producers to sell at a price significantly below the market value of timber (Rahman and Devadoss 2002). While the USITC initially collected evidence indicating that Canadian lumber imports were causing harm to the U.S. lumber industry, in 1983, the International Trade Administration (ITA) of the USDoC determined that the subsidies amounted to *"de minimis"* (less than 0.5%) and that Canadian stumpage programs were not specific to a single industry, making them non-countervailable (Lindsey et al. 2000). As a result of this negative CVD determination, the initial CVD investigation concluded without imposing any duties on Canadian softwood lumber imports.

After Lumber I, the U.S. lumber industry group, the Coalition for Fair Lumber Imports (Coalition), initiated another countervailing duty (CVD) investigation into Canadian softwood lumber imports in 1986. This time, the ITA reversed its previous determination[4] and found evidence that Canadian lumber producers had received a subsidy of 15% ad valorem (Gorte and Grimmett 2006). One day before the final CVD determination, the two countries negotiated and signed the Memorandum of Understanding (MOU). Under the terms of the MOU, the Canadian government committed to imposing a 15% tax on lumber exported to the U.S. for the subsequent five years. The Coalition withdrew its petition immediately after the signing of the MOU (Rahman and Devadoss 2002). This marked the first legal agreement of the ongoing dispute.

The third investigation of Canadian softwood lumber imports, or Lumber III, began with the termination of the MOU when the Canadian government withdrew in September 1991. In the following October, the U.S. Trade Representative (USTR) self-initiated an investigation upon the pending completion of the previous CVD investigation. This investigation confirmed substitutability between lumber species groups in the U.S. market and concluded with a final determination that Canadian softwood lumber causes material injury to the U.S. industry, and exports are countervailable. As a result, the U.S. government imposed a 6.5% ad valorem duty on Canadian lumber imports (Congressional Research Service 2001). Canada appealed the final determination of material injury and the final CVD of 6.5% under the U.S.-Canada Free Trade Agreement (FTA). This phase of the dispute was settled with a final FTA panel decision ordering the DoC to revoke the CVD order (Gagné 2003).

In 1996, both countries were interested in finding a durable solution to the dispute.

[4]Based on a new guideline issued by DoC that outlines which subsidies are subject to the CVD laws.

Upon numerous negotiations, the first softwood lumber agreement emerged, known as the 1996 Softwood Lumber Agreement (SLA1996). Under SLA1996, a tariff-rate quota of 14.7 million board-feet was established for Canadian softwood lumber imports produced in British Columbia, Alberta, Ontario, and Quebec (Gagné 2003). The Canadian government was responsible for collecting fees on over-quota shipments, in return, the U.S. government would alleviate any trade restrictions throughout the agreement. However, when SLA1996 expired in March 2001, the countries failed to settle for a new agreement, leading to the fourth phase of the dispute.

The Coalition submitted another petition for both CVD and Antidumping Duty (AD) shortly after the expiration of SLA1996. In August 2001, the DoC announced an initial CVD of 19.3% and an AD of 3.1% on Canadian softwood lumber imports, and the final determination of CVD and AD amounted to 27.2% in May 2002 (Yin and Baek 2004). In April 2006, the U.S. and Canada reached a framework agreement to form an end to the dispute. Under this tentative agreement, Canadian lumber products would account for no more than 34% of the U.S. lumber market share, and an export tax would be imposed on lumber products exported to the U.S. with a price lower than $355 per thousand board-feet (CBC News 2006). In addition, the U.S. agreed to refund 80% of the $5 billion duties collected in the previous four years and to withhold from initiating trade actions or imposing new tariffs (CBC News 2006). This agreement officially took effect on October 12, 2006, known as the 2006 Softwood Lumber Agreement (SLA2006). It was to last seven years, with a possible two-year extension.

As of the writing of this paper, no further agreement has been reached since the expiration of SLA2006 on October 12, 2015. From 2016-2023, the trade dispute has escalated on several occasions with various duty rates imposed over the years. In July 2023, The DoC set the rate on Canadian softwood lumber imports at 7.99% (Global Affairs Canada 2023). The Canadian government appealed under dispute-resolution provisions in the North American Free Trade Agreement (NAFTA) to the World Trade Organization (WTO) to challenge the levitation of duties. Most recently, the USITC issued its five-year (sunset) review decision on the newest investigation on Canadian softwood lumber on November 30, 2023. All four commissioners arrived at the conclusion that lumber imports from Canada are unfairly subsidized, therefore, remaining a threat to the U.S. lumber industry. Their decision supports the continuation of the countervailing and anti-dumping duty on Canadian softwood imports (USITC Office of External Relations 2023).

2.3 Methodology

We are interested in discovering substitutability between lumber products imported from Canada and those domestically produced in the U.S. The first step in the model specification involves testing the statistical properties of the price series, specifically testing for stationarity and cointegration. A stationary series is characterized by statistical properties, such as mean and variance, that remain time-invariant, and nonstationary series exhibit trends or irregular fluctuations over time. We apply the standard Augmented Dickey-Fuller (ADF) test (Dickey and Fuller 1979) to check the presence of a unit root. The lag length is selected using modified information criteria by Ng and Perron (2001). We also include the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al. 1992) for stationarity, recognizing the potential issue of size distortion in the ADF test. The null hypothesis of the ADF test is that the series has a unit root. Rejection of the null hypothesis indicates that the series is stationary. The null hypothesis of the KPSS test is that the time series is stationary, therefore, rejecting the null indicates the presence of a unit root. If a nonstationary process becomes stationary after the first difference, then this process is denoted as *integrated of order one*, or *I(1)*.

Once we confirm the nonstationarity property of the price series, the natural next step is to check for cointegration. The concept of cointegration was introduced by Engle and Granger in 1987. Given a system containing two or more time series, even the individual time series is nonstationary, if all components in the system are I(d) and there exists a linear combination of them that is stationary, then we call these time series cointegrated. Engle and Granger (1987) proposed a two-step method to test for cointegration between two non-stationary time series. The first step is to regress one variable on the other with OLS and obtain the residuals, and the second step of the procedure is to test residuals for stationarity. If the residual is stationary, then cointegration is confirmed.

Considering two price series of lumber produced in Canada and the U.S., p_t^1 and p_t^2. The linear regression procedure among prices models the long-run relationship is given by

$$p_t^1 = \alpha + \beta p_t^2 + u_t \tag{2.1}$$

where the residual $u_t = \psi u_{t-1} + \xi$ is the error correction term. Note that if both prices are in logarithmic terms, the coefficient β represents the price transmission elasticity, which is the responsiveness or sensitivity of one price to changes in another. If u_t is stationary, then p_t^1 and p_t^2 are cointegrated. The cointegration test was further extended by Johansen (1995) to

test for one or more cointegration relationships for systems with multiple time series. This multivariate modeling framework is based on a maximum-likelihood procedure. Although we only focus on pairwise systems with a maximum of one cointegration relationship, we include Johansen's trace test to confirm the presence of cointegration.

Engle and Granger (1987) further introduced the Granger representation theorem, which states that if a system of time series is cointegrated, then there exists a Vector Error Correction (VEC) representation. The application of error correction mechanisms to econometrics was first introduced in the 1950s (see, Phillips 1958). Thereon, these models have received great interest in economic empirical applications (see works of Davidson et al. 1978; Dawson 1981; Salmon 1982). Consider the following VEC representation with a single lag for two variables that are both integrated of order one,

$$\begin{bmatrix} \Delta p_t^1 \\ \Delta p_t^2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} u_{t-1} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} p_{t-1}^1 \\ p_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \tag{2.2}$$

where $u_{t-1} = p_{t-1}^1 - \beta\, p_{t-1}^2$ is the error correction term from Equation 2.1. The coefficient vector a in front of the error correction term estimates the adjustment to deviations from the long-run equilibrium.

It should be pointed out that the Engle-Granger procedure assumes the cointegration relationship, or the adjustment of short-run dynamics toward the long-run equilibrium, is linear, symmetric, and instantaneous (Balke and Fomby 1997). However, it is possible that the adjustment does not occur in every period. For example, in the context of market integration, a no-arbitrage condition may exist when the price difference is less than the transaction costs. (Balke and Fomby 1997) describe this phenomenon in which economic agents only start the adjustment process when the deviation from the equilibrium exceeds a certain threshold as *"discrete adjustment"*. In the event of substitutability, the relationship between SPF and SYP would theoretically change from unrelated to substitutes only when their price differential exceeds a certain threshold value. The parity relationship between two lumber prices can be described as an autoregressive process:

$$\tilde{p}_t = \delta\, \tilde{p}_{t-1} + v_t \tag{2.3}$$

where $0 < \delta < 1$ and $\tilde{p}_t = p_t^1 - p_t^2$ is defined as the forcing variable, which triggers a regime-switching behavior when passes the threshold value, and v_t is a white-noise error term.

The price parity term $\tilde{p}$ represents the error correction term in Equation 2.1 when the cointegration relationship is at equilibrium where $\alpha = 0$ and $\beta = 1$ are expected.

As mentioned earlier, because of consumer preference on types of lumber products purchased, consumers' responses to price disparities ($\tilde{p}_t$) may not be linear. To test for the presence of nonlinear and asymmetric behaviors in $\tilde{p}_t$, we implement four linearity tests with the null of a linearity Autoregressive (AR) model against a Threshold Autoregressive (TAR) model. The first test is a neural network test for neglected nonlinearity by Lee et al. (1993). This neural network test uses a single hidden layer network, and the null hypothesis of linearity is not rejected if the optimal network weights are zero. The second test is Keenan (1985) One-Degree Test for nonlinearity, which is based on the correlation of estimated residuals and the squared forecast. If the squared forecast has additional forecasting ability, i.e., correlated with the residuals, then the process is considered nonlinear. The third test, introduced by Tsay (1986), is similar to Keenan's test as it directly tests departures from linearity in the mean. Lastly, we perform the likelihood ratio test for threshold nonlinearity introduced by Chan (1993). If the null hypothesis is not rejected, indicating linearity, the price differentials should be modeled using a linear AR model, and the pairwise price system can be evaluated using a VEC model. Otherwise, if the null hypothesis is rejected, suggesting nonlinearity, threshold models are implemented.

The confirmation of nonlinearity indicates the degree of substitutability changes given different levels of price disparity. We then follow the technique proposed by Balke and Fomby (1997) and apply the Self-Exciting Threshold Autoregression (SETAR) to the price differentials. They introduced the concept of threshold cointegration that combines the idea of cointegration with nonlinear threshold modeling developed by Tong (1983, 1990) and Chan (1993). Balke and Fomby (1997) showed that standard tests for unit roots and cointegration are valid under threshold cointegration[5].

The SETAR model is an extension of the linear autoregressive (AR) model, allowing the autoregressive coefficients to assume different values depending on a specific threshold. Let θ be the threshold, the SETAR(1) model with two regime *(L, H)* and one lag can be written as:

$$\tilde{p}_t = (\gamma_L + \delta_L \tilde{p}_{t-1} + v_{L,t-1})\mathbb{1}(\tilde{p}_{t-1} \leq \theta) + (\gamma_H + \delta_H \tilde{p}_{t-1} + v_{H,t-1})\mathbb{1}(\tilde{p}_{t-1} > \theta) \tag{2.4}$$

where $\mathbb{1}$ is the indicator function and equal to 1 if the condition is true and is zero otherwise. Compared to a regular TAR, instead of using an external variable as the forcing variable,

[5]Balke and Fomby (1997) also noted that these tests may lack power in the presence of threshold effects.

the lagged value of $\tilde{p}_t$ is used, hence the name “self-exciting". As can be observed, having one threshold turns the linear AR(1) process into a piecewise linear model containing two regimes for which each regime is described as a unique AR(1) process depending on the state of the forcing variable. In the empirical analysis, we identify the number of thresholds through a test proposed by Hansen (1999) that examines linearity against thresholds, and the SETAR(1) model against the SETAR(2) model. Specifically, we test the null hypothesis of no threshold against one threshold, no threshold against two thresholds, and one threshold against two thresholds. The value of the threshold is determined using a grid search across all possible threshold values as defined by the data.

The VEC model in Equation 2.2 can be generalized to a Threshold VEC (TVEC) model where the adjustment vector a changes depending on the state of the error correction term (Goodwin and Holt 1999; Goodwin and Piggott 2001; Ramsey et al. 2021),

$$\begin{bmatrix} \Delta p_t^1 \\ \Delta p_t^2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_1^L \\ a_2^L \end{bmatrix} u_{t-1}^L + \begin{bmatrix} a_1^H \\ a_2^H \end{bmatrix} u_{t-1}^H + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} p_{t-1}^1 \\ p_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{bmatrix} \tag{2.5}$$

where the error correction term is split into two regimes. This system-based approach to the paired-price system allows for the estimation of individual price responses to shocks. Hansen and Seo (2002) derived a maximum-likelihood (ML) estimator that estimates the threshold and cointegration parameters through a two-dimensional grid search. In addition, Hansen and Seo suggested a sup-LM cointegration test for the null of the linear VEC model against a TVEC model. We apply this test prior to the estimation of the TVEC model.

Impulse response analyses offer a visual interpretation of the dynamic price patterns in response to exogenous shocks and the overall model dynamics. We conduct both standard (orthogonalized) Impulse Response Functions (IRFs), and the Generalized Impulse Response Functions (GIRFs) following the steps outlined by (Koop et al. 1996). We examine the response of the price parity relationship (using linear autoregressive models) and individual prices (using bivariate error correction models) within the softwood lumber industry. If responses diminish quickly after an initial shock, this implies that shocks are transitory and markets behave somewhat independently. Conversely, persistent effects of shocks indicate a strong relationship among the markets.

The standard IRF illustrates how an exogenous shock at time t impacts the process at time $t+h$ and, therefore, measures the persistent effect of shocks on the price series.

Mathematically, an IRF for a process y can be defined as:

$$IRF_y(h,\rho,\omega_{t-1}) = E[y_{t+h}|\epsilon_t = \rho, \epsilon_{t+h} = 0, \omega_{t-1}] - E[y_{t+h}|\epsilon_t = 0, \epsilon_{t+h} = 0, \omega_{t-1}] \quad (2.6)$$

where ω_{t-1} is the observed history of the process, ρ is the size of the shock at time t, and ϵ_t is the error term. One important thing to note is that for the standard IRF, $\epsilon_{t+h} = 0$, that is, no further shocks are assumed to affect the system. Using parameters from each regime of the threshold models, we perform regime-specific impulse responses to observe distinctive lumber price behaviors under different regimes.

Koop et al. (1996) pointed out that the standard IRF is less useful when applied to nonlinear models. In the case of the threshold model, the impacts of the shock for the standard IRF would differ depending on the size and starting regime of ω_{t-1}. Therefore, they introduced GIRFs that allow for the examination of responses not only to one standard deviation shock in a single variable but also to shocks in multiple previously observed variables simultaneously. Consider the following GIRF on process y:

$$GIRF_y(h,\rho,\omega_{t-1}) = E[y_{t+h}|\epsilon_t = \rho, \omega_{t-1}] - E[y_{t+h}|\omega_{t-1}] \quad (2.7)$$

With GIRF, instead of using y_{t+h} without shock ρ as the benchmark, the compared realization of y_{t+h} is an average of future shocks. For the calculation of the realized GIRFs, we follow Koop et al.'s suggestion and draw random bootstrap samples of the estimated error term for the shock ρ and observed price for ω_{t-1}. Similar to the standard IRFs, we consider the regime-specific GIRFs for both responses of the price parity relationship and individual lumber prices.

In summary, our estimation methodology can be described as follows. First, we evaluate the stationarity and cointegration of the time series. If the time series is nonstationary and cointegrated, we check for nonlinearities in the process. Confirmation of linearity in the process implies the application of linear AR models to price differentials and VEC models to pairwise price systems. If threshold effects exist, then SETAR and TVEC models are considered. Lastly, IRFs and GIRFs are presented to analyze the dynamics of price adjustments.

2.4 Data and Model Specification

The weekly lumber price from January 1995 to April 2020 is collected from a private commodity price reporting agency *Fastmarkets*. We intentionally truncated the dataset to exclude the period following extreme price swings in the lumber industry during the COVID-19 pandemic because softwood lumber prices surged in May 2020 attributed to structural changes in the residential construction and labor sectors. Therefore, by excluding these prices, we are able to analyze the relationship between domestically produced and imported lumber without the influence of abnormal exogenous shocks from the economy. We focus on Western Spruce-Pine-Fir (SPF) shipped from Canada[6], and Southern Yellow Pine (SYP) produced in the westside[7] of the Southern U.S. The selected lumber dimensions for both products are 2x10 (inches) random lengths with a lumber grade of 2. All prices are reported in U.S. dollars per thousand board-feet. SYP prices are free-on-board (fob) at the mills and SPF prices include duties or export taxes when applicable.

SPF is the largest imported lumber species group from Canada, constituting 86% of total Canadian softwood lumber imports. While SPF is also produced in the U.S., it only accounts for 5.4% of total softwood production (Shook et al. 2007). SYP, on the other hand, is the largest single species of lumber produced in the U.S., and no SYP is commercially produced in Canada (Shook et al. 2007). It is worth noting that lumber products used in the analysis are kiln-dried and untreated, indicating that they have not undergone chemical processing to resist rot, moisture, and decay. Treated lumber is commonly used for outdoor purposes whereas untreated lumber is primarily used for structural framing in residential constructions. As noted by Nagubadi et al. (2004), treated and untreated lumber products may be complements, since they are often used in the same construction site.

There has been an increasing trend in housing demand in the past decade, particularly in the Southern U.S. region. The new privately owned housing starts for the Midwest, Northeast, South, and West regions are plotted in Figure 2.1. We can observe the housing demand in the Southern region of the U.S. is growing at a larger and faster rate compared to the other three regions. In fact, the South accounts for over half of new home constructions nationwide. To acknowledge the different levels of housing demand, we include a regional analysis of lumber prices to investigate whether the regional aspect of prices reflects a distinct price parity relationship with Canadian softwood lumber imports. We consider

[6]Western SPF: shipped from Prince George, British Columbia, Canada. Duties/export taxes included when applicable.

[7]Westside: mills in Texas, Arkansas, Oklahoma, and Louisiana west of the Mississippi River. Prices are FOB at the mills.

weekly prices from December 2021 to November 2023, incorporating two additional SYP products produced by mills in the eastside[8] and central region[9] of the Southern U.S. We have to limit the analysis to a shorter time frame due to data availability. Because of the shorter time series, we anticipate these studies to yield different results compared to prices collected over a longer timeframe. Nonetheless, we think the shorter time series provides valid insights into price relationships between Canadian lumber imports and domestically produced lumber from different U.S. regions.

The price series collected from January 1995 to April 2020 comprises 1321 weeks and will be occasionally denoted as the “long series" in later discussions. Likewise, prices collected from December 2021 to November 2023 are referred to as the “short series", containing only 104 observations. The summary statistics for all price series are given in Table 2.1. We observe that in the long price series, the means of SPF and SYP are close, with SYP prices showing higher volatility. In contrast, average prices and standard errors of the short series are significantly higher than those in the long series. These differences are expected, given the substantial price swings in the softwood lumber industry during the pandemic. The regional analysis contains data beginning in December 2021, although the most recent lumber prices have reverted to their historical levels, as shown in Figure 2.3, part of the price series reflects the lingering effect of the pandemic.

When comparing each price series in the regional analysis, prices are more differentiated. Out of the four lumber prices, Canadian SPF has the highest price on average and it is the most volatile. For SYP, the U.S.-East region has the highest lumber price on average but the lowest volatility. Lumber prices in the U.S.-West and U.S.-Central regions are very similar, with the U.S.-Central having slightly higher mean and standard deviation. Time series plots of all lumber prices are presented in Figure 2.2 and Figure 2.3, for the long series and short series, respectively. We observe that in general, SPF and SYP follow similar patterns regardless of the time frame, which is an indication of possible cointegration. The higher volatility patterns in the short series are also apparent from the plots.

All prices are transformed to logarithmic form for the rest of the analysis, which allows us to interpret price differentials and marginal effects in proportional terms. We conduct stationarity tests for individual lumber prices and price differentials using the ADF and KPSS tests. When implementing the ADF test, it is important to decide whether to include

[8]Eastside: mills in Georgia, Florida, South Carolina, North Carolina, and Virginia. Prices are FOB at the mills.

[9]Central: mills in Mississippi, Alabama, and Louisiana east of the Mississippi River. Prices are FOB at the mills.

deterministic components of a drift and a deterministic trend. Time series with a unit root, such as a random walk, would exhibit a linear trend graphically when a drift term is included, and a quadratic increasing pattern when a deterministic trend is included. We apply ADF test specifications with and without a non-zero constant (drift). Test statistics are presented in Table 2.2, and p-values are included in brackets. Rejecting the null hypothesis of the ADF test implies stationarity. It is evident that all the long price series are nonstationary without the inclusion of a drift term and become stationary once the drift is included. This suggests that the time series is stationary with a drift, contrary to the properties found in most commodity prices. As an alternative, we include the KPSS test to check for level stationary. The null hypothesis of the KPSS test is strongly rejected for all lumber prices, indicating the price series contains a unit root. For prices in short series, there are no contractions in the results of the two unit root tests, implying that within the later time frame, the lumber prices are nonstationary.

Next, we assess the stationary property of the price differentials. Theoretically, the price differentials should be stationary, as demonstrated by the ADF test result of the long series. However, the price differentials for the short series are nonstationary, which could be attributed to the short data availability and abnormal price fluctuations within each lumber species resulting from the pandemic. Lastly, both ADF and KPSS indicate that prices and price differentials become stationary after the first difference, suggesting that they are integrated of order one. Due to the disagreement within the two unit root tests and for the sake of maintaining consistent analysis for each paired price system, we proceed with the analysis assuming lumber prices are nonstationary and the price differences are stationary and check for cointegration.

The process of testing for cointegration begins with Engle-Granger's two-step approach, by first estimating a linear regression between each paired price, followed by testing for stationarity of the estimated residuals. As suggested by Engle and Granger, if the residuals are stationary, then the two prices are considered cointegrated. Cointegration test results are presented in Table 2.5. The null hypothesis of the ADF test for estimated residuals for each paired price is rejected at least at 10% level, with the exception of SPF and SYP produced in the Central U.S. This indicates that, in most cases, there is evidence supporting a weak form of cointegration between the Canadian and U.S. softwood lumber prices. We then conducted Johansen's trace test for cointegration. The existence of one cointegration relationship in Canada-US$_{East}$ price system and no cointegration relationships are found in the other two regional price systems. It is worth mentioning that for the long price series, both of Johansen's trace tests of $r = 0$ and $r \leq 1$ are rejected, indicating two cointegra-

tion relationships within the system. Given that this is a pairwise system, the number of cointegration relationships equals the number of time series, implying stationary time series processes. This finding contradicts results obtained from Engle-Granger's two-step approach.

We further apply Hansen and Seo (2002) test of linear cointegration against threshold cointegration. P-values are calculated from 1000 bootstrap replications. The null hypothesis is rejected by the long price series system, which perhaps conflicts with Johansen's test results, indicating the presence of threshold cointegration. For the regional price pairs, no evidence supports the presence of a threshold effect in the cointegration. Given the mixed findings from various approaches, we suggest the existence of at least some degree of cointegration within the lumber prices. Therefore, to maintain consistency in the analysis, we have opted to proceed with error correction models for the pairwise price system[10].

We proceed with the model specification procedure, conducting tests to identify nonlinear behaviors in the price differentials. Linearity test results are described in Table 2.6. None of the tests can be rejected for the regional short series, indicating the linear property in regional prices. However, in the long series, evidence of nonlinear patterns is mixed, with two out of the four tests rejecting the null hypothesis of linearity at the 1% level. Based on these results, we apply linear models for the short series and threshold models for the long series. To determine the number of thresholds, we conduct three tests as proposed by Hansen (1999) test for linear Autoregressive (AR) model against a Self-Exciting Threshold Autoregressive (SETAR) model. The three tests are as follows, with test results listed in Table 2.7: no threshold against one threshold (AR vs. SETAR(1)), no threshold against two thresholds (AR vs. SETAR(2)), and one threshold against two thresholds (SETAR(1) vs. SETAR(2)). The results favor one threshold for the long price series and confirm no thresholds in the regional price differentials.

To determine the number of lags included in each model, we compute Akaike information criterion (AIC) and Bayesian information criterion (BIC) for up to four lags. The computed criteria are listed in Table 2.8. There are some disagreements between AIC and BIC in determining the optimal lag length. For instance, in the AR models of east and west regions, AIC suggests a lag length of four, whereas BIC suggests a lag length of three. To maintain a more parsimonious model, we proceed with the analysis following BIC. The

[10]We acknowledge that applying a VEC model to a stationary paired price system could result in model misspecification. However, given that VEC models are a restricted version of VAR models, the coefficient estimates remain valid even if prices are stationary. In addition, we consider it more important to maintain consistency in the analysis across all paired price systems, as this makes easy comparisons across different regions and time frames.

final model specifications are SETAR(1) with four lags and TVEC(3) for the long price series, AR(3) and VEC(2) for the east region, AR(3) and VEC(1) for the west region, and AR(4) and VEC(1) for the central region.

2.5 Empirical Findings

In this section, we interpret the estimated parameters of models specified for each pairwise price system and analyze the pricing dynamics in response to exogenous shocks. Table 2.4 contains the ordinary least squares (OLS) estimates of a simple cointegration regression, where we regress the logarithmic price of SPF on SYP. We observe relatively strong relationships between all pairwise prices, indicated by the high adjusted R^2 values. More specifically, regressions of the short series show stronger relationships than the long series. Since all prices are in logarithmic terms, the slope coefficient β represents the price transmission elasticity, measuring the percent change in one price for every one percent increase in the other prices.

In the conventional spatial market integration analysis, price transmission occurs when speculators transfer goods from one market to another when arbitrage opportunities arise. However, in this study, price transmission behavior is driven by substitutions in the consumption of different lumber species. We observe a close-to-unity price transmission elasticity in the "CA-US$_{West}$" and "CA-US$_{Central}$" markets, suggesting a strong degree of substitution between SPF and SYP within these regional markets. The slope coefficient of "CA-US$_{East}$" is greater than 1, which indicates an over-adjustment of SPF prices to a price shock in SYP. Interestingly, the slope coefficient of the long series is 0.568, suggesting an inelastic price adjustment of SPF to SYP. We can draw two conclusions from this simple price transmission analysis. First, there are significant yet imperfect substitutions between SPF and SYP across all markets and time frames. Although the magnitude of substitution and relationships between lumber prices cannot be precisely depicted by price transmission elasticity, the positive and significant correspondence between prices confirms the price linkage among Canadian and U.S. lumber prices. Second, the inconsistent slope coefficients between the short and long series suggest the presence of nonlinearity in the price linkages, as confirmed by multiple tests described in the previous section.

We continue the analysis with univariate model estimations of the price differentials. As mentioned, no evidence of nonlinearity was found in price differentials within the shorter time frame. Therefore, we proceed with estimation using linear autoregressive (AR)

models. The estimated parameters of AR models are presented in Table 2.9. All estimated parameters are statistically significant except for the fourth lag in the "CA-US$_{Central}$" model, implying a relatively persistent adjustment to price parities in all three regions. Similar coefficient estimates are observed in all three AR models, indicating that the regional aspect of SYP production does not significantly affect its price relationships with SPF. This result is not surprising given the limited 104 observations for the regional analysis and the relatively uniform nature of SYP prices across the three regions during this period. To visually examine how price parity responds to exogenous shocks from itself, we include the orthogonalized (standard) impulse responses (IRFs) for the three AR models, plotted in Figure 2.4. As anticipated, IRFs are similar across all paired price differentials for the three regions. Specifically, we notice the impact of the shock continues to increase in the second week and then gradually decreases. The shocks exhibit high persistence as the impacts linger above zero even after 12 weeks. This suggests that if any exogenous shocks cause a shift in the price differentials, the impact on the price parity is permanent. The persistence responses found in regional IRFs support the hypothesis that SPF and SYP lumber are substitutes.

Since nonlinearity is confirmed within the long price series, we apply a Self-Exciting Threshold Autoregressive (SETAR) model with a single threshold. The estimations of the SETAR model are presented in Table 2.9. The threshold θ is searched over the $[\theta_{5\%}, \theta_{95\%}]$ interval, where $\theta_{5\%}$ corresponds to the 5% quantile of the threshold value[11]. The estimated threshold value of the SETAR model is -0.263. With only one threshold, the model identifies two regimes, which we refer to as the "high" or "low" regime. The low regime occurs when the SPF price is more than 26.3 percentage points below the SYP price. This regime is relatively uncommon, comprising only 5.58% (77 observations) of the dataset. Based on our hypothesis, under the low regime, when the SPF price is below the SYP price by more than 26.3 percent, the two lumber markets behave independently. On the other hand, the high regime contains 94.15% (1239 observations) of the dataset, and it occurs when the SPF price is above the SYP price, or when below, the price disparity is less than 26.3 percent. Under the high regime, we hypothesize that the two lumber species behave like substitutes.

A threshold value of -0.263 may seem arbitrary, therefore, it is helpful to put this number into the context of lumber substitutions. Home builders have a preference for using SPF lumbers for wall framing because they are light in weight and easy to nail, and use SYP lumbers for horizontal structural framing such as beams and joists given their high struc-

[11] Choosing a 5% quantile ensures a minimum of 66 observations in each regime for statistical accuracy.

tural capacity. Although treated wood is not the focus of this study, it is worth mentioning that because of the density characteristics in SYP, it dominates the pressure-treated lumber market, accounting for approximately 85% (AIFP Lumber 2019). When SPF prices are more than 26.3 percentage points lower than SYP prices, consumers tend to choose the lumber type based on their intended usage. However, as SPF prices approach or surpass those of SYP, consumers may switch from SPF to SYP for purposes such as wall framing due to the cost increase, even though the denser nature of SYP makes it more challenging to work with. This is when we observe substitutability between the two lumber products.

It is worth noting that the threshold value found in the long series may explain the absence of nonlinearity in the short series. In theory, the substitution between the two lumber species is triggered in the high regime, and in the low regime, the two lumber species are independent. As observed in Figure 2.3, for the majority of the time, SPF prices are higher than SYP prices, indicating that during this time frame, prices are in the high regime when substitution is found. This is further confirmed by the coefficient estimates, with AR model coefficients closely comparing to those estimated by the high regime of the SETAR model. In the high regime, the estimated parameter for the first autoregressive lag is higher than in the low regime, while the parameter of the second lag is smaller. Moreover, the coefficient estimates of the low regime become insignificant starting at the third lag, compared to a significant fourth lag in the high regime. This implies a relatively faster adjustment to price parity in the low regime.

Our hypothesis is confirmed by results from impulse response analyses. The calculation of IRFs is less trivial for threshold models compared to linear models, as responses depend on the state of the regime. We first consider regime-specific IRFs, constructed using parameters estimated from the SETAR model. The difference in the persistence profile is confirmed by the regime-specific IRFs, plotted in Figure 2.5. The top figure illustrates the IRF in the low regime, where the effects of the shocks dissipate much faster than in the high regime, implying that shocks are transitory. In addition, the IRF in the high regime has a much higher persistence. This result aligns with our hypothesis that the two lumber products are unrelated in the low regime, while become substitutes in the high regime.

One drawback of the standard IRF is that it only reflects the effect of a one-time positive shock at time $t+h$, compared to innovation at time $t=0$. As pointed out by Koop et al. (1996), when nonlinearity is present, the nature of the responses may also depend on the direction and timing of the shocks. Therefore, we include the Generalized Impulse Response Functions (GIRFs) for the SETAR model, following the unified approach suggested by Koop et al. (1996). They recommend constructing GIRFs using different sets of history and shocks,

as these two values jointly determine the response paths. We implement Monte Carlo simulations with 500 bootstrapped histories drawn from the observed data and 500 shocks drawn from the estimated residuals. GIRFs are then computed using 100 replications from the estimated based line and shocked path as described in Equation 2.7[12].

Perhaps it is more meaningful to consider the simulated GIRF replications based on the regime and direction of the shocks. We divide the replicated GIRFs into four groups to explore density plots and aggregated responses. First, we obtain the regime-specific GIRFs by separating the replications based on whether simulations were started in the high or low regimes. We further separate the regime-specific GIRFs given the direction of shocks, which allows us to observe if price differentials respond symmetrically to positive and negative shocks.

Koop et al. (1996) suggested comparing the density of different subsets of future periods, as the persistence of the process can be assessed by observing the convergence of the density. Therefore, we plot the densities for 1, 5, and 10 periods ahead of each subset, as shown in Figure 2.6. The density plots reveal several notable observations. What is the most obvious is the fact that convergence is faster and more uniform in the low regime, as indicated by a narrower density around zero, although the response is not fully absorbed after 10 periods. In addition, densities in the high regime are more volatile. No significant asymmetric is observed between densities when comparing responses to positive and negative shocks. This suggests that lumber consumers' substitution decisions remain consistent for both positive and negative signals in the market.

Next, we calculate the aggregated path of GIRFs[13]. As presented in Figure 2.7, GIRFs are aggregated by regimes and then by the direction of the shocks. From the aggregated GIRF plots, we make the following observations. First, shocks in the high regime show a significantly larger magnitude and higher persistence than those in the low regime, which holds true for both positive and negative shocks. This finding aligns with findings from density plots. When compared to regime-specific IRFs in Figure 2.5, the convergence in the low regime is more persistent in the GIRFs. Second, we find symmetric responses in the high regime to positive and negative exogenous shocks, while in the low regime, positive shocks induce a slightly larger adjustment than negative shocks. This finding is again, consistent with findings from the density plots. In summary, the GIRFs results of the SETAR model

[12]This procedure requires high computational intensity. With 500 histories, 500 shocks, and 100 replications for each, amounting to 25 million simulations.

[13]GIRFs are aggregated by mean, a typical approach in empirical applications. We acknowledge that aggregation may lead to a loss of information conveyed in the response. Nevertheless, we believe the aggregated paths offer valuable insights when comparing across regimes and directions of shocks.

suggest lumber prices in the low regime exhibit a lower degree of substitutions than in the high regime. That is, even if substitution behaviors are observed in the low regime, the degree of substitutability is less pronounced compared to the high regime.

While it is useful to investigate the univariate model based on the price parity relationship, a bivariate approach can provide a more comprehensive analysis of specific changes for each price within the system. We examine a bivariate error correction model for each of the linked lumber markets. Similar to the univariate case, we identify threshold cointegration in the pairwise price system of the long series, while no linearity is observed for the three regional markets in the short system. Therefore, a Threshold Vector Error Correction (TVEC) model is applied to the long series system, and linear Vector Error Correction (VEC) models to the short series. The model estimations are presented in Table 2.10. In general, the error correction terms are larger for SYP than SPF, although no statistical significance is found in models with short series. We once again find that the parameter estimates of the three VEC models are similar, supporting the evidence that there are no significant differences between the three regional paired-price systems. In addition, VEC estimates are similar to the TVEC model estimates in the high regimes, indicating SPF and regional SYP are substitutes to each other.

The TVEC model is estimated in a similar fashion compared to the univariate SETAR model with one threshold and, therefore, the model identifies two regimes. In TVEC models, the regime-switching depends jointly on the cointegrating vector $(1, \beta)$ and the threshold parameter θ. A grid search is conducted over each value of (β, θ) within the 5% and 95% quantiles, and the pair minimizing the sum of residual squared errors is selected. The estimated cointegration vector is (1, -1.004), and the selected threshold value is -0.284, closely aligning with the threshold selected by the SETAR model.

The forcing variable of the TVEC model is the error correction term (ECT). Therefore, the low regime is triggered when $ECT = SPF_t - \beta \times SYP_t$ is below the threshold, i.e., $SPF_t \leq 1.004 \times SYP_t - 0.284$. Since the estimated cointegration vector β is close to $(1, -1)$, the ECT values become close to price differentials and we can interpret the threshold value similarly to the SETAR model. A threshold of -0.284 implies that the low regime occurs when SPF prices are more than 28.3% lower than SYP prices. This regime is relatively uncommon, comprising 6.1% of the observations. The high regime happens when the SPF prices get closer to or surpass SYP prices.

The impulse response analysis of the multivariate system provides richer information as it traces the impacts of changes in one variable on others in the system. IRFs for the three VEC models are presented in Figure 2.8, Figure 2.9, and Figure 2.10. It is not surprising

that the impulse responses look similar across the regions, where SYP prices in all regions respond similarly to shocks from SPF, and vice versa. The effects of shocks are persistent, except for the SPF's response to a shock from SYP, indicating that SPF is less responsive to price changes in SYP. For the TVEC model, the regime-specific IRFs are plotted in Figure 2.11. The panels on the left show impulse responses in the low regime, and the panels on the right show the responses in the high regime. The difference in the speed of adjustment between the two regimes is apparent. The impulse responses in the high regime show no sign of convergence after 12 weeks, whereas in the low regime, the responses become insignificant after week six in most cases. The magnitude of the responses is also much larger in the high regime in both variables. These results suggest similar implications as those from the univariate models. The rapid adjustments in the low regime indicate the independence of the two lumber markets, while the sustained persistence in the high regime suggests substitutability between the two lumber products.

Following a similar procedure as in the univariate case, we compute GIRFs over 100 bootstrap replications from 500 draws of the history and estimated residuals. The shock we consider is a system-wide shock, meaning residuals from both SPF and SYP are contemptuously selected. Since no constraints were applied when selecting the shocks, a replication can contain a positive and a negative shock from SPF and SYP. This implies that we cannot divide the GIRFs by the direction of the shocks. Therefore, the subset was done by regime and then by variable. Density plots for each variable and each regime are presented in Figure 2.12, and there is no evidence of convergence for both SPF and SYP as the densities are less concentrated around zero as the periods after the shocks increase. Moreover, we observe much higher volatilities in the high regime than in the low regime, with no distinct difference in volatility across the two variables.

We then discuss the aggregated GIRF paths, which confirm the different levels of shock persistence found in the regime-specific IRFs, as presented in Figure 2.13. We observe that, for both SPF and SYP, an average shock in the high regime yields stronger and more persistent impacts. In the low regime, the response of SYP converges back to zero after seven weeks, but the response of SPF increases with no sign of convergence. We believe this unexpected path of SPF in the low regime is a result of some extreme behaviors. As pointed out by Stigler (2020), extreme behaviors are found in some GIRF paths under the low regime. Therefore, aggregating over the median instead of the mean helps to avoid this issue. We include the aggregated GIRF using median, as shown in Figure 2.14. Compared to aggregating over the mean, this eliminates the abnormal behaviors in SPF in the low regimes, while there is not much change in response behaviors in the high regimes for both

lumber species, except for slightly higher response magnitudes. Lastly, we observe a faster convergence in the low regime, especially for SPF. This is expected as it aligns with our underlying hypothesis that the two lumber species are independent in the low regime.

Overall, our findings in the univariate price differential approach align with those from the multivariate error correction approach. With the confirmed threshold effect in the long price series in both cases and the lower persistence profile in the low regime in response to a shock, the evidence supports our hypothesis that, in the low regime, SPF and SYP exhibit a much lower degree of substitution.

2.6 Conclusion

Trade in softwood lumber between the U.S. and Canada has significant impacts on both economies, as the lumber market is closely linked to construction, manufacturing, and labor industries. In recent years, global forest products production and consumption have experienced an upswing, driven by increased residential constructions, highlighting the importance of addressing the trade dispute between the U.S. and Canada. The history of the U.S.-Canada Softwood Lumber Dispute involves a recurring pattern between trade litigation and agreements, and the determination of injury to the U.S. industry from Canadian lumber imports is a crucial part of the development of a new trade agreement. Motivated by this aspect of the trade dispute, this paper examines the substitutability between U.S. and Canadian softwood lumber species through a threshold analysis. Theoretically, if the degree of substitutability is not substantial, the implementation of trade restrictions on Canadian softwood imports is likely to have minimal impact on the U.S. market.

For lumber prices collected between 1995-2020, we find strong evidence supporting the presence of threshold effects between Spruce-Pine-Fir (SPF) imported from Canada and Southern Yellow Pine (SYP) produced in the U.S., which can be well characterized by models with one threshold. Our framework examines the nonlinearity in prices through two approaches: a univariate approach that applies a Self-Exciting Threshold Autoregressive (SETAR) model to the price differentials, and a multivariate approach that fits the pairwise price system to a Threshold Vector Error Correction (TVEC) model. The threshold values estimated by both models are similar. SETAR selects a threshold of -0.263, and TVEC selects a threshold of -0.284. This implies that in the low regime, the two lumber species are hypothesized to be unrelated if the price of SPF is more than 26.3 (28.4) percentage points lower than SYP prices. Additionally, the shock persistence in the low regime is much lower

than in the high regime, which supports the hypothesis that the two lumber species are unrelated when their price differentials are below the threshold.

Furthermore, we analyze the pairwise price systems linking SPF prices to regional SYP prices from the east, west, and central regions of the Southern U.S. We find no evidence supporting nonlinear behaviors. There are also no significant differences across different regions. This is not surprising, as the regional analysis contains only 104 observations for each price series, and the majority of the observations are in the high regime according to the estimated thresholds.

Our findings are largely consistent in both approaches and all dynamic impulse response analyses, providing important implications for the ongoing softwood lumber dispute. We confirm that the degree of substitution between SPF and SYP is lower in the low regime, which occurs when SPF prices are more than 26.3 (28.4) percentage points lower than SYP prices. This implies that keeping SPF prices higher than SYP may induce substitutions between the two species, thereby increasing competition. These dynamics of market responses contribute important understanding to the broader discourse on trade policy in the softwood lumber industry.

While this study only focuses on two lumber species, the analysis can be easily generalized to other lumber species produced in other regions of the U.S. and Canada. Overall, we bring an important view on the substitutability of two lumber species, which is a great addition to the literature on the softwood lumber dispute and can provide additional information looking into the future of the dispute. Understanding how price parities between domestic products and imports could potentially provide valuable information on future trade restrictions. For example, U.S. trade commissions can take price parities as a factor to determine the appropriate threshold when placing trade duties on Canadian lumber imports. This can lead to a mutually beneficial stage for all participants in the trade dispute, therefore, producing a more efficient economic outcome.

2.7 Tables and Figures

Table 2.1: Summary statistics for lumber prices

		No. Obs.	Mean	St. Dev.	Min.	Max.
Long-Series[1]	US_SYP	1321	351.82	82.59	193	596
	CA_SPF	1321	353.29	70.09	145	601
Short-Series[2]	US_East	104	538.78	217.72	325	1165
	US_West	104	491.26	227.89	282	1120
	US_Central	104	496.79	233.69	290	1155
	Canada	104	627.2	273.64	308	1250

[1]The long series contains weekly lumber prices from January 1, 1995, to April 24, 2020.
[2]The short series contains weekly lumber prices from December 10, 2021, to November 30, 2023.

Table 2.2: Unit Root Test for Individual Price Series

	Level Data			First-Differenced Data		
	ADF (No Drift)[1]	ADF (With Drift)	KPSS Level[2]	ADF (No Drift)	ADF (With Drift)	KPSS Level
Long-Series (N=1321)						
US-SYP	-0.256 [0.594][3]	-3.490 [0.008]	6.317 [0.010][4]	-19.54 [0.000]	–	0.022 [0.100][5]
CA-SPF	-0.106 [0.647]	-3.438 [0.010]	1.137 [0.010]	-14.41 [0.000]	–	0.021 [0.100]
Short-Series (N=104)						
US-East	-1.736 [0.078]	-1.693 [0.432]	1.478 [0.010]	-4.128 [0.000]	-4.223 [0.001]	0.097 [0.100]
US-West	-1.266 [0.188]	-1.783 [0.387]	1.423 [0.010]	-4.140 [0.000]	-4.254 [0.001]	0.098 [0.100]
US-Central	-1.355 [0.162]	-1.772 [0.392]	1.375 [0.010]	-4.517 [0.000]	-4.636 [0.000]	0.092 [0.100]
Canada	-1.416 [0.145]	-0.827 [0.806]	1.992 [0.010]	-5.504 [0.000]	-5.704 [0.000]	0.115 [0.100]

[1]Null hypothesis of the ADF test is the series contains a unit root. vs. Alternative: The series is stationary.
[2]Null hypothesis of the KPSS test is: The series is Level (Trend) stationary. Vs. Alternative: The series contains a unit root.
[3]Brackets contains p-values.
[4]Computed statistics are outside the table of critical values listed in Kwiatkowski (1992), therefore, actual p-values are less than 1% for the level data and greater than 10% for differenced data.
[5]For level-stationary, the critical value at 10% is 0.347, and the critical value at 1% is 0.739. For trend-stationary, the critical value at 10% is 0.119, and the critical value at 1% is 0.216.

Table 2.3: Unit-Root Tests for Price Differentials

	Level Data			First-Differenced Data		
	ADF (No Drift)	ADF (With Drift)	KPSS Level	ADF (No Drift)	ADF (With Drift)	KPSS Level
Long-Series (N=1321)						
Canada-US	-4.213 [0.000]	-4.229 [0.001]	6.671 [0.010]	–	–	0.011 [0.100]
Short-Series (N=104)						
Canada-US$_{East}$	-1.363 [0.159]	-1.513 [0.523]	1.211 [0.010]	-5.414 [0.000]	-5.425 [0.000]	0.070 [0.100]
Canada-US$_{West}$	-1.119 [0.238]	-1.696 [0.430]	0.908 [0.010]	-5.161 [0.000]	-5.166 [0.000]	0.080 [0.100]
Canada-US$_{Central}$	-1.136 [0.231]	-1.612 [0.472]	0.889 [0.010]	-5.366 [0.000]	-5.367 [0.000]	0.088 [0.100]

Brackets contains p-values.

Table 2.4: Linear OLS Estimates of Long-run Relationships

Price Pairs	α	β	*Adjusted* R^2
CA-US (N=1321)	2.529*** (0.104)[1]	0.568*** (0.018)	0.437
CA-US$_{East}$ (N=104)	-0.593 (0.353)	1.116*** (0.057)	0.790
CA-US$_{West}$ (N=104)	0.311 (0.329)	0.988*** (0.054)	0.766
CA-US$_{Central}$ (N=104)	0.499 (0.344)	0.955*** (0.056)	0.738

[1]Parenthesis contains standard errors.
***Asterisk: Coefficient is significant at 1%.

Table 2.5: Cointegration Test Results

Price Pairs	Test	Test Statistics
CA-US (N=1321)	ADF test of error term	-4.425*** [0.000][1]
	Johansen's trace test (r=0)	41.560***
	Johansen's trace test (r<=1)	16.630***
	Hansen and Seo's test[2]	35.440** [0.024]
CA-US$_{East}$ (N=104)	ADF test of error term	-1.764* [0.074]
	Johansen's trace test (r=0)	15.820*
	Johansen's trace test (r<=1)	2.910
	Hansen and Seo's test	18.294 [0.322]
CA-US$_{West}$ (N=104)	ADF test of error term	-1.668* [0.090]
	Johansen's trace test (r=0)	14.290
	Hansen and Seo's test	11.541 [0.603]
CA-US$_{Central}$ (N=104)	ADF test of error term	-1.523 [0.119]
	Johansen's trace test (r=0)	11.560
	Hansen and Seo's test	13.837 [0.324]

[1]Brackets contains p-values.
[2]Hansen and Seo's (2002) test of linear cointegration vs. threshold cointegration. P-values are calculated from 1000 bootstrap replications.
***, **, *Asterisk: Coefficient is significant at 1%, 5% and 10%.

Table 2.6: Linearity Tests of Price Differentials

Price Differentials	White[1]		Keenan[2]		Tsay[3]		Chan[4]	
	Stat	*P-val*	*Stat*	*P-val*	*Stat*	*P-val*	*Stat*	*P-val*
CA-US (N=1321)	2.194	0.334	1.699	0.193	6.624	0.000	28.165	0.001
CA-US$_{East}$ (N=104)	0.080	0.961	0.029	0.864	0.513	0.798	7.554	0.332
CA-US$_{West}$ (N=104)	0.661	0.719	0.001	0.974	2.614	0.022	5.014	0.238
CA-US$_{Central}$ (N=104)	0.351	0.839	0.051	0.819	0.517	0.753	5.672	0.306

[1]White's Neural Network Test. N_0: The time series is linear in mean.
[2]Keenan's One-Degree Test. N_0: The time series follows some AR process.
[3]Tsay's Nonlinearity Test. N_0: The true model is an AR process.
[4]Chan's Likelihood Ratio Test. N_0: The time series follows a normal AR process.

Table 2.7: Hansen's Test for Linearity Against Threshold

Price Differentials	AR vs SETAR(1)		AR vs SETAR(2)		SETAR(1) vs SETAR(2)	
	Stat	*P-val*[2]	*Stat*	*P-val*	*Stat*	*P-val*
CA-US (N=1321)	37.867	0.000	55.219	0.000	16.867	0.124
CA-US$_{East}$ (N=104)	7.554	0.814	22.815	0.537	–	–
CA-US$_{West}$ (N=104)	5.782	0.950	16.644	0.873	–	–
CA-US$_{Central}$ (N=104)	9.025	0.827	39.996	0.134	–	–

[1]If the null hypothesis of a linear AR model is rejected, then proceed to test the number of thresholds.
[2]P-values are calculated from 1000 bootstrap replications.

Table 2.8: Model Specification: Information Criteria

	AIC				BIC			
	Lag = 1	Lag = 2	Lag = 3	Lag = 4	Lag = 1	Lag = 2	Lag = 3	Lag = 4
CA-US (N=1321)								
AR	-9028.07	-9506.29	-9542.81	-9553.91	-9017.70	-9490.73	-9522.07	-9527.98
SETAR	-9026.11	-9531.33	-9573.69	-9588.27	-9000.18	-9495.03	-9527.01	-9531.22
VECM	-19998.36	-20343.76	-20334.59	-20320.54	-19951.69	-20276.37	-20246.47	-20211.71
TVECM	-20026.83	-20376.60	-20379.80	-20369.36	-19938.69	-20247.00	-20208.76	-20156.89
CA-US$_{East}$ (N=104)								
AR	-653.49	-693.41	-706.66	-708.08	-648.20	-685.48	-696.08	-694.86
SETAR	-649.56	-692.01	-704.16	-706.97	-636.33	-673.50	-680.36	-677.88
VECM	-1423.69	-1436.60	-1416.86	-1398.56	-1400.06	-1402.60	-1372.58	-1344.07
TVECM	-1424.32	-1450.13	-1436.62	-1408.39	-1379.70	-1384.75	-1350.65	-1301.99
CA-US$_{West}$ (N=104)								
AR	-655.68	-701.09	-707.09	-707.57	-650.39	-693.16	-696.99	-694.25
SETAR	-652.56	-698.24	-703.89	-702.85	-639.34	-679.73	-680.09	-673.76
VECM	-1436.06	-1440.58	-1420.89	-1408.42	-1412.43	-1406.59	-1376.60	-1353.92
TVECM	-1437.29	-1440.50	-1424.49	-1414.72	-1392.67	-1375.12	-1338.52	-1308.32
CA-US$_{Central}$ (N=104)								
AR	-646.00	-686.05	-688.82	-692.42	-640.72	-678.11	-678.24	-679.20
SETAR	-642.38	-683.51	-686.80	-689.41	-629.16	-665.00	-663.00	-660.32
VECM	-1400.22	-1405.56	-1391.66	-1375.40	-1376.60	-1371.56	-1347.38	-1320.90
TVECM	-1401.41	-1399.19	-1399.30	-1383.07	-1356.78	-1333.81	-1313.33	-1276.67

Table 2.9: SETAR[1] and AR Parameter Estimates

Price Differentials	Model Specification	Intercept	t-1	t-2	t-3	t-4
Canada-US (N=1321)	Low Regime	-0.028	1.112***	-0.312**	0.125	-0.028
	(No. Obs: 5.85%)	(0.170)	(0.087)	(0.126)	(0.122)	(0.073)
	High Regime	0.000	1.656***	-0.786***	-0.012	0.124***
	(No. Obs: 94.15%)	(0.001)	(0.029)	(0.056)	(0.056)	(0.029)
CA-US$_{East}$ (N=104)	Linear AR	0.003	1.711***	-1.030***	0.293***	–
		(0.004)	(0.093)	(0.161)	(0.093)	
CA-US$_{West}$ (N=104)	Linear AR	0.006	1.689***	-0.909***	0.192**	–
		(0.005)	(0.096)	(0.169)	(0.096)	
CA-US$_{Central}$ (N=104)	Linear AR	0.001	1.668***	-0.989***	0.443**	-0.154
		(0.005)	(0.098)	(0.188)	(0.189)	(0.099)

[1]The selected threshold value is -0.263.
Parenthesis contains standard errors.
***, **, *Asterisk: Coefficient is significant at 1%, 5% and 10%.

Table 2.10: TVECM[1] and VECM Parameter Estimates

Price Pairs		Equation	ECT	Intercept	SPF_{t-1}	SYP_{t-1}	SPF_{t-2}	SYP_{t-2}	SPF_{t-3}	SYP_{t-3}
Canada-US (N=1321)	Low Regime (No. Obs. 6.1%)	SPF	-0.113***	-0.041**	0.217***	0.259**	-0.093	0.228	-0.134*	-0.322***
			(0.001)	(0.012)	(0.005)	(0.046)	(0.245)	(0.112)	(0.075)	(0.001)
		SYP	-0.043	-0.022	0.120*	0.868***	-0.011	-0.278**	-0.010	-0.006
			(0.237)	(0.102)	(0.061)	(0.000)	(0.872)	(0.02)	(0.877)	(0.942)
	High Regime (No. Obs. 93.9%)	SPF	-0.007	-0.000	0.618 ***	0.074**	-0.064*	-0.024	-0.059	-0.006
			(0.088)	(0.954)	(0.000)	(0.042)	(0.074)	(0.629)	(0.064)	(0.877)
		SYP	0.009***	0.000	0.149***	0.990***	-0.097***	-0.345***	0.005	-0.108***
			(0.007)	(0.894)	(0.000)	(0.000)	(0.001)	(0.000)	(0.844)	(0.000)
$CA\text{-}US_{East}$ (N=104)		SPF	-0.024	-0.004	0.772***	0.312***	-0.330***	-0.166	–	–
			(0.016)	(0.003)	(0.097)	(0.113)	(0.096)	(0.110)		
		SYP_E	0.007	-0.002	0.100	1.035***	0.011	-0.391***	–	–
			(0.014)	(0.003)	(0.086)	(0.100)	(0.085)	(0.097)		
$CA\text{-}US_{West}$ (N=104)		SPF	-0.024	-0.005	0.604***	0.0019	–	–	–	–
			(0.017)	(0.003)	(0.090)	(0.0841)				
		SYP_W	0.018	-0.002	0.136*	0.700***	–	–	–	–
			(0.014)	(0.003)	(0.074)	(0.070)				
$CA\text{-}US_{Central}$ (N=104)		SPF	-0.020	-0.005	0.536***	0.089***	–	–	–	–
			(0.016)	(0.003)	(0.089)	(0.078)				
		SYP_C	0.020	-0.003	0.061	0.713***	–	–	–	–
			(0.015)	(0.003)	(0.087)	(0.076)				

[1]The selected threshold value is -0.284, the estimated cointegration vector is (1, -1.004)
Parenthesis contains standard errors.
***, **, *Asterisk: Coefficient is significant at 1%, 5% and 10%.

Figure 2.1: Regional Newly Privately-Owned Housing Starts: Total Units

This figure plots the total number of new privately owned housing units started in the Midwest, Northeast, South, and West regions of the U.S. Data is measured in thousands of units with no seasonal adjustments. Data is collected by the U.S. Census Bureau and the U.S. Department of Housing and Urban Development and retrieved from the Federal Reserve Bank of St. Louis.

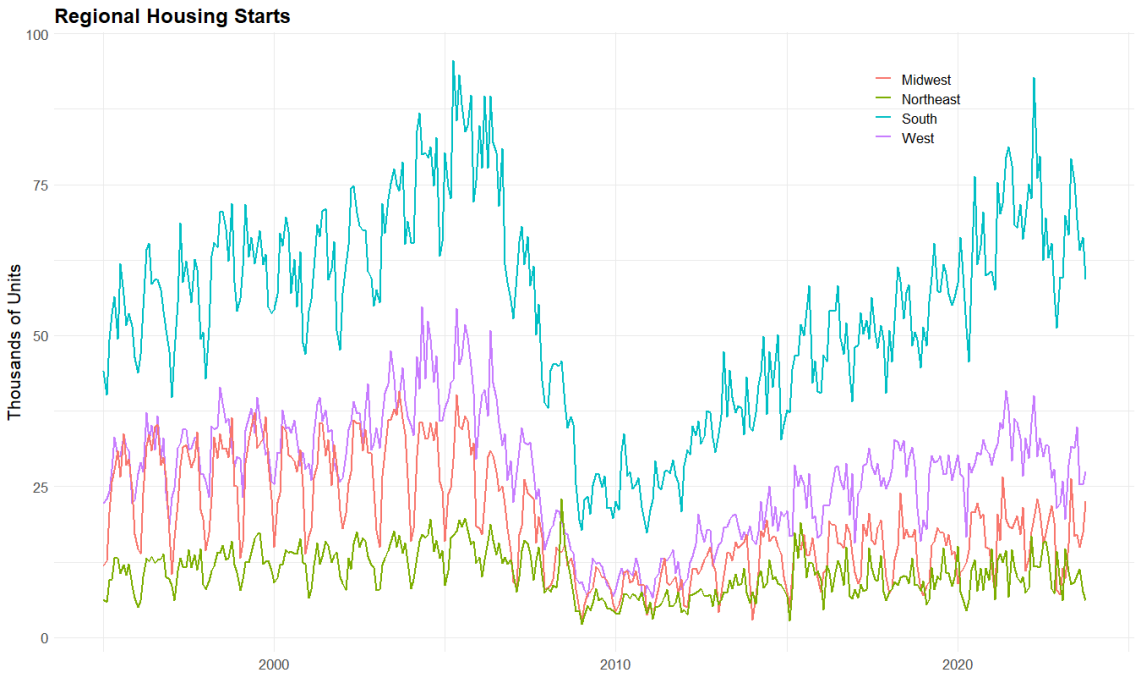


Figure 2.2: Weekly Price Movements and Price Differentials of Canada SPF and U.S. SYP January 6, 1995, to April 24, 2020 (No. Observation: 1321)

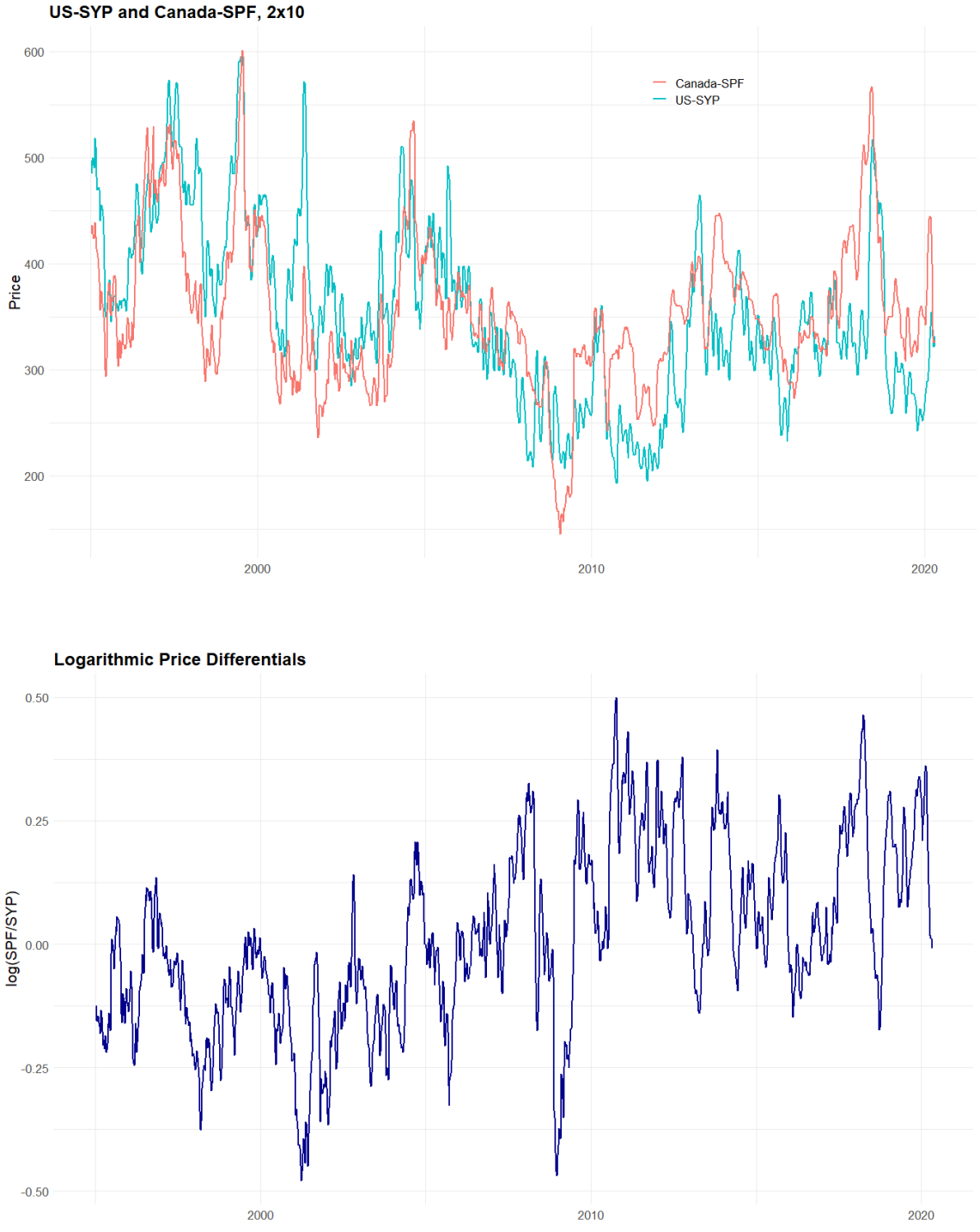
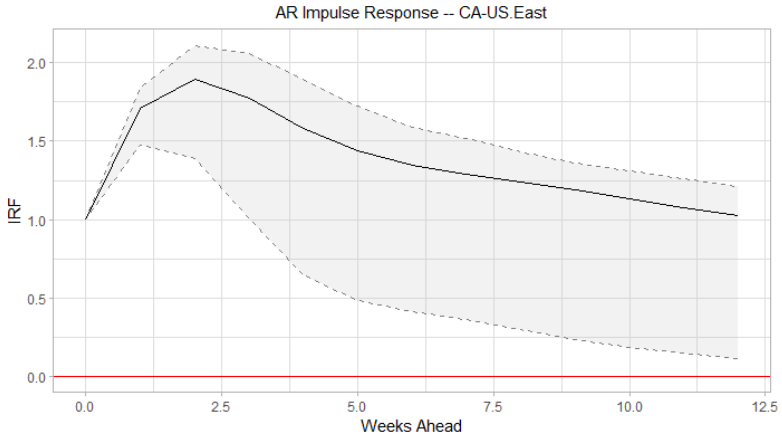


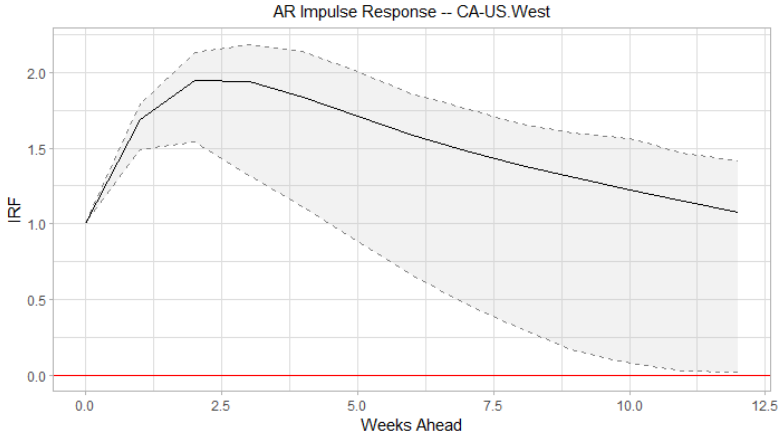
Figure 2.3: Regional Weekly Price Movements and Price Differentials of Canada SPF and U.S. SYP December 10, 2021, to November 30, 2023 (No. Observation: 104)

Figure 2.4: Impulse Response of AR Models

(a) Canada – U.S. East



(b) Canada – U.S. West



(c) Canada – U.S. Central

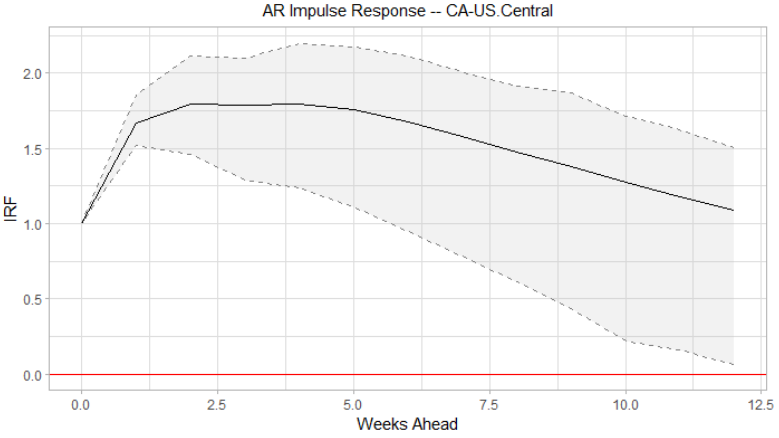
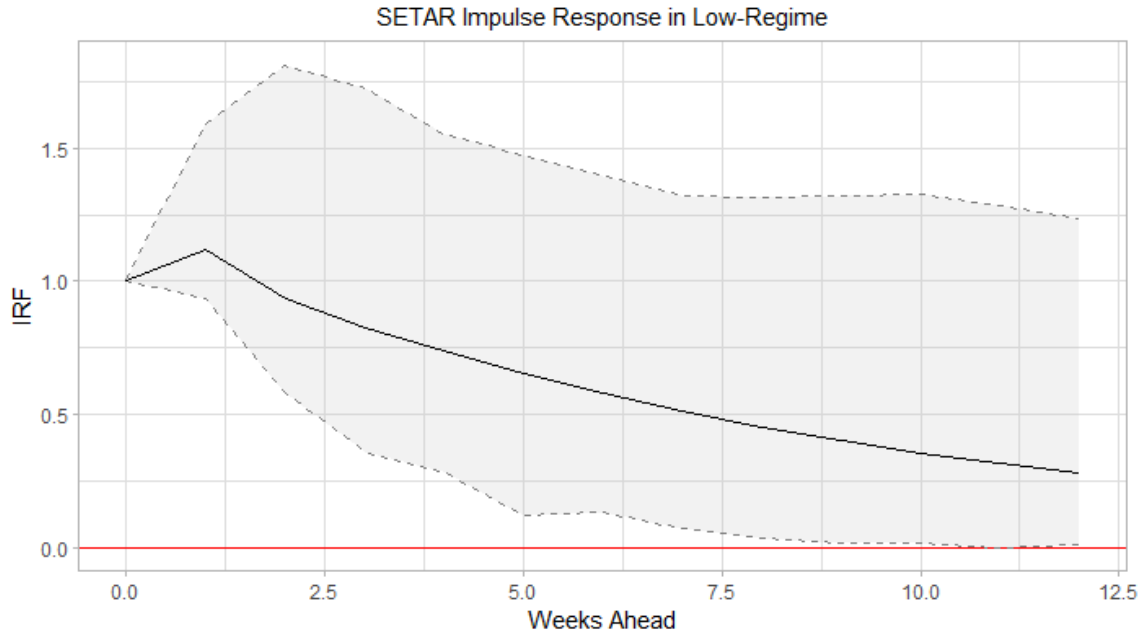


Figure 2.5: Regime-Specific Orthogonalized Impulse Response of SETAR Model



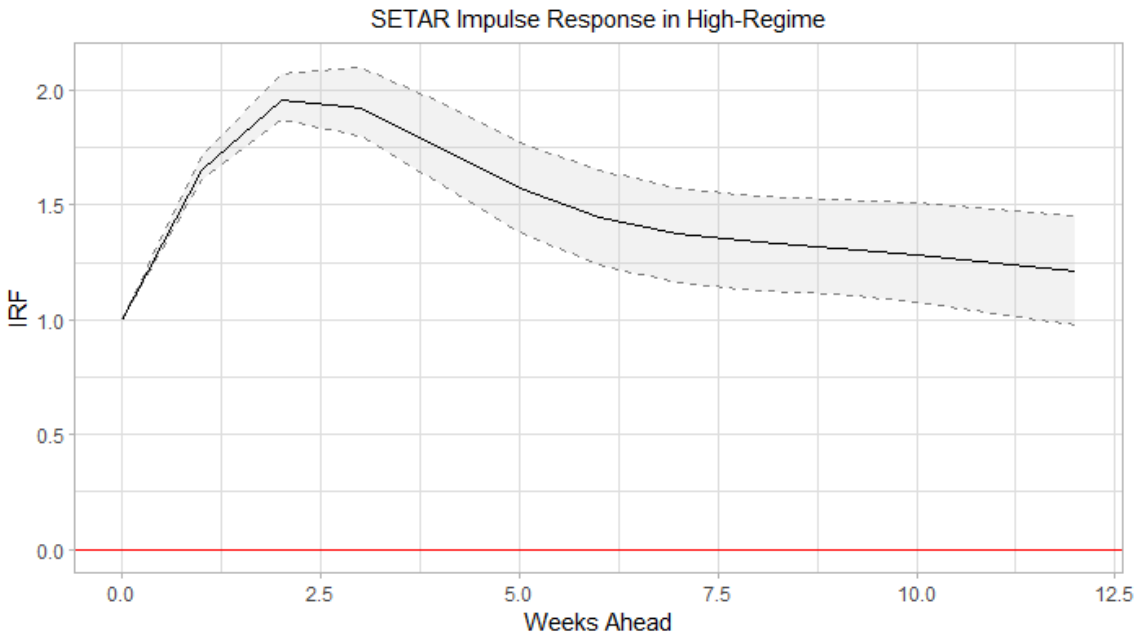


Figure 2.6: Generalized Impulse Response Density Plots of SETAR Model

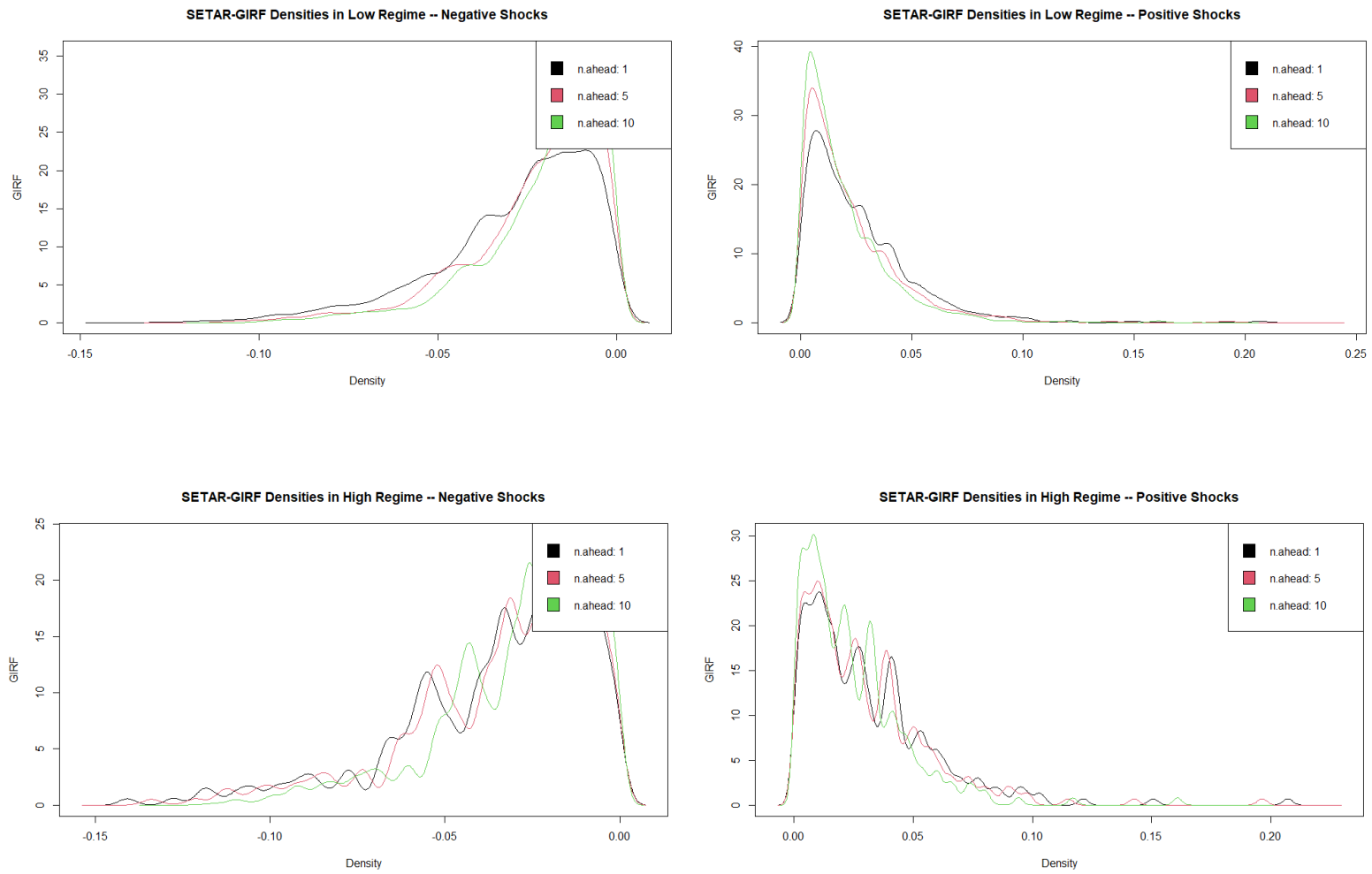
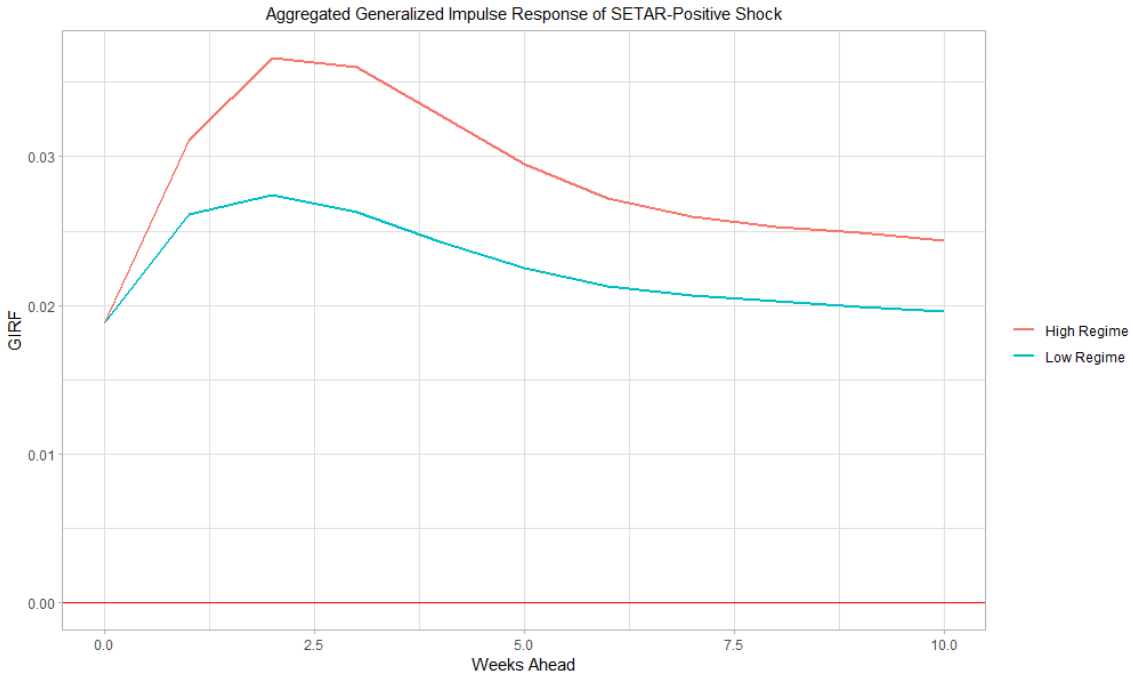


Figure 2.7: Aggregated Generalized Impulse Response of SETAR Model

(a) Positive Shock



(b) Negative Shock

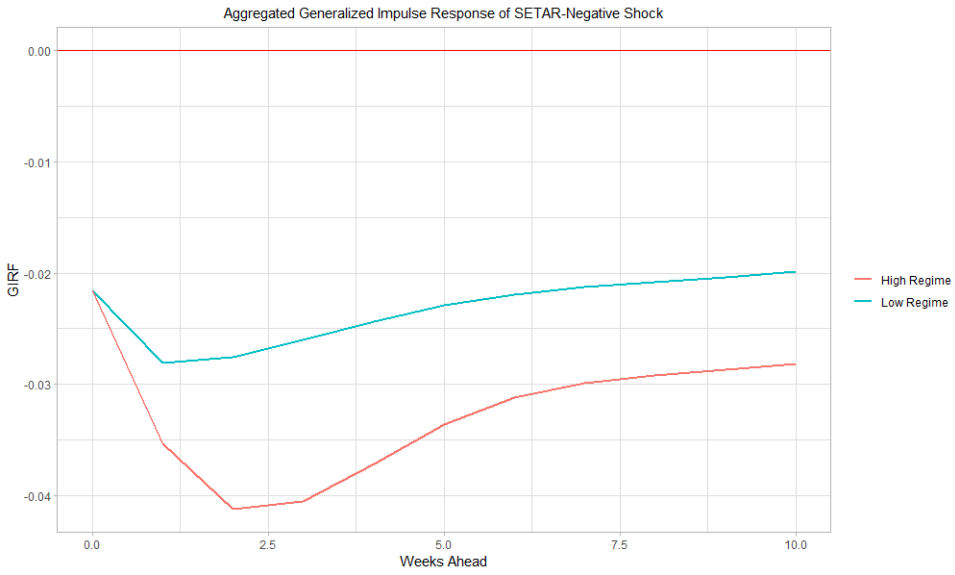


Figure 2.8: Impulse Response of VEC Model – East Region

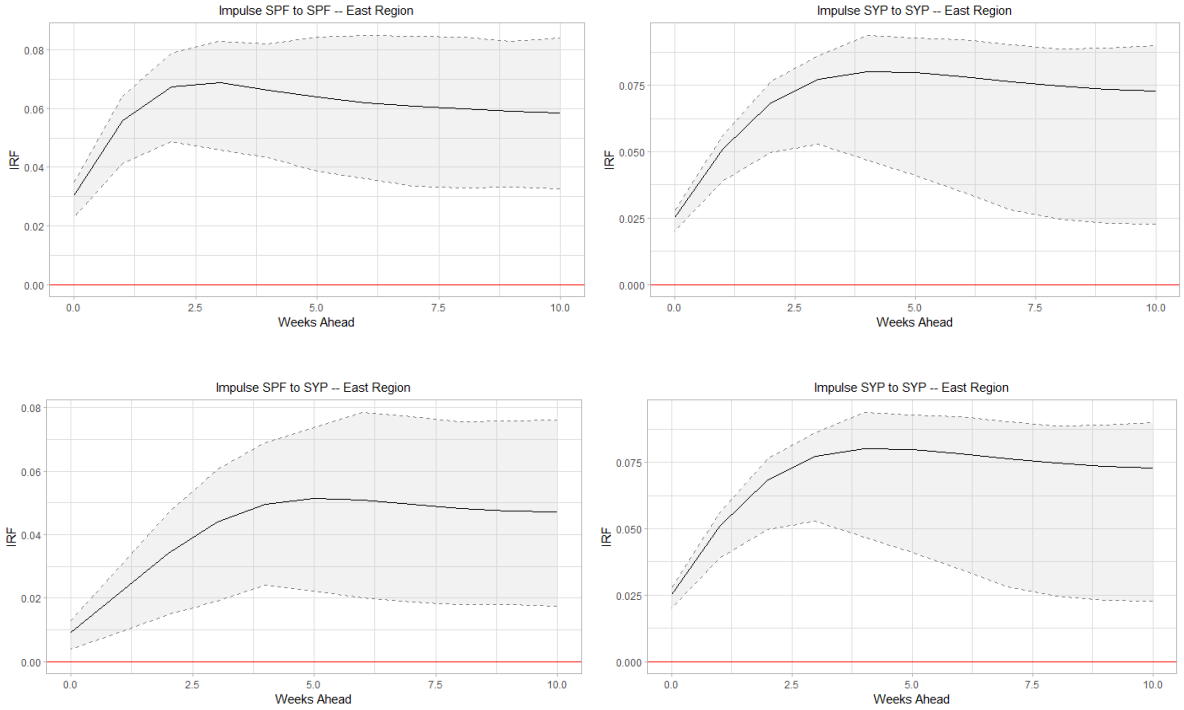


Figure 2.9: Impulse Response of VEC Model – West Region

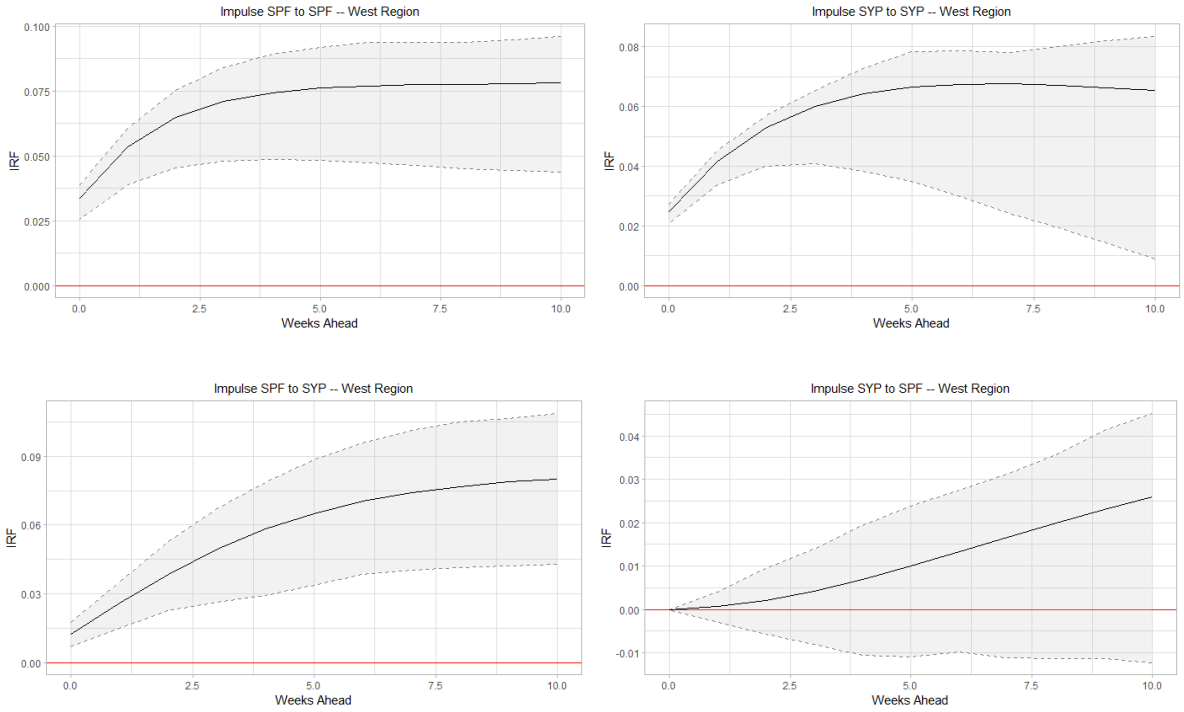


Figure 2.10: Impulse Response of VEC Model – Central Region

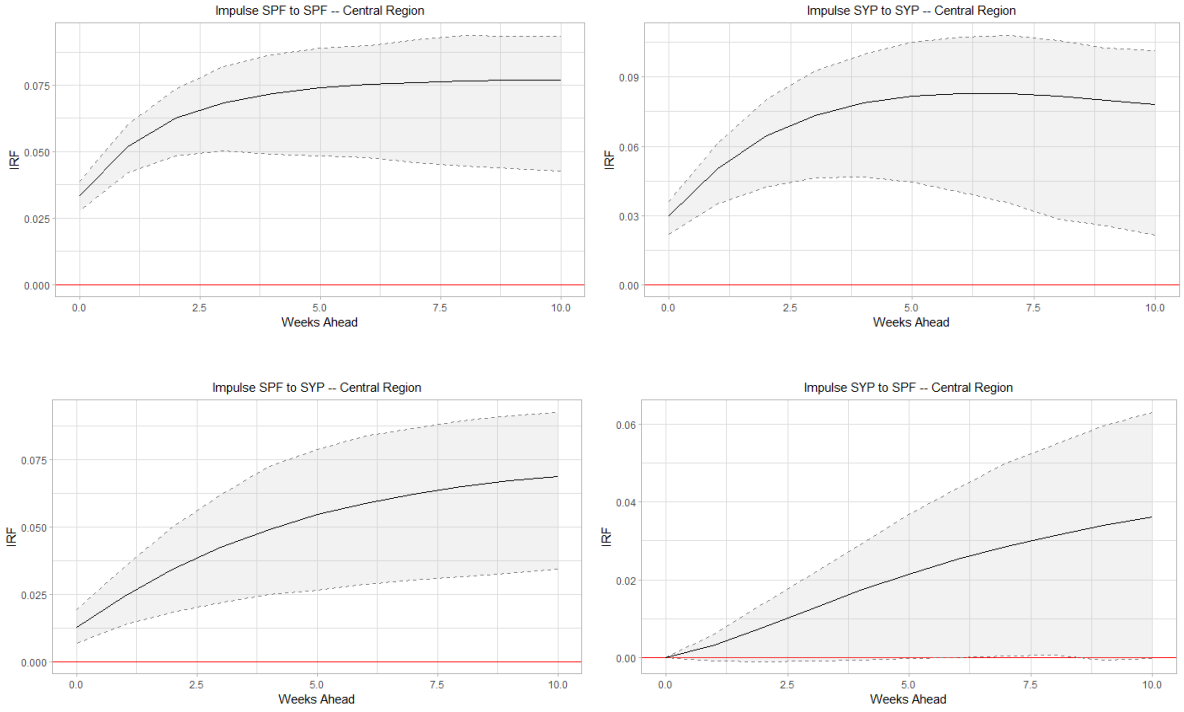


Figure 2.11: Regime Specific Orthogonalized Impulse Response of TVECM Model

This figure presents the regime-specific orthogonalized impulse response of TVECM. Left panels under (a) present impulse responses in the low regime, and right panels under (b) present impulse responses in the high regime. IRFs are generated based on a TVEC(3) model. The data frequency is weekly. All variables are transformed into natural logs. The shaded areas represent 95% bootstrapped error band.

(a) Low Regime

(b) High Regime

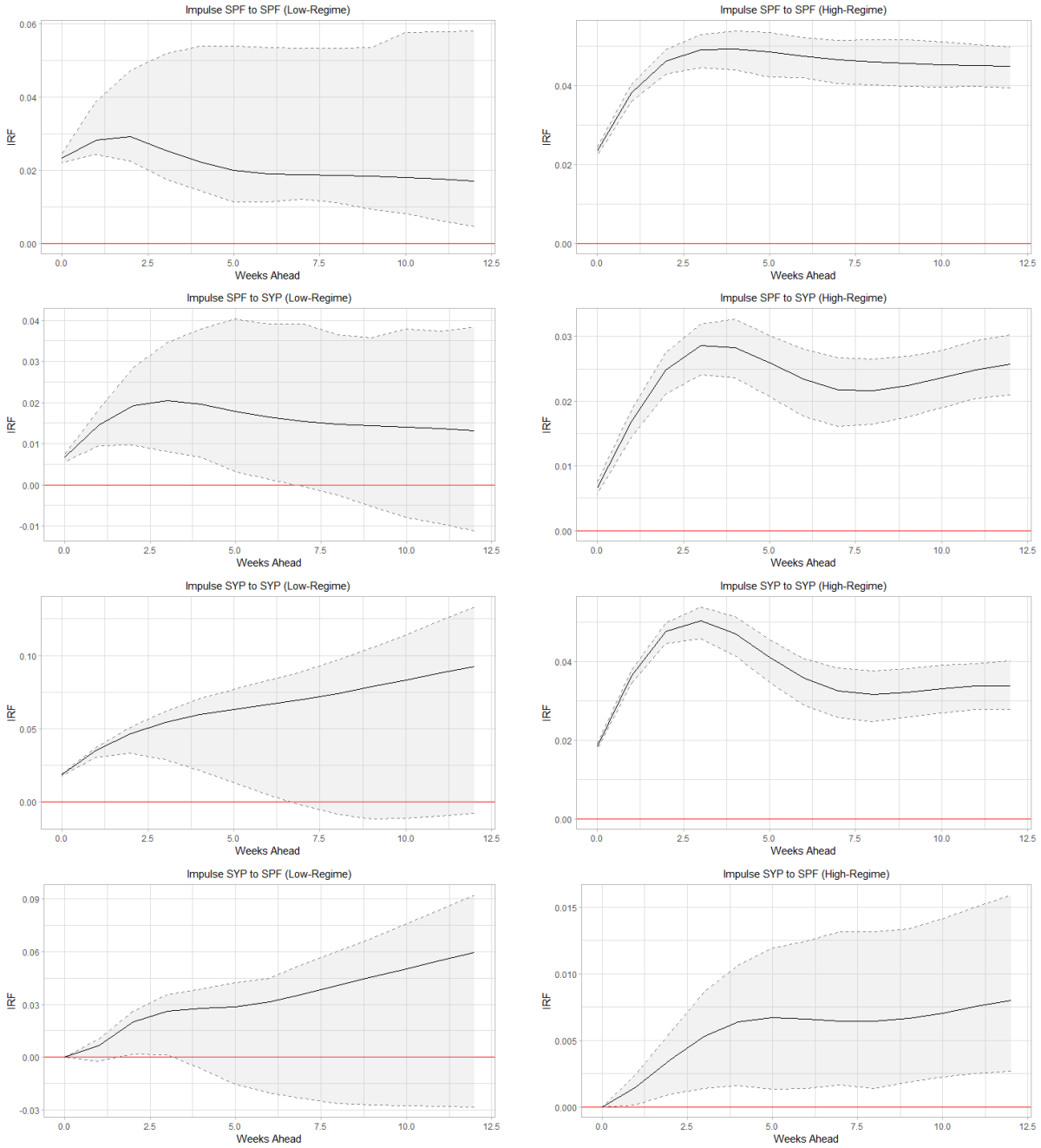


Figure 2.12: Generalized Impulse Response Density Plots of TVEC Model

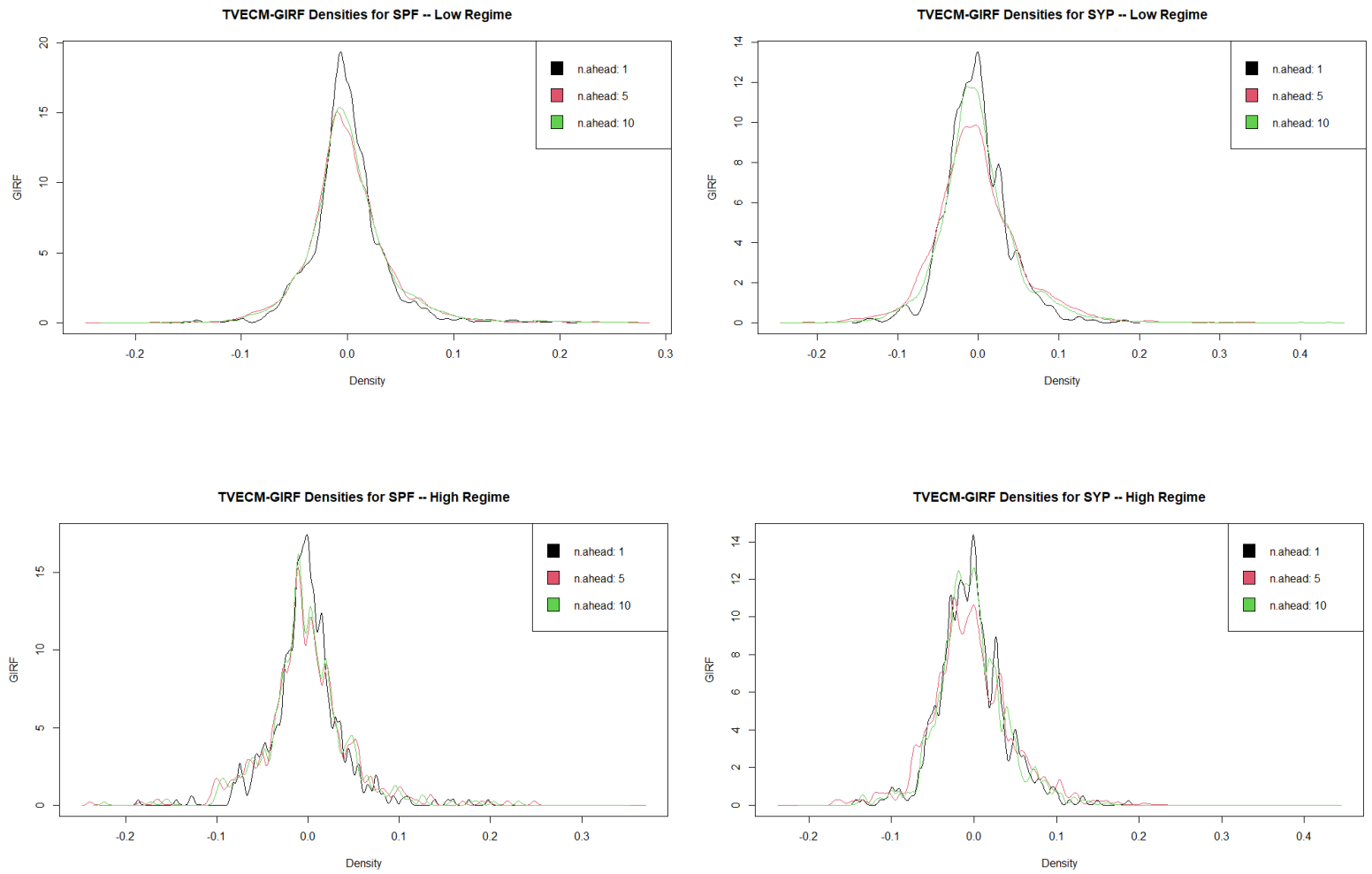


Figure 2.13: Aggregated Generalized Impulse Response of TVEC Model

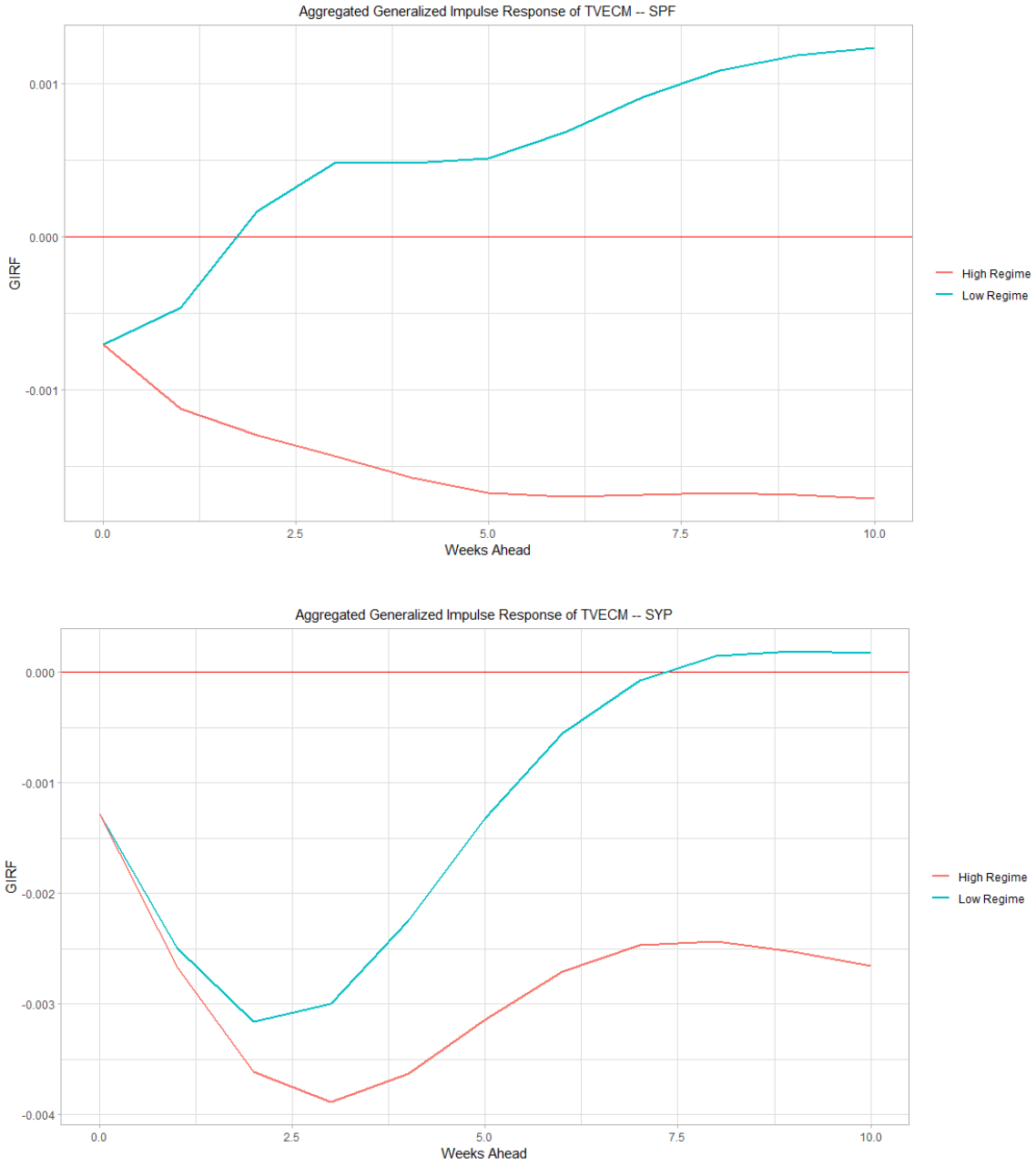
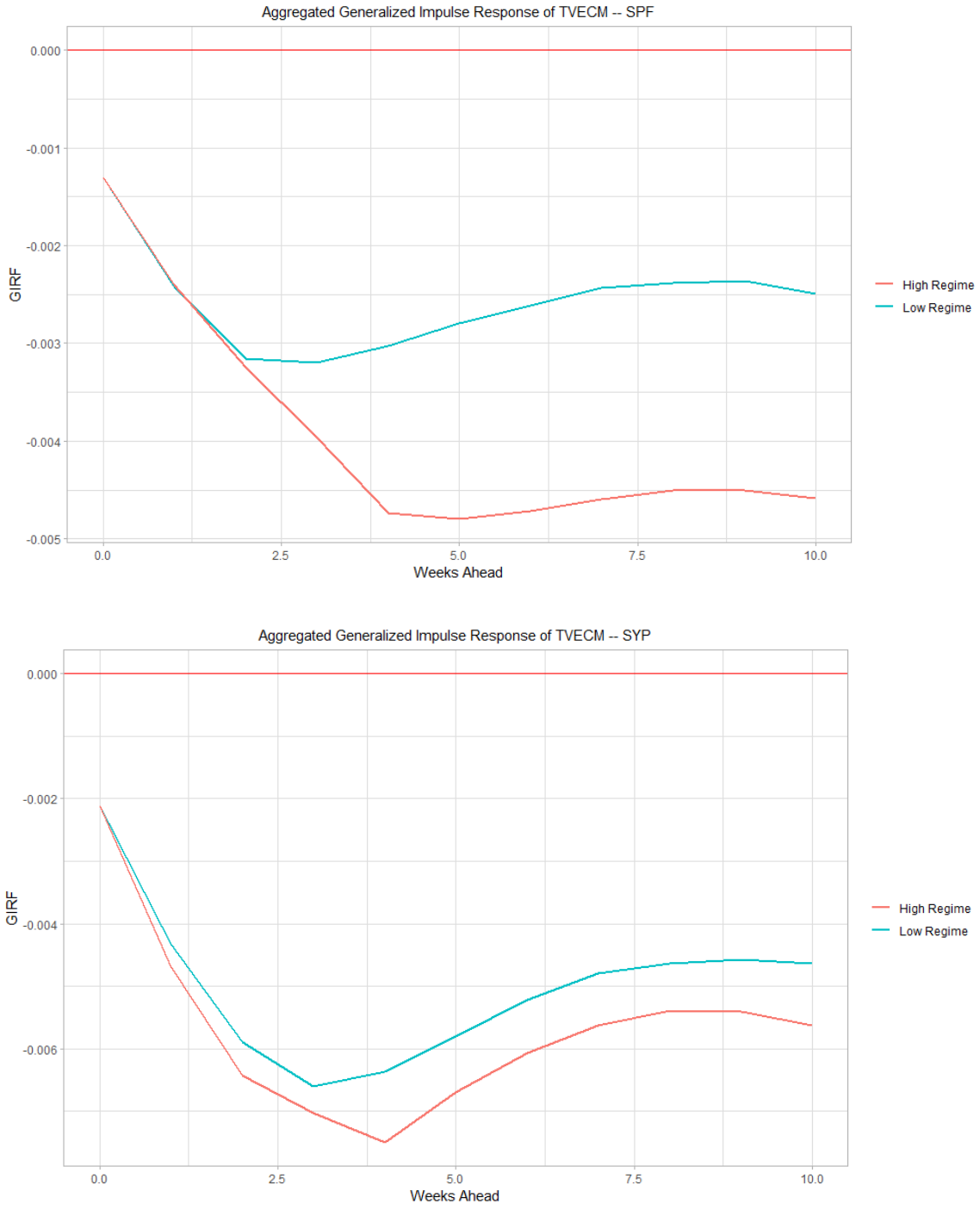


Figure 2.14: Aggregated Generalized Impulse Response of TVEC Model (Averaging Over Medain)



CHAPTER

3

IMPACTS OF COVID-19 ON U.S. MANUFACTURED WOOD INDUSTRY: EVIDENCE FROM SOFTWOOD LUMBER AND STRUCTURAL PANELS

3.1 Introduction

The Coronavirus-2019 (COVID-19) pandemic, caused by the SARS-CoV-2 virus (CDC), has profoundly impacted the global economy. COVID-19 is highly contagious, spreading through small liquid particles from one infected person to another. While most infected individuals experience mild respiratory symptoms such as coughing and fever, some may progress to pneumonia and become seriously ill. Therefore, the most effective way to prevent or slow down the spread of the virus is to limit social interactions. In March 2020, Tedros Adhanom Ghebreyesus, Director-General of the World Health Organization (WHO 2020), declared COVID-19 a worldwide pandemic. President Donald Trump quickly fol-

lowed suit, declaring COVID-19 a national emergency for the United States (U.S.). Many states implemented "shelter in place" protocols, urging residents to stay indoors and prohibiting group gatherings. Consequently, numerous ongoing projects were halted, putting significant pressure on various sectors, including the construction industry and related product manufacturing sectors.

Prices of commodities produced in the forest industry have experienced extreme fluctuations since the beginning of the pandemic. For instance, Spruce-Pine-fir (SPF) softwood lumber prices surged from $326 per thousand board-feet (mbf) in April 2020 to $1610/mbf in May 2021, reaching a historical high since 1995. The prices then stabilized at $350/mbf at the end of August 2021. Subsequently, in March 2022, SPF lumber prices experienced another whimsical ride, reaching a peak of $1250/mbf, then moderating to $348/mbf, returning to the pre-pandemic average price level in May 2023, three years after the pandemic. The price of wood-based panels, such as Oriented Strand Board (OSB), echoes a similar pattern. The significant price swings in softwood lumber and structural wood panels not only drive up the cost of new constructions and home renovations but, more importantly, create challenges in making production decisions and managing forest resources for companies involved in construction, woodworking, and forest industries.

Several channels of structural changes have occurred during the pandemic that can contribute to these large swings in wood product prices. First, the labor supply of unskilled and semi-skilled workers was significantly reduced due to the declaration of a national emergency and the implementation of many restrictive policies. Albanesi and Kim (2021) classify occupations based on their flexibility to work remotely and contact intensity. The occupation group that includes manufacturing, production, and transportation jobs falls under the inflexible/low-contact category and was estimated to experience a 30% decline in total employment during 2020. In addition, DeLuca and Pinheiro (2023) find uneven recovery across occupations and industries. In particular, they discover slower recovery in construction, production, and transportation occupation groups.

Taking a closer look at the employment status of these industries, the labor market disruption is reflected in the National Employment Statistics reported by the U.S. Bureau of Labor Statistics (BLS OEWS 2022). The total number of employees in the manufactured wood products industry decreased from 405.3 to 373.7 thousand one month after the lockdown, a 7.5% drop. A similar decline can be found in truck transportation, from 1492.2 to 1413.1 thousand (BLS 2020, 2021). The primary reason for this labor supply contraction is illness. According to estimates from the Survey of Occupational Injuries and Illnesses (SOII) conducted by the BLS, the total reported illness cases from the private industry

more than quadrupled in 2020 compared to 2019, increasing from 127,200 cases to 544,600 cases (BLS 2020). The significant rise in cases is mainly driven by a 4,000% increase in employer-reported respiratory illnesses, from 10,800 in 2019 to 428,700 in 2020 (BLS 2021). Although the total number of employees has returned to the pre-pandemic level a few months after the sharp drop, the lack of labor in transportation and wood manufacturing creates stagnation in the supply chain of forest products, leading to a shortage in the market and driving up prices.

In addition to labor market disruptions, a shift in consumers' housing preferences plays a significant role in explaining large fluctuations in wood product prices. More than a third of the U.S. annual consumption of softwood lumber and structural panels comes from new housing construction (Howard and Liang 2019). Therefore, the strong pandemic-induced housing demand would apply upward pressure on the demand for manufactured wood products. Economists have been studying the behaviors of the housing market during the pandemic. The increased share of jobs that allowed for working from home (Davis et al. 2021), and the "stay-at-home" shocks during the lockdown period changed household consumption patterns (Gamber et al. 2023), supporting the notion of an increase in demand for houses. It is worth mentioning private investment in housing spiked upward in 2020, evident in the private fixed investment estimated by the U.S. Bureau of Economic Analysis. Specifically, the private residential fixed investment (PRFI) was approximately $815 billion in the second quarter of 2020 and steadily climbed to a historic peak of $1218 billion in the first quarter of 2022. Although there has been a subsequent decrease, the current estimated investment stands at about $1076 billion, representing a substantial 32% increase compared to the second quarter of 2020. PRFI offers a clear perspective on the private sector's expenditure on fixed assets, serving as an indicator of the demand for housing.

Corresponding to the excess demand for houses at that time, there was a notable increase in the existing home median sales price starting in May 2020. Figure 3.1 plots the median sales price from September 2018 to July 2023 for both new construction homes and existing homes. The sharp increase in both price series indicates a clear disruption in the housing market during the first six months of the pandemic, although the new construction home market shows more volatility than the existing home market. In addition to the evidence of increasing housing demand, the National Association of Home Builders (NAHB) points out a growing demand to repair and remodel homes during the lockdown, further driving up wood product prices. Based on a quarterly survey of NAHB remodeler members, the Remodeling Market Index (RMI) in the first quarter of 2020 is 47 (NAHB 2023). This index jumped to 73 in the second quarter, indicating a strong increase in demand in

the remodeling market. Overall, disruptions in the housing market, arising from increased demand in both new construction homes and renovation projects, may partly explain the highly fluctuating wood product prices.

With ample resources, the wood product manufacturing sector is one of the most significant employers in U.S. manufacturing sectors. According to the National Industry-Specific Occupational Employment and Wage Estimates published by the BLS OEWS (2022), the wood product manufacturing sector employs about 424,000 people with an annual payroll income of approximately $19.2 billion, making this sector fundamental building block of the U.S. economy. In addition to being the world's leading producer of wood products, the U.S. is also the largest wood product consumer in the world. the U.S. total lumber consumption in 2017 is estimated to be 54.7 billion board feed, of which 42.2 billion board-feet (bf) is produced domestically, and the rest comes from imports (Howard and Liang 2019). With Canada being the largest lumber importer to the U.S.. In 2022, around 80% U.S. softwood lumber imports are from Canada, and the rest come from countries in Europe such as Germany and Sweden. As one of the most important players in the global forest product market, studying the U.S. manufactured wood industry not only provides valuable insight into the U.S. general economy but also the global market.

The objectives of this study are threefold. First, we examine the impact of the pandemic utilizing a Vector Error Correction (VEC) model on three selected manufactured wood products, Spruce-Pine-Fir (SPF), Southern Yellow Pine (SYP), and Oriented Strand Board (OSB). Figure 3.2 plots their monthly price movements from 1995 to 2023, with the beginning of the pandemic indicated by a vertical dotted line. While wood product prices have returned to their pre-pandemic levels, it is evident that manufactured wood product prices have been the most volatile during the pandemic in the past decades. Motivated by these abnormal behaviors, we are interested in recovering how much of the price volatility was introduced by the pandemic. To accomplish this, we separate our dataset into pre- and post-pandemic subsets and conduct an event study to measure the impact.

An event study is a statistical tool originally developed in finance and accounting research to examine changes in financial assets to a specific event (MacKinlay 1997). This method has since migrated to other research fields such as economics. The initial task of conducting an event study is to define the event of interest and the estimation window. Defining the start of the pandemic might not be universal since this worldwide pandemic appeared at different times for each country. Given our focus is on the U.S. market, we define the start of the event, i.e. the pandemic, as March 13, 2020. When the WHO declared COVID-19 a pandemic on March 11, 2020, the Trump Administration quickly responded

and announced the national emergency, along with lockdown policies and travel bands. Thus, information prior to this date is included in our estimation window, which is used to estimate the normal performance model. The parameters estimated from the model are then used to forecast the expected prices. We compare the model forecasted prices to the observed actual prices after the event occurs, any prices outside of the expected price bands are considered abnormal. Estimating the model proceeding to the event and comparing it to the outcomes of the event illustrate how prices deviate from modeled behavior, which allows us to assess the impact of the event on prices.

The second objective is to assess how exogenous shocks from other sectors contribute to the high volatility in wood product prices, under the framework of the COVID-19 pandemic. Characterizing price behaviors of manufactured wood products during the pandemic is a complicated task, as the wood product sector is impacted by multiple exogenous shocks. In addition, during the pandemic, there were significant policy interventions such as fiscal aid to households and increased softwood lumber tariff on Canada, which could cause unintended direct or indirect effects on the market. These policies could have increased the market turbulence, thus making modeling the impact of policy interventions on price transmission a challenge. We construct the VEC model including exogenous variables from multiple sectors that are closely related to forest product markets. This allows for a nice representation of the market conditions to disentangle the effects of information affecting wood product prices. We consider impulse response functions to indicate how individual wood product prices respond to sector-specific shocks.

The third objective of this study is to analyze the cross-commodity price transmission, which determines the price relationships across three manufactured wood products. Co-movements between commodity prices are mostly driven by changes in consumption behaviors, therefore, indicating substitutability and complementary relations among the products (Saadi 2011). SPF is the largest consumed lumber product in the U.S., although it is primarily imported from Canada, and SYP is the largest softwood species produced domestically in the U.S. (Nagubadi et al. 2004). The study of price relationships between wood product prices has important international trade policy implications between the U.S. and Canada. Given the large trade volume and value of softwood lumber imports coming from Canada, there has been a long and complex trade dispute between the two countries in the past decades. The U.S. lumber industry claims that Canadian lumber producers for selling below cost, therefore, imposing unfair competition to the market. Our analysis of cross-commodity price transmission provides insights into one important dimension of the trade dispute. More specifically, if the two lumber products are independent, then

trade restrictions implemented on Canadian softwood imports would be less effective in protecting the domestic lumber industry.

Several studies have conducted the analysis of the COVID-19 impact on forest products from various perspectives. For example, Zanello et al. (2023) study the price movement of lumber while evaluating multiple events such as the COVID-19 pandemic, the Great Financial Crisis of 2008, the Pacific Northwest floods in 2021, and softwood lumber agreements signed in 1981 and 2006. They find that the pandemic has a sizeable impact on the softwood lumber industry, although the magnitude of the impact is less than the impact of the flooding event, a significant supply chain disruption. van Kooten et al. (2022) view the pandemic impacts on the U.S. lumber market in terms of its welfare effects and estimate U.S. lumber producers gained $5.3 billion while U.S. consumers lost $7.3 billion per quarter. Gan et al. (2022) analyze the synchronized movements of U.S. lumber futures and southern pine saw-timber stumpage prices and their response to multiple COVID-19-related events and economic stimuli. They estimate the magnitude of cumulative abnormal returns separately using a market model and an autoregressive integrated moving average model and reveal different magnitudes of response to the events between two prices, indicating asymmetric price adjustment and benefit distributions among different agents. While these studies provide valuable information on the forest industry's response to the pandemic, not many studies analyze the industry in a multivariate setting.

Our study takes on a new perspective to quantify the pandemic's impact through a multivariate time series analysis. The results derived from our event study and impulse response analyses offer new insights into the effect of the pandemic effect on the manufactured wood industry. Although the market volatility induced by the pandemic has dissipated, our study shed light on how the expectations of wood product markets have shifted after experiencing extreme events such as the pandemic. We aim to provide valuable insights for researchers, policymakers, wood producers, contractors, stakeholders, and other market participants to have a better understanding of wood price behaviors and the impacts of subsequent shocks from other sectors.

The remaining sections are organized as follows. In Section 3.2, we provide a more detailed discussion of cross-commodity price transmission and its implication to the U.S.-Canada softwood lumber dispute. We then describe the econometric models used to accomplish the analysis in Section 3.3. We summarize the variables used in the models in Section 3.4. Section 3.5 presents the empirical analysis of the manufactured wood product price behaviors during the pandemic, including results of the event study, impulse response analysis, and price transmission across wood products. Finally, the last section concludes.

3.2 Cross-Commodity Price Transmission and Its Implication on U.S.-Canada Softwood Lumber Dispute

Price transmission has been a long-studied phenomenon in economics. The adjustment of price reflects the efficiency and resilience of a market, it is also a way to measure market responses to exogenous shocks. Since in many situations, it is difficult to observe the dynamics of market supply and demand, economists believe that price is an element that carries important market signals that is observable. The literature associated with price transmission can be categorized into two dimensions, vertically and horizontally. In general, vertical price linkage studies examine the price linkage of the same commodity across the supply chain. In the context of the forest industry, researchers have conducted numerous studies to examine price relationships among sawtimber, lumber, and paper markets (Ning and Sun 2014; Haynes 1977; Zhou and Buongiorno 2005).

Horizontal price linkage studies usually refer to how the price of the same commodity is transmitted in different locations at the same stage of the supply chain. How price shocks in one market are transmitted to another can be viewed as an indicator of whether efficient arbitrage exists between the markets, building on a fundamental economic concept, the Law of One Price (LOP). LOP holds if regional markets share a common, unique price, abstracting from transaction costs, and violations of LOP may indicate a less integrated market and unstable trading relationships (Fackler and Goodwin 2001). The spatial price linkage among various U.S. forest products is studied, such as softwood lumber (Sun and Ning 2014), OSB sheathing (Goodwin et al. 2011), southern pine stumpage (Prestemon 2003; Daniels 2011), and paper (Tang and Laaksonen-Craig 2007). On the contrary, while the theoretical framework and econometric implications are similar to spatial price transmission, horizontal price linkage sometimes is referred to as cross-commodity price transmission, which focuses on the price relationship between different commodities within the same market or the same level of the supply chain. This includes analyses of price co-movements, which implies the substitutability and complementary relations among commodities. Overall, the study of price transmission plays a crucial role in understanding market dynamics and informing economic policy.

This study evaluates the cross-commodity price transmission between Spruce-Pine-Fir (SPF), Southern Yellow Pine (SYP), and Oriented Strand Board (OSB) in the U.S. market. As one of the most used forest products in the U.S. construction industry, learning the price linkage among these manufactured wood products can provide insights into an important

dimension of forest product trade policies. Specifically, we ask the question of whether the market of imported softwood lumber is related to the market of domestically produced wood products. We include OSB sheathing, a wood structural panel product, to gain a more comprehensive view of the manufactured wood products market[1].

There has been a longstanding trade conflict between the U.S. and Canada over the importation of softwood lumber products since 1982. In Canada, most softwood lumber comes from publicly-owned forests, which are managed by provincial governments. This is opposite to the U.S., which has a higher share of privately owned forests. Due to this different regulatory framework, the U.S. lumber producers petitioned the U.S. Department of Commerce to set a tariff on Canadian softwood lumber imports as they think that the Canadian government unfairly subsidizes the lumber sector. the U.S. lumber industry claims that Canadian lumber producers for selling below cost, therefore, imposing unfair competition to the market.

Over the years, the U.S. government has imposed various trade restrictions on Canadian softwood lumber imports, including countervailing duties and anti-dumping duties, while two agreements were taken in place in the history of the dispute. The first Softwood Lumber Agreement (SLA) took place in 1996. Upon its expiration, the two countries signed another agreement in 2006, which expired in 2015. After one year of no anti-dumping investigations against Canadian softwood lumber products following the expiry of the SLA, the U.S. government started the imposition of duties on Canadian softwood lumber products in early 2017. Until the present day, there is no other agreement has taken place, leaving the U.S. government the freedom to control the tariff.

The U.S.-Canada softwood lumber dispute has significant economic implications for both countries. Since Canadian softwood lumber imports would be detrimental to the U.S. lumber market only if the two products are related. That is, if the two lumber products are not related, then any trade restrictions on Canadian softwood would not affect U.S. SYP markets. Several studies have analyzed the substitutability between the U.S. and Canadian softwood lumber through a demand system. For instance, Nagubadi et al. (2004) employ a restricted translog subcost function to softwood consumption and prices and find the SPF group appears to be substitutes for untreated SYP and structural panels. Similar results are also found by Gan (2006) that U.S. domestic and imported forest products have become

[1]Plywood sheathing is a structural wood-based panel that competes with OSB sheathing. We conducted an analysis to include plywood and found that its inclusion does not introduce additional dynamics to the model. More specifically, the impulse response analysis indicates that the other three manufactured wood products show insignificant responses to a shock from plywood. Therefore, plywood is omitted from the study.

more substitutable over time. Shook et al. (2007) estimate the cross-price elasticity of demand of the U.S. softwood lumber species and did not find strong evidence that Canadian softwood species behave as substitutes to the U.S. domestic species. Our study takes on a different approach by examining the cross-commodity price transmission, in the hope of providing more insights from this perspective.

3.3 Econometric Methods

As one of the most popular methods to use in multiple time series analysis, Vector Autoregressive (VAR) model offers a convenient way to compute the dynamic behaviors among multiple time series, yet the model is estimated with the assumption that the underlying time series data is stationary, which in many cases, does not represent the behavior of real-world data. If each component of the VAR system is integrated, the conventional method requires some type of data transformation to make the time series stationary. A common approach is to take differences to meet the stationarity requirement. Taking differences for each component individually may distort useful relationships between the variables and complicate interpretability, especially if a trend needs to be added back to the predicted values after differencing. One solution to retaining the cointegration relationships among variables is to rewrite the VAR representation including a cointegration restriction. This modified version of VAR is proposed by Engle and Granger (1987) who suggest vector error correction (VEC) models when cointegration is present. VEC models allow one to determine the long-run relationship through an error correction term while keeping the short-run relationship flexible.

We start our analysis by assessing the time series properties of each series included in the models, then fit models to capture how prices are transmitted. Stationarity and cointegration tests are standard tests to do before the estimation of time series econometric models for most price transmission analyses to ensure appropriate model selection. We apply the standard augmented Dickey-Fuller (ADF) test to test the presence of a unit root in a signal time series Dickey and Fuller (1979). Rejecting the null hypothesis infers that the series is generated by a stationary process. In practice, however, the standard ADF test is being criticized for lack of power when detecting a process that is close to having a unit root yet not a pure random walk. It can also suffer from size distortion which leads to gross over-rejection of the null hypothesis Schwartz and Wachter (2022). Understanding the limitation of the standard ADF test, we include test statistics from Phillips-Perron (PP)

unit root test that is valid under heteroscedasticity. The null hypothesis of the PP test is a random walk, the same as the ADF test. When implementing unit root test like the ADF test, one important issue is the specification of optimal lag. Inappropriate selection of lag length may bias the test. We first set an upper bound for the optimal lag length selection, determined using a rule of thumb suggested by Schwert (1989):

$$p_{max} = \left[12 \left(\frac{T}{100} \right)^{1/4} \right] \tag{3.1}$$

where T is the number of observations in the sample. Then the optimal lag length is selected according to the minimized Schwarz Information Criterion (SIC).

After confirming the nonstationary behaviors in the time series, we continue our analysis by checking for cointegration. The concept of cointegration was introduced by Engle and Granger in 1987. If two non-stationary time series are cointegrated, there exists a linear combination of the variables that is stationary. There are several ways to test for cointegration. Engle and Granger (1987) propose a two-step residual-based test procedure to check for cointegration between two time series. Consider two nonstationarity time series x_t and y_t that are integrated of order one, then there exists a linear long-run relationship such that

$$y_t = \alpha + \beta x_t + u_t \tag{3.2}$$

where u_t is referred to as the error correction term. The first step is to estimate Equation 3.2 and obtain the residuals. Then check the stationarity of the residuals to determine if the two series are cointegrated. Engle and Granger proposed ADF-type regression without an intercept on the residuals to test for unit root. If the time series is cointegrated, the Engle and Granger method will show the stationarity of the residuals. We apply the ADF test on the residuals to check for stationarity.

Although easy to implement, when dealing with a system that contains more than two endogenous variables, there might be more than one linear combination that is stationary. The number of cointegrating relations is referred to as the rank order, denoted as r. This issue cannot be addressed by Engle and Granger's two-step procedure. One method to determine the cointegrating rank order in a system with multiple non-stationary time series is Johansen (1995) maximum eigenvalue and trace test. Johansen's methodology starts in the VAR model of order p, which is given by

$$y_t = a + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \epsilon_t \tag{3.3}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})$ is a $K \times 1$ vector of endogenous variables, a is a $K \times 1$ vector of intercept terms, A_i are $K \times K$ coefficient matrices, and the error term $\epsilon_t' = (\epsilon_{1t}, \ldots, \epsilon_{Kt})$ is a K-dimensional white noise[2]. Similar to the procedure of stationarity tests, the lag length p may be chosen using a model selection criterion such as Akaike Information Criterion (AIC).

We can obtain the Vector Error Correction (VEC) representation from Equation 3.3 by subtracting y_{t-1} on both sides and rearranging terms, which gives us:[3]

$$\Delta y_t = a + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i y_{t-i} + u_t \tag{3.4}$$

where $\Pi = -(I_K - A_1 - \cdots - A_p)$ and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$, $i = 1, \ldots, p-1$. p is the model's lag length in its VAR form, thus the VEC representation takes the order of $p-1$. If $rk(\Pi) = r < K$, then the coefficient matrix Π can be decomposed as $\alpha\beta'$, where α and β are $K \times r$ matrices of rank r. As referred by Lütkepohl (2005), the matrix β is called a cointegration matrix that represents the long-run relationships among the endogenous variables, and matrix α is sometimes called the loading matrix, which describes the speed at which an endogenous variable converges back to the equilibrium values. Notice that the decomposition is not unique, therefore, researchers often propose restrictions on β and/or α to obtain unique cointegration relations[4].

If the VEC model is an adequate description of the data generation process, then it is common to apply Johansen's two likelihood ratio type tests to determine the number of cointegration ranks of the matrix Π. These tests are the maximum eigenvalue test and trace test, and their test statistics are defined as follows:

$$\lambda_{max} = -T\, log(1 - \lambda_{r+1}) \tag{3.5}$$

$$\lambda_{trace} = -T \sum_{i=r+1}^{k} log(1 - \lambda_i) \tag{3.6}$$

where T is the sample size, and λ_i are the estimated eigenvalues of matrix Π. We determine the cointegration rank by testing a sequence of null hypotheses that there are r cointegrating

[2] ϵ_t is white noise, that is, $E(\epsilon_t = 0), E(u_t u_t') = \Sigma_u$, and $E(u_t u_s') = 0$ for $s \neq t$. Σ_u is the covariance matrix.

[3] To see that VEC models are essentially VAR models with some cointegration restrictions, please refer to Appendix, where a simple bivariate model with two lags is presented to illustrate this point.

[4] More details on obtaining unique cointegration relations are discussed in the Appendix.

vectors, starting with $r = 0$,

$$H_0 : r\,k(\Pi) = r \text{ versus } H_1 : r\,k(\Pi) > r \tag{3.7}$$

If the test is rejected, then repeat the test with $r + 1$ until the test is not rejected, and the cointegration rank is chosen[5].

Once the time series properties are defined, we implement an extended approach of the standard vector autoregressive (VAR) model to characterize the joint dynamic behavior of a collection of variables. The basic form of the VAR system, therefore its VEC representation, describes the joint generation process of all variables. In practice, however, the generation process may be affected by observables that are determined outside of the system. Since wood product prices are affected by other sectors that are regarded as exogenous, we consider an extension of the basic form of the VAR system by adding a series of exogenous variables[6]. We also include seasonal dummies, which are classified as a group of non-stochastic exogenous variables, to account for possible seasonal effects. The implemented model is presented as follows,

$$\Delta y_t = a + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \sum_{i=1}^{s} B_i x_{t-s} + \epsilon_t \tag{3.8}$$

where Δy_t is the first difference of the variables in vector y_t, $\Pi = \alpha\beta'$ is the coefficient of cointegrating relationships, and Γ_i are the coefficient matrices of the lags of differenced variables of y_t. x_t is the vector of all deterministic terms and B_i are the corresponding coefficient matrices. s is the number of lags for exogenous variables[7]. Lastly, ϵ_t is the error term with mean zero and variance-covariance matrix Σ.

After model estimation, we conduct an event study to compare the forecast results from econometric models to the observed value to access abnormal behaviors in wood product prices. We evaluate the dynamic patterns of price adjustment to shocks from other variables using impulse response functions, which allow us to have a straightforward view

[5]Note that if cointegration exists, the maximum number of cointegration vectors available is $K-1$. If $r\,k(\Pi) = 0$, this implies no cointegration, the VEC model reduces to VAR(p-1) in the first differences that are stable.

[6]Note that it is possible that all exogenous variables and the deterministic seasonal dummy variable do not enter the cointegration relation relationships, as described in Lütkepohl (2005), Chapter8

[7]It is possible that $s = 0$, which is assumed for this study. It is appropriate to not include these variables in the cointegration relations as we assume the lag order of exogenous variables is 0. This implies that only the current value of the exogenous variables would affect the current endogenous variables. For a more detailed discussion, please see subsection 3.5.1 Model Specification.

of the magnitude and persistence of price response to a certain shock.

3.4 Data

We focus on three of the most used manufactured wood products in the U.S. construction industry, Spruce-Pine-Fir (SPF) softwood lumber, Southern Yellow Pine (SYP) softwood lumber, and oriented strand board (OSB) sheathing. SPF lumber is made up of spruce, pine, and fir tree species. The physical and mechanical characteristics of the wood of these three tree species are similar, and they are therefore often harvested and processed together. There are two species of SPF in North America, eastern and western. Eastern SPF species are sourced from Black spruce, Red spruce, White spruce, Jack pine, and Balsam fir trees. These tree species grow more slowly and yield wood with great strength properties. On the other hand, the western species is sourced from White spruce, Engelmann spruce, Lodgepole pine, and Alpine fir trees, these species are available in larger quantities due to their enormous heights, because of the favorable climate of the geographic region that these species grow in. In this project, we use the price of 2x10 kiln-dried Western SPF in our analysis[8]. SPF prices collected are end-of-the-week prices in Western Canada[9]. SPF lumber has many great properties such as a good strength-to-weight ratio, making it a desirable material to be used as a structural framing material in all types of building applications (AIFP Lumber 2020). SYP is used to describe native pine species whose major range is in the south of the U.S., including the four most common species that make up 90% of the southern pine timber industry, loblolly shortleaf, longleaf, and slash (AIFP Lumber 2019). As the largest softwood product produced in the U.S., SYP is mainly used for construction framing and decking. We choose end-of-the-week 2x10 kiln-dried SYP prices from the west side of the U.S. for the analysis[10]. It is worth noting that the lumber prices included in the study are for untreated lumber. Untreated lumber products are specifically used for indoor projects, such as residential construction framing, while treated lumbers are used in outdoor decking and other outdoor applications. OSB is a structural panel engineered by cutting strands of wood from small logs. OSB sheathing is primarily used as a "load-

[8]Western SPF is also the primary deliverable species of Chicago Mercantile Exchange Random Length Lumber Futures Contract.

[9]Item code: LCBT. Kiln-dried Western Spruce-Pine-Fir (Western Canada), grade 2&btr, 2x10, random lengths mill prices. Prices are for paper-wrapped stock, and shipments from the Prince George, B.C., area. Duties/export taxes included when applicable.

[10]Item code: LAES. Kiln-dried Southern Yellow Pine (Westside), grade 2, 2x10, random lengths, free-on-board mill prices. Mills in Tex., Ark., Okla., and La west of the Mississippi River.

bearing structure component", to cover roofs, walls, and floors. We selected 7/16-inch OSB sheathing prices in the Southwest region of the U.S. for analysis[11]. Forest product prices are collected by a privately owned price reporting agency Random Lengths based on industry feedback.

Besides the wood products, we include six exogenous variables to analyze the impact of shocks from other sectors. Since the market of manufactured wood products is closely tied to the construction industry in the U.S., we include new privately owned housing starts to represent changes in the housing market. As mentioned in the introduction section, a supply shortage in the wood product industry may be one of the main contributors to changes in their prices. Therefore, the wood-product manufacturing sector employment data is included to account for labor market disruptions. Housing starts are collected by the U.S. Census Bureau and the U.S. Department of Housing and Urban Development, and employment data is collected by the U.S. Bureau of Labor Statistics (BLS). Industrial Production Index (IPI) measures the level of production and capacity utilization rates in manufacturing, mining, and electric and gas utilities (Federal Reserve Board 2017). This index is included to account for changes in the overall economy. The IPI data is provided by the Board of Governors of the Federal Reserve System.

Copper and aluminum are two popular metals used not only in construction but also in other industries. Therefore, changes in their prices can potentially impact the prices of manufactured wood products, but not vice versa. Both metal prices used in the analysis are the three-month rolling forward contract prices from the London Metal Exchange (LME) collected from Bloomburg. On a related note, dynamics in the energy market can potentially cause spillover effects on wood product prices since the domestic supply relies heavily on trucks. There were large supply-side shocks that were undergone in the crude oil industry during the pandemic. The West Texas Intermediate (WTI) crude oil prices first plunged in Spring 2020 and hit negative territory for the first time in April 2020, then quickly rebound in the second half of the year. Therefore, it would be interesting to see if changes in crude oil prices have any significant impact on the selected wood products. We choose the WTI crude oil as our product choice since it serves as one of the main global oil benchmarks. The free-on-board (FOB) spot price is collected from the U.S. Energy Information Administration.

Each price series is adjusted for inflation using the Consumer Price Index (CPI) retrieved from BLS. For more information and data sources regarding the variables, please refer to

[11]Item code: PACS. Oriented Strand Board Sheathing (Southwest) 7/16-inch, free-on-board mill prices.

Table 3.1. All time series are collected every month from January 1995 to May 2023. One thing to note is that the wood product prices are updated weekly, while crude oil, aluminum, and copper prices are collected daily. We aggregate the frequency into monthly by taking the average per month, because the macro variables are only available in monthly frequency.

Table 3.2 presents the summary statistics before the pandemic and the event period. We define March 20, 2020, as the start of the event since this is when the lockdown policies are taken into effect in the U.S. Therefore, our pre-pandemic subset contains monthly data from January 1995 to March 2020, and the event period contains data from April 2020 to May 2023. It is not surprising to see the surge in the average wood prices for all three wood products during the event period. In particular, the increase in price of OSB sheathing is the most substantial at 170% increase on average, followed by SPF at 102%, then SYP at 67%. Another thing that is worth mentioning is the extreme increase in the standard deviation of wood product prices during the event period.

Copper and aluminum prices have also increased significantly, while the average price for crude oil has increased slightly. The summary statistics do not suggest an overall decrease in employment, in fact, the average employment was higher during the event period. Lastly, we observed higher numbers of residential housing started during the pandemic.

The plot of each time series is presented in Figure 3.2 and Figure 3.3. Prices of SPF lumber and OSB sheathing since 1995 are presented in Figure 3.2, note the substantial movements in prices that occurred after the pandemic. In fact, the prices have reached a historical high. In addition to the price surge and large movements, we observe similar movement patterns in the two wood products. This co-movement is more pronounced after the pandemic. This suggests that there is a long-run relationship between the lumber and sheathing prices, i.e. cointegration. Plots in Figure 3.3 suggest strong seasonal patterns in some processes such as housing starts and employment, which implies the need to address seasonality in the model in order to reveal nonseasonal patterns that are interesting to the analysis.

3.5 Empirical Analysis

3.5.1 Model Specification

Our objective is to choose a parsimonious model that accurately characterizes the system's dynamics. The model selection process is based on various key aspects. We first consider the time series proprieties of the variables to check for stationarity and cointegration. Prior

to implementing unit root tests, we determine the maximum number of lags to include for the optimal lag selection. Given our sample of the pre-pandemic subset contains 303 observations, the maximum lag length is 16 based on Equation 3.1. Stationarity test results are present in Table 3.3. Given the test statistics, we conclude that all time series indicate the presence of a unit root. The test results are consistent in all unit root tests[12], where we fail to reject the null hypothesis that the series contains a unit root. Each series becomes stationary after first differencing, implying all series are integrated of order one. We then check for the cointegration relationship between the three endogenous variables using Engle and Granger's two-step procedure and Johansen's trace test. Note that the Johansen procedure only tests the existence of a cointegration relation between two variables, therefore, the test is applied to three possible pairs of wood products. Test statistics are presented in Table 3.4, where we observe that the ADF test is rejected in all cases, indicating a cointegration relationship. It is worth noting that Johansen's trace test suggests three cointegration relationships in the system, equal to the number of endogenous variables. This implies that the processes are stationary, which contradicts the results of Engle and Granger's two-step procedure.

Sims et al. (1990) show that unrestricted VAR in the levels (VARL) produces consistent estimates of cointegrated data without the determination of cointegration relationships. Suharsono et al. (2017) and Zhang et al. (2010) also show that fitting a VARL model can provide better out-of-sample forecasting than implementing a cointegration constraint. Therefore, we investigate both VARL and VEC models and compare their performance. One important component of model specification for time series models is to select the correct number of lags. In addition, our model involves several exogenous variables, which also require lag selection. There is not much guidance in the literature on how to select lags for exogenous variables, therefore, we choose the model that generates the lowest information criterion. We use three types of information criteria to determine the best model fit, namely the corrected Akaike Information Criterion (AICC), Bayesian Information Criterion (BIC), and Hannan-Quinn Criterion (HQC). The calculated information criteria are presented in Table 3.5[13]. We can see that model VEC(2, 0) is the most appropriate given it

[12]Unit root tests conducted here do not incorporate a drift term or a deterministic trend. If a deterministic component is included, the test statistics for the three wood product price series reject the null hypothesis, implying stationarity. Additionally, we examined level stationarity using the KPSS test, and the results were contrary to those of the ADF and PP tests. As for exogenous variables, the presence of a unit root is confirmed across all tests.

[13]Note that a VEC(p) model has an equivalent VAR(p). Please see section A.1 for details

minimizes all three information criteria[14]. We implement another model diagnostic check, which is the Portmanteau test, to see whether correlations remain in the model residuals. The test results are presented in Table 3.6. The null hypothesis of the Portmanteau test is that all autocorrelations and cross-correlations in the residuals are zero. Looking at the p-values, the null hypothesis is rejected only for the third and fifth lag, and all other lags are uncorrelated.

Lastly, we compare model performance by evaluating their out-of-sample fit. Forecasting accuracy is measured using the following three metrics: Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Theil's U statistics.

$$RMSE = \sqrt{\sum_{t=1}^{N}(A_t - F_t)^2} \tag{3.9}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| \tag{3.10}$$

$$\textit{Theil's } U = \sqrt{\frac{\sum_{t=1}^{N-1}(\frac{F_{t+1}-A_{t+1}}{A_t})^2}{\sum_{t=1}^{N-1}(\frac{A_{t+1}-A_t}{A_t})^2}} \tag{3.11}$$

where A is the actual value, and F is the forecasted. $N = 38$. Unlike RMSE and MAPE, which emphasized the magnitude of the forecast error, Theil's U statistics compare the model's forecast to naive guessing. If Theil's U statistics is less than 1, then the forecasting technique is better than guessing. Comparison of forecasting accuracy metrics is in Table 3.7. Overall, VEC model gives better out-of-sample forecasts for the two lumber products, where VARL does a better job predicting OSB sheathing.

Note that in addition to VARL and VEC models, we examine the performance of a Bayesian approach using Minnesota priors (Litterman 1986), for which the optimal priors are selected with a grid search. Bayesian-type models have shown successes in Macroeconomics models and other fields for forecasting (Bergman et al. 2017; Karlsson 2013), which motivates us to consider whether the Bayesian approach is suitable in the setting of this study. The details of the implementation and results are described in Appendix B. To summarize, we find that Bayesian models provide an equally good or slightly better forecast. This is evident from the comparison of forecasts across all models, which is presented in Table A.1. However, the differences in forecasting metrics are not prominent.

[14]Models with longer and shorter lag length are examined, although the diagnose results are not included in the table as these models do not appear to give a better fit than the two models presented here.

Considering the calculating complexity and low interpretability of the Bayesian models, we think the classic VEC model with a specification of exogenous variables is the most appropriate model for the purpose of this study. Overall, we select VEC(2, 0) as our model for the following empirical studies.

3.5.2 Empirical Results

To empirically assess the effect of the COVID-19 pandemic on the manufactured wood industry, we conduct an event study by comparing forecasted manufactured wood prices from the specified VEC(2,0) model to the observed price under the pandemic. We transform the time series into logs before conducting analyses. We choose March 20, 2020, as the event start date. This date is right after the declaration of COVID-19 being a national pandemic. This date is indicated in Figure 3.2 and Figure 3.3 by the vertical line, where it exhibits a notable change in the behavior across all variables. To further confirm the occurrence of a structural break around this date, we perform a Chow Test (Chow 1960). The test statistics presented in Table 3.8 confirm that a significant structural break has occurred in all price series at the start of the pandemic. The result of the event study is displayed in Figure 3.4, where the expected prices forecasted based on model estimates are plotted against the actual prices observed during the event period.

There are several things to note after observing the graphs. First, the model appears to fit well before the start of the event, given that the fitted values follow close to the observed values. Second, it becomes evident that, even after accounting for possible structural changes from related sectors, the model forecasts still fail to capture the extreme volatility in prices during the pandemic; the observed price significantly deviates from the forecasted prices. This confirms the extreme price volatility introduced to the manufactured wood product markets during the pandemic. An interesting point to note is that spikes in wood product prices correspond to spikes in death rates due to the pandemic, as described in Figure 3.5. The data on the weekly percentage of deaths due to COVID-19 in the U.S. is retrieved from the Centers for Disease Control (CDC). We can observe similar spikes as those seen in wood product prices. In fact, the spikes in death rates almost exactly overlap with spikes in Southern Yellow Pine prices with a four-month delay. Lastly, it is clear that all three wood product prices have returned to the expected level around three years into the pandemic, suggesting the shocks from the pandemic are transitory.

In addition, one of the increases in prices during the event window coincides with the announcement of the doubling of tariffs on Canadian softwood, as indicated in Figure 3.6.

Forest product prices, especially softwood lumber, can be significantly affected by changes in the trade policy between the U.S. and Canada. Given different economic conditions, quite interesting to observe how the wood product prices react differently when faced with substantial tariff increases on Canadian softwood at different points in time. Figure 3.6 indicates two large increases in the softwood lumber tariff since the expiry of the SLA in 2015. The first incident here happened on April 24, 2017, when U.S. Commerce Secretary Wilbur Ross announced that his agency would impose new anti-subsidy duties averaging 20 % on Canadian softwood lumber imports (Reuters 2017). This is the first official release regarding the softwood trade dispute since the SLA2006. According to Ross, this move will "affect $5 billion worth of lumber coming in from Canada, which is about 31.5 % of the total U.S. market" (Reuters 2017). This high tariff placement affects the Canadian and the U.S. markets, driving a disparity between the two market prices. Market reaction to a tariff increase is different, however, given the COVID-19 pandemic background. On November 24, 2021, the U.S. Commerce Department placed a 17.99 % tariff, more than twice the 8.99% rate imposed during the Trump administration. Instead of experiencing a drop in price, the market responded with a sharply increased price at the time of this incident, making it reasonable to assume that the effect of tariffs on the U.S. market is more prominent during the pandemic.

To quantify the magnitude and direction of the impacts of the pandemic channeled through different sectors, we conduct impulse response analysis, and results are shown in Figure 3.10-Figure 3.12. These plots contain impulse responses of the three manufactured wood products to a one percent shock in the exogenous variables. There are several things to note based on the impulse response plots. All three wood product prices respond significantly to shocks from housing starts, the number of employees in the manufactured wood sector, and the Industrial Production Index (IPI). The effects of shocks from these three exogenous variables are transitory as the wood product prices converge back to zero or near-zero levels within 12 months. We do not observe significant responses to shocks from copper, aluminum, and crude oil prices, suggesting that the dynamics in the energy market and the metal market do not contribute much to price swings in wood product prices. Second, the directions of the responses are consistent with what we expect. In particular, all three wood product prices respond positively to an increase in housing starts. When there is an increase in the number of housing starts, it indicates an increase in housing demand, which subsequently drives up the demand for wood products and, in turn, leads to higher prices. On the other hand, the prices respond negatively to shocks in number of workers employed in the manufactured wood industry. This is also aligned with economic

intuitions since more workers implies more supply, therefore, bringing down the price. These results suggest that in the short run, the manufactured wood product markets can realize significant structural adjustments from other sectors such as sharp reductions in existing stocks, substantial expansion in new housing demand, and general economic conditions.

Another important aspect of the analysis is the assessment of relationships between these three wood products. The results are presented in Figure 3.7, Figure 3.8, and Figure 3.9. We can see that SYP price responds positively to a shock in SPF, while SPF does not respond to a price increase in SYP. This indicates there is a one-direction substitution between the two lumber markets, that SYP treats SPF as a substitute but not the other way around. Lastly, OSB responds negatively to a positive shock in both lumber products, and vice versa, indicating that sheathing and softwood lumber are complements of each other.

3.6 Conclusion

the U.S. manufactured wood industry has experienced abnormally high price volatility since the beginning of the COVID-19 pandemic. The price variations are substantial even three years after the declaration of this worldwide pandemic. As wood product price movements reflected changes in the labor and housing market, studying the industry becomes a crucial part of understanding the general economic growth and recovery. The rapid changes in wood prices intensify the underlying risks and uncertainty for those in construction industries and, at the same time, create challenges for the government to place recovery policies. We examine the impact of the pandemic on three popular manufactured wood products, namely Spruce-Pine-Fir, Southern Yellow Pine, and Oriented Strand Board, through an event study. The forecasted values are generated after fitting a Vector Error Correction model with exogenous variables. While the event study results suggest that during the pandemic, the wood product prices deviate significantly from the expected price based on the model, our contention is that the prices of manufactured wood products have returned to their pre-pandemic levels.

Studying the price behaviors of wood products in response to outside shocks as the forest industry reflects supply and demand dynamics, making it an important economic indicator. The inclusion of exogenous variables allows us to study the relationships among multiple time series variables from other sectors, and analyze how exogenous shocks are transmitted across the manufactured wood product prices through an impulse response

analysis. We find that all three wood product prices respond positively to an increase in housing starts, and negatively to an increase in employment of the manufactured wood sector. These results confirm our theory that disruptions in both labor and housing markets during the pandemic play a significant role in explaining the abnormally volatile wood product prices.

In addition, the fact that a large share of softwood lumber consumed domestically is imported from abroad makes lumber price movements especially important when making trade policies. In particular, the ongoing trade dispute between the U.S. and Canada. This makes assessing the cross-commodity price transmission among manufactured wood product markets interesting, as it provides a general idea of market resilience and price signals pass-through among those markets, and provides insights into the relationships between these products. We find that in general, SYP views SPF as substitutes, but not the other around. Moreover, OSB sheathing appears to be complements to both lumber products. This result provides implications for future trade analysis of the U.S.-Canada lumber trade issue.

Perhaps some long-run effects remain unidentified, but the majority of the pandemic-induced market volatility has diminished. While markets haven't fully returned to their normal conditions in the long term, evidence suggests that the manufactured wood product industry can reduce deviations from equilibrium caused by significant unexpected shocks within three years. We believe that an extended time series analysis spanning more than five years post-pandemic is preferable to ensure the complete dissipation of lingering effects stemming from these external shocks. Our analysis can be extended to include additional lumber species and related wood products, and it can also be applied to other international trade commodities to examine relationships between imported and domestically produced products.

3.7 Tables and Figures

Table 3.1: Variable Descriptions and Data Source

	Descriptions	Data Source
Spruce-Pine-Fir	$/Thousands board-feet	Random Lengths, part of Fastmarkets
Southern Yellow Pine	$/Thousands board-feet	Random Lengths, part of Fastmarkets
OSB Sheathing	$/Thousands Square Feet	Random Lengths, part of Fastmarkets
Aluminum	LME 3-Month Rolling Forward, $/ton	Bloomberg
Copper	LME 3-Month Rolling Forward, $/ton	Bloomberg
Crude Oil	WTI Crude Oil, $/Barrel	U.S. Energy Information Administration
Housing Starts*	Thousands of Units	U.S. Census Bureau, U.S. Department of Housing and Urban Development
Employment*	Thousands of Persons	U.S. Bureau of Labor Statistics
IPI*	Industrial Production Index, 2007=100	Board of Governors of the Federal Reserve System (U.S.)
CPI*[1]	Consumer Price Index for All Urban Consumers, 1982-1984=100	U.S. Bureau of Labor Statistics

*Asterisk: Data retrieved from FRED, Federal Reserve Bank of St. Louis.

[1] CPI is used to deflate all prices for the analysis.

Table 3.2: Summary Statistics, Before and After April 2020

	Pre-Pandemic (Jan. 1995 - Mar. 2020)					Event Period (Apr. 2020 - May 2023)				
	N	Mean	St. Dev.	Min.	Max.	N	Mean	St. Dev.	Min.	Max.
Spruce-Pine-Fir	303	353.36	69.47	155.25	582.60	38	714.58	308.32	329.50	1592.50
Southern Yellow Pine	303	351.96	81.77	209.50	588.50	38	588.86	265.13	326.75	1234.00
OSB Sheathing	303	227.85	80.37	132.50	515.20	38	613.38	346.87	205.50	1386.25
Aluminum	303	1859.28	419.89	1186.50	3114.00	38	2371.58	471.42	1494.50	3491.00
Copper	303	4715.26	2499.59	1374.00	9885.00	38	8386.03	1353.89	5188.50	10375.00
Crude Oil	303	53.13	28.92	11.35	133.88	38	70.16	23.54	16.55	114.84
Housing Starts	303	109.53	38.58	31.90	197.90	38	125.90	17.53	84.90	164.30
Employment	303	753.10	124.10	528.00	1037.30	38	876.28	53.83	697.80	945.90
IPI	303	93.31	8.39	69.86	106.08	38	99.32	4.67	82.64	104.81

Note: Data frequency is monthly.

Table 3.3: Augmented Dickey-Fuller and Phillips-Perron Unit Root Tests[1]

	Level Data				
	Optimal Lag[2]	ADF Tau Statistics	ADF P-value	PP Tau Statistics	PP P-value
Spruce-Pine-Fir	2	-0.514	0.493	-0.564	0.472
Southern Yellow Pine	1	-0.740	0.396	-0.743	0.395
OSB Sheathing	2	-0.463	0.514	-0.431	0.526
Aluminum	1	-0.830	0.357	-1.031	0.272
Copper	3	-0.096	0.650	-0.113	0.645
Crude Oil	2	-0.452	0.518	-0.311	0.573
Housing Starts	15	-0.259	0.593	-0.013	0.678
Employment	15	0.111	0.718	0.565	0.838
IPI	16	0.934	0.907	1.303	0.952
	First Difference				
	Optimal Lag[2]	ADF Tau Statistics	ADF P-value	PP Tau Statistics[3]	PP P-value
Spruce-Pine-Fir	1	-11.954	<.0001	-14.427	<.0001
Southern Yellow Pine	0	-16.334	<.0001	-16.337	<.0001
OSB Sheathing	2	-10.738	<.0001	-15.172	<.0001
Aluminum	0	-17.308	<.0001	-17.307	<.0001
Copper	0	-15.475	<.0001	-15.463	<.0001
Crude Oil	1	-9.084	<.0001	-12.119	<.0001
Housing Starts	14	-2.420	0.015	-16.463	<.0001
Employment	14	-2.095	0.035	-7.685	<.0001
IPI	15	-3.902	<.0001	-26.319	<.0001

[1]Unit root tests are performed on pre-pandemic data (Jan. 1995–Mar. 2020). The null hypothesis is the same for two tests. H_0: The time series contains a unit root.
[2]Optimal lags are selected according to minimized Schwarz Information Criterion; the upper bound maximum number of lag used is 16. It is determined using Schwert's (1989) rule of thumb.
[3]PP test must be performed with a minimum of 1 lag. Therefore, for variables that have an optimal lag length of 1, the test is performed with 1 lag.

Table 3.4: Cointegration Test Statistics

Engle and Granger's 2-Step Procedure			Johansen's Trace Test		
	ADF Tau	P-Value		Test Stastistics	P-Value
SPF-SYP	-5.09	<.0001	H_0: r = 0	95.34	<.0001
SPF-OSB	3.54	0.0005	H_0: r = 1	49.02	<.0001
SYP-OSB	-2.841	0.0046	H_0: r = 2	15.07	0.0002

Table 3.5: Goodness of Fit Comparison Across Models

	VARL(2, 0)	VARL(2, 1)	VEC(2, 0)	VEC(2, 1)
AICC[1]	-3537.83	-3501.70	-3544.16	-3512.47
BIC[2]	-3304.18	-3237.11	-3323.36	-3256.83
HQC[3]	-3477.63	-3450.59	-3483.46	-3456.97

[1]AICC: Corrected Akaike Information Criterion
[2]BIC: Bayesian Information Criterion
[3]HQC: Hannan-Quinn Criterion (HQC)

Table 3.6: Model Diagnose: Portmanteau Test for Cross Correlations of Residuals

VEC(2, 0)		
Up To Lag	Chi-Square Statistics	P-Value
3	22.47	0.008
4	24.96	0.126
5	40.38	0.047
6	44.94	0.146
7	52.00	0.220
8	59.99	0.268
9	71.81	0.209
10	79.83	0.247

H_0: All auto-correlation and cross-correlations are zero.

Table 3.7: Forecasting Accuracy Metrics Comparisons

Variable	Model	RMSE	MAPE	Theil's U
Spruce-Pine-Fir	VARL(2, 0)	0.475	0.066	0.087
	VEC(2, 0)	0.452	0.061	0.083
Southern Yellow Pine	VARL(2, 0)	0.477	0.063	0.090
	VEC(2, 0)	0.473	0.065	0.089
OSB Sheathing	VARL(2, 0)	0.740	0.108	0.140
	VEC(2, 0)	0.832	0.129	0.158

Table 3.8: Chow Structural Break Test[1]

	Date of Structural Break	F Statistics	P-value
Spruce-Pine-Fir	4/1/2020	128.86	0.00
Southern Yellow Pine	4/1/2020	135.72	0.00
OSB Sheathing	4/1/2020	116.34	0.00

[1] H_0: There is no breakpoint at the defined date.

Figure 3.1: U.S. Home Supply and Median Sales Price for New (top) and Existing (bottom) Homes

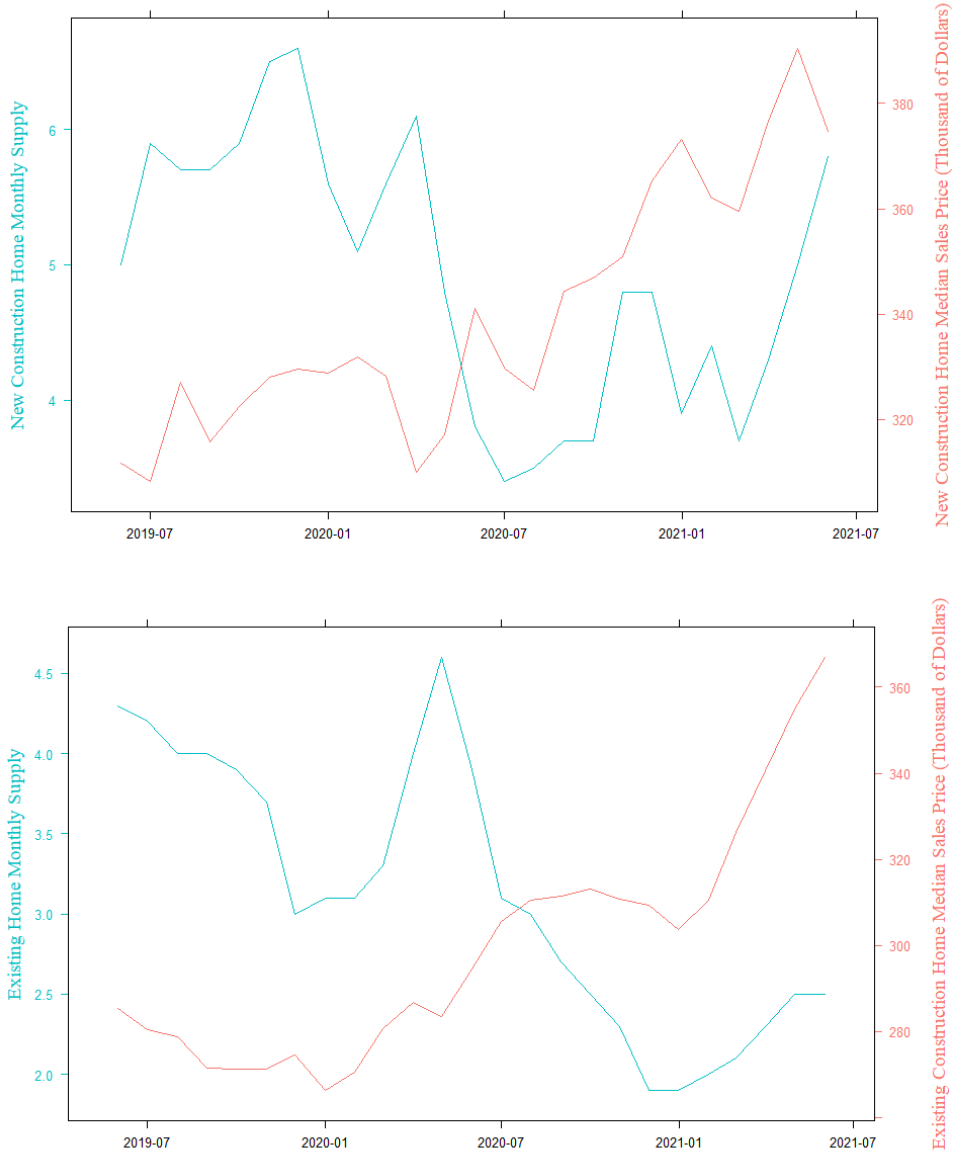


Figure 3.2: Price Movements of Spruce-Pine-Fir, Southern Yellow Pine, and OSB Sheathing

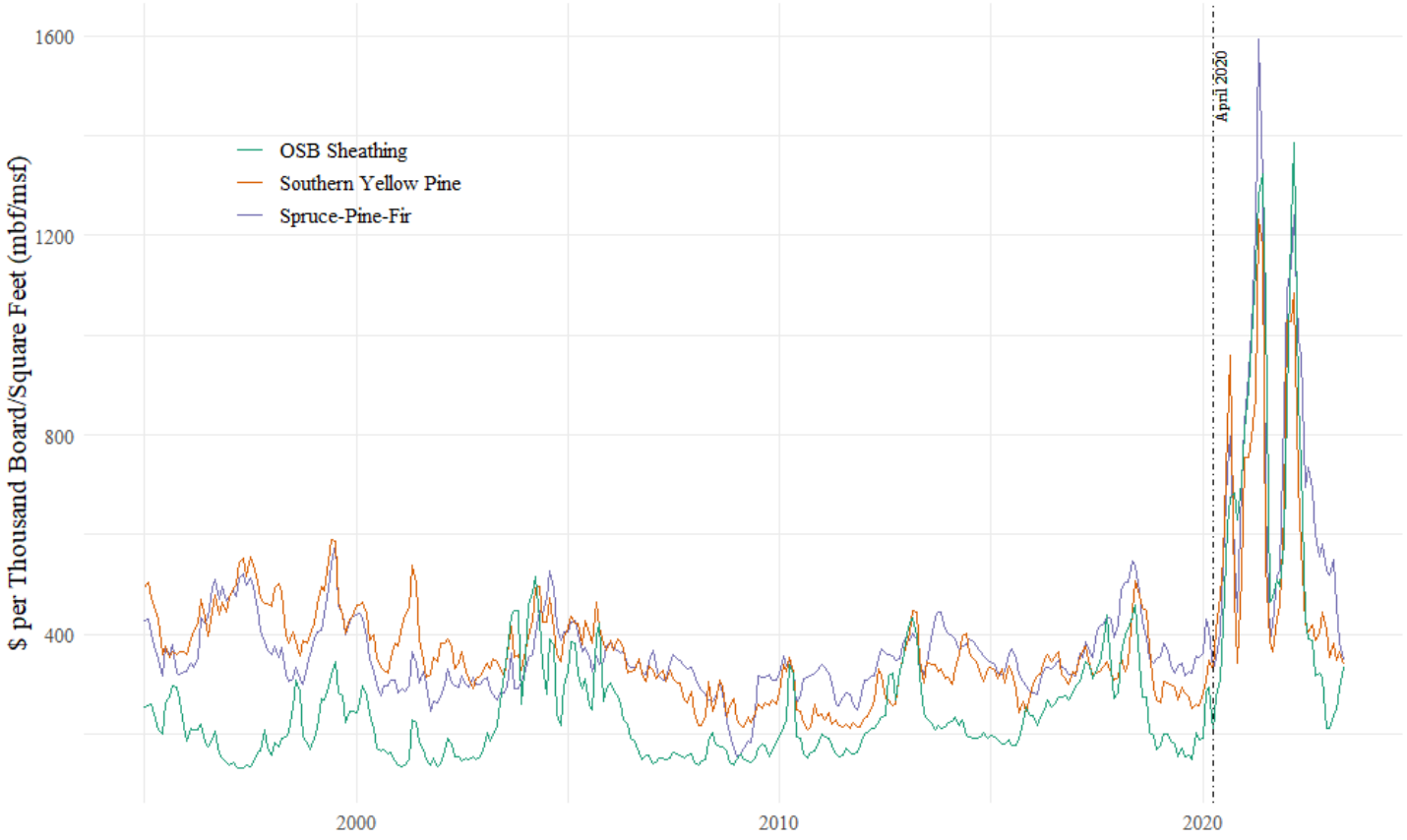
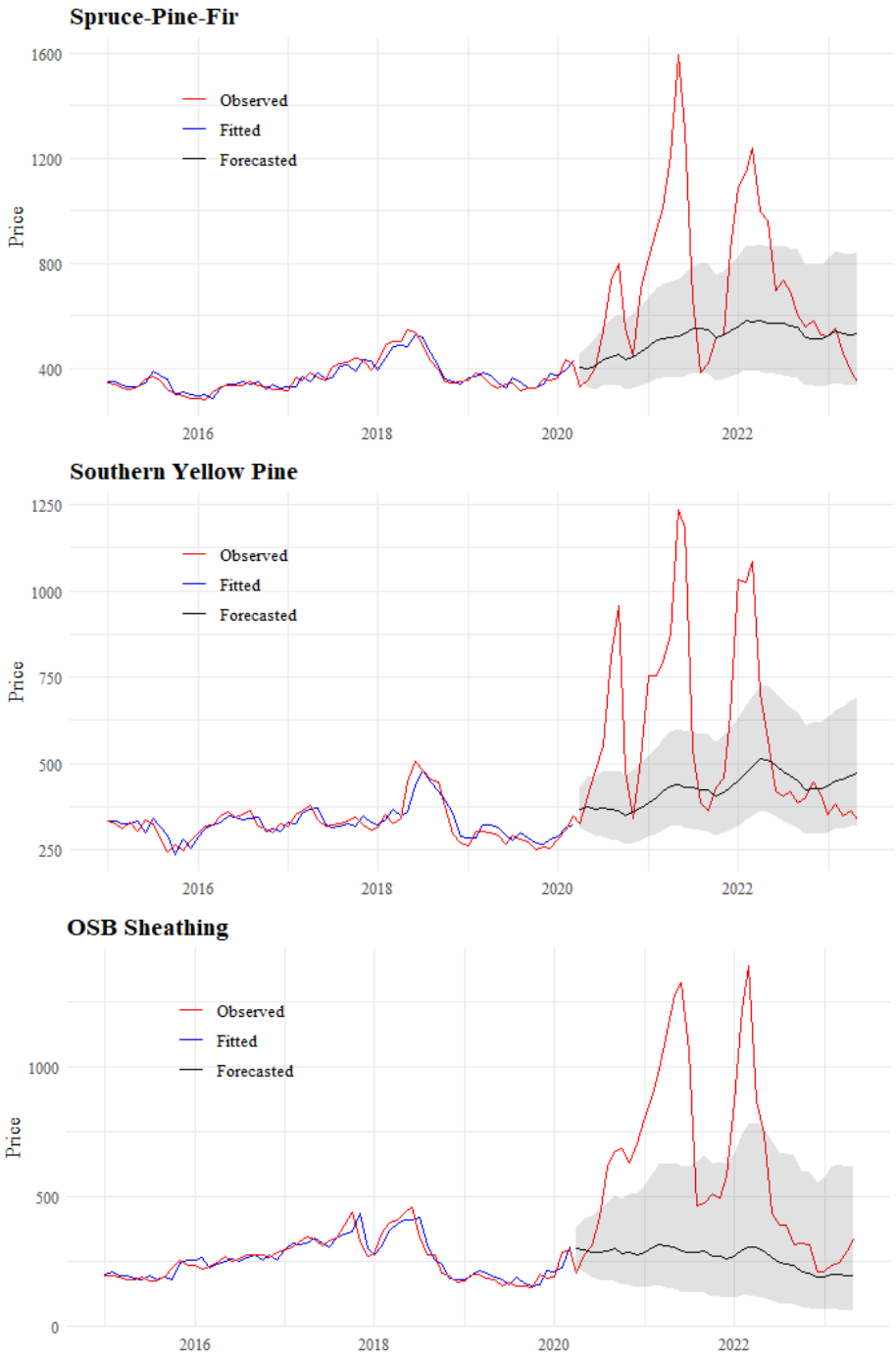


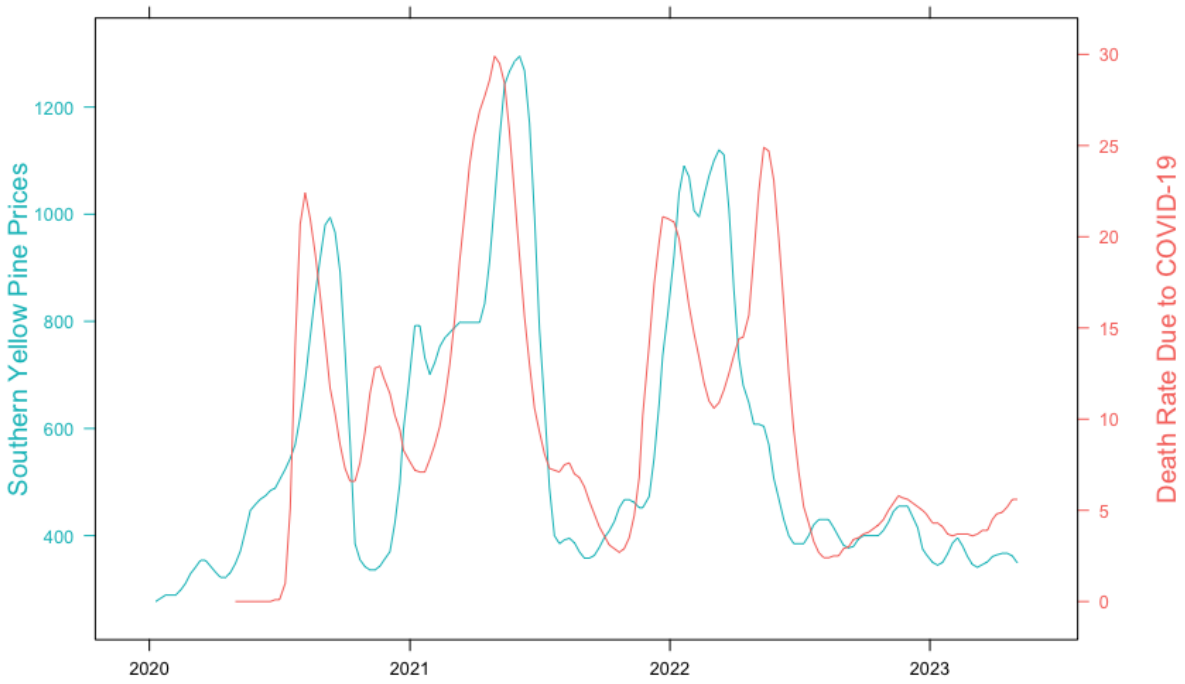
Figure 3.3: Time Series Plots for All Exogenous Variables

Figure 3.4: Expected and Actual Price Movement during the Event Period



Notes: The shaded area represents the 95% confidence band. The red line indicates the actual observed price, blue line indicates the model fitted values, and the black line is the model forecasted values. The event period is from April 2020 to May 2023. Forecasts are based on a VEC(2, 0) model, forecasting period is 38 months.

Figure 3.5: Southern Yellow Pine Prices and Percentage of Deaths Due to COVID-19



Notes: This figure overlays Southern Yellow Pine prices with death rates due to the COVID-19 pandemic. The data frequency is weekly.

Figure 3.6: Changes in Wood Product Price When Tariff Increased in 2017 and 2021

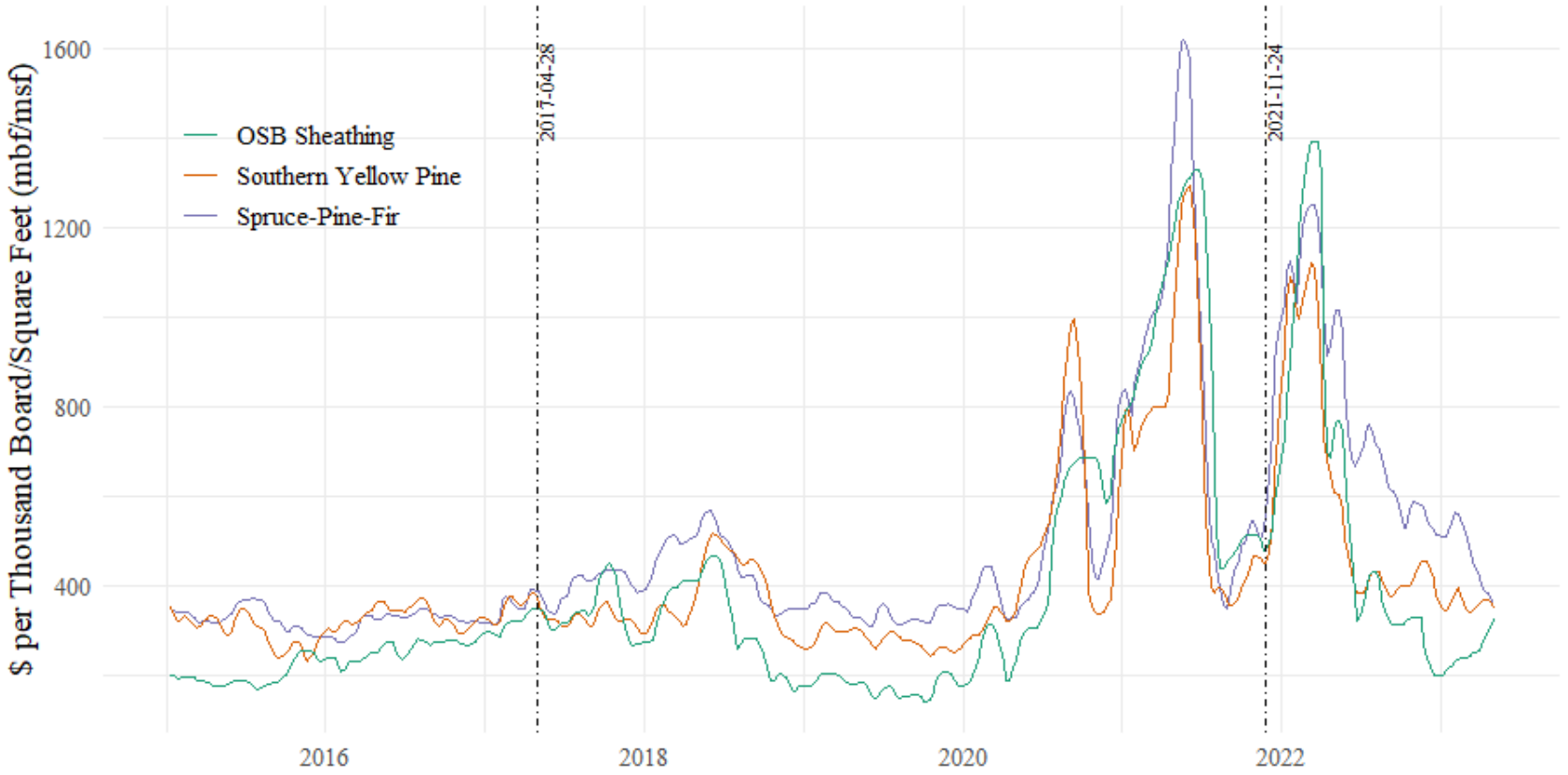
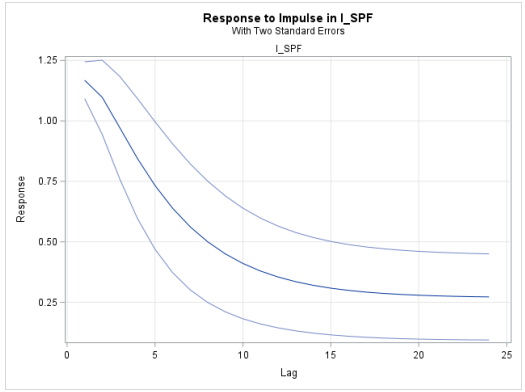
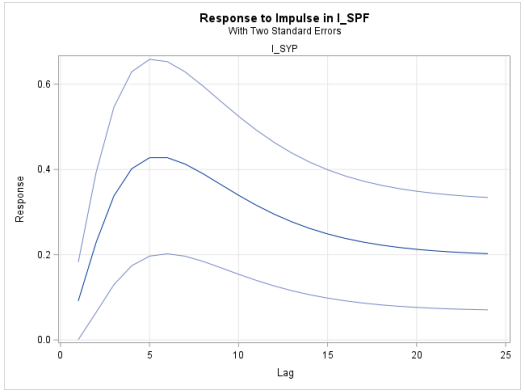


Figure 3.7: Responses to a Shock in Spruce-Pine-Fir

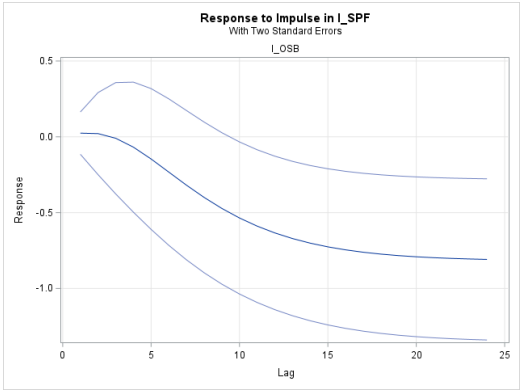
(a) Spruce-Pine-Fir



(b) Southern Yellow Pine



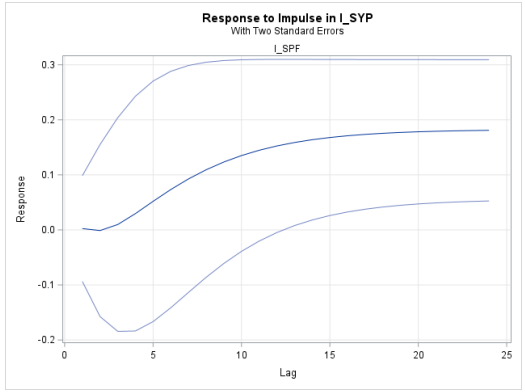
(c) OSB Sheathing



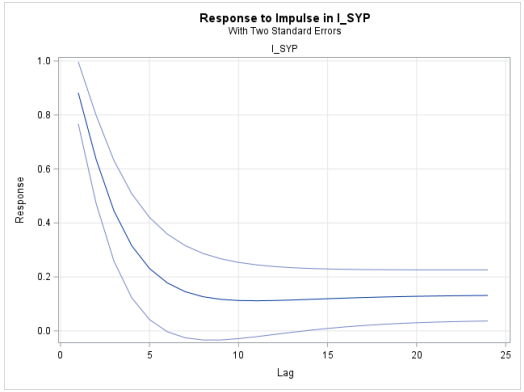
This figure presents responses to a shock in the price of Spruce-Pine-Fir. Panel (a) presents the self-response of SPF. Panel (b) and (c) present responses of SYP and OSB, respectively. Responses are generated based on a VEC(2, 0) model with seasonal dummies. The data frequency is monthly. All variables are transformed into natural logs.

Figure 3.8: Responses to a Shock in Southern Yellow Pine

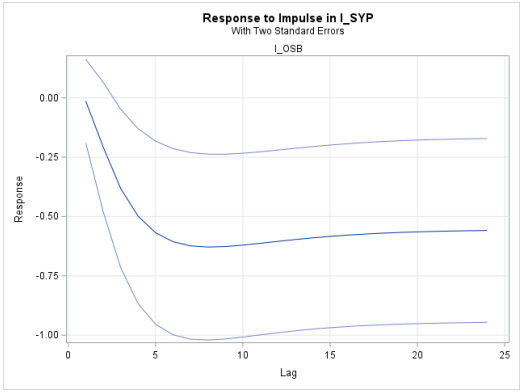
(a) Spruce-Pine-Fir



(b) Southern Yellow Pine



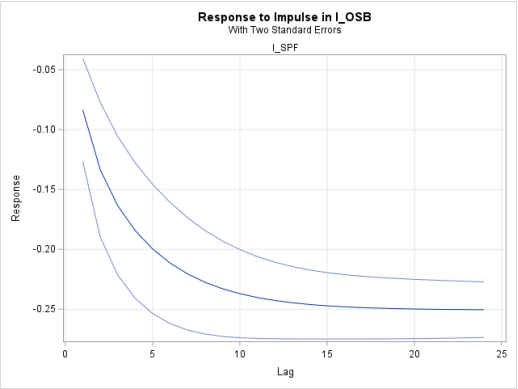
(c) OSB Sheathing



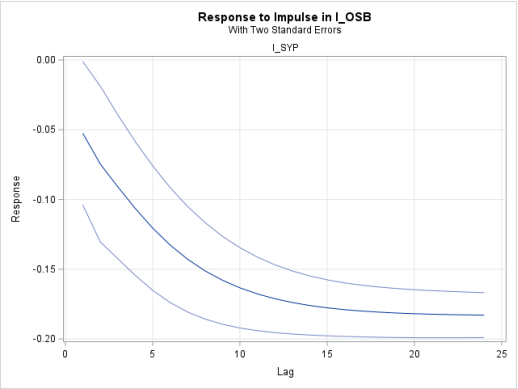
This figure presents responses to a shock in the price of Southern Yellow Pine. Panel (b) presents the self-response of SYP. Panel (a) and (c) present responses of SPF and OSB, respectively. Responses are generated based on a VEC(2, 0) model with seasonal dummies. The data frequency is monthly. All variables are transformed into natural logs.

Figure 3.9: Responses to a Shock in Oriented Strand Board

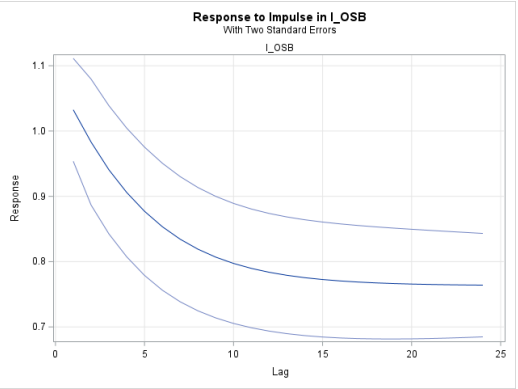
(a) Spruce-Pine-Fir



(b) Southern Yellow Pine

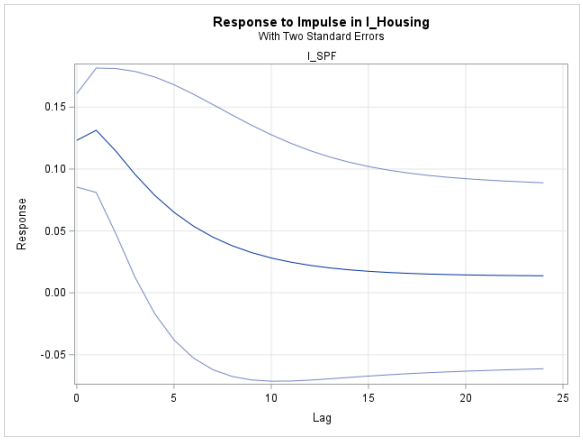


(c) OSB Sheathing

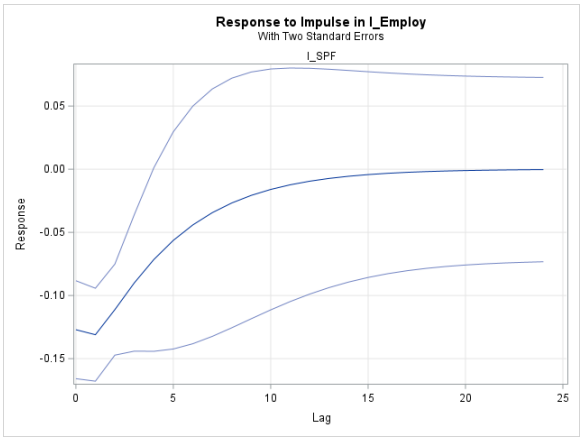


This figure presents responses to a shock in the price of the Oriented Strand Board. Panel (c) presents the self-response of OSB. Panel (a) and (b) present responses of SPF and SYP, respectively. Responses are generated based on a VEC(2, 0) model with seasonal dummies. The data frequency is monthly. All variables are transformed into natural logs.

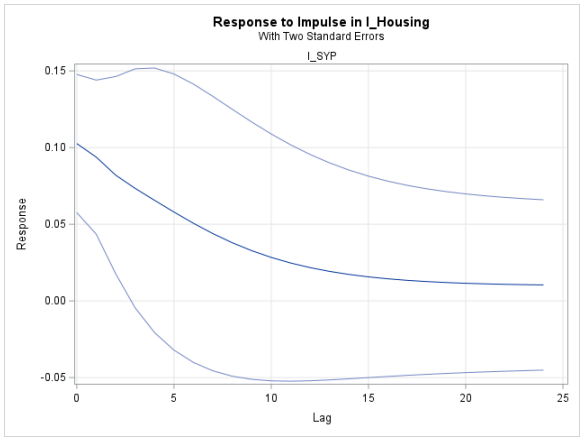
(a) Spruce-Pine-Fir



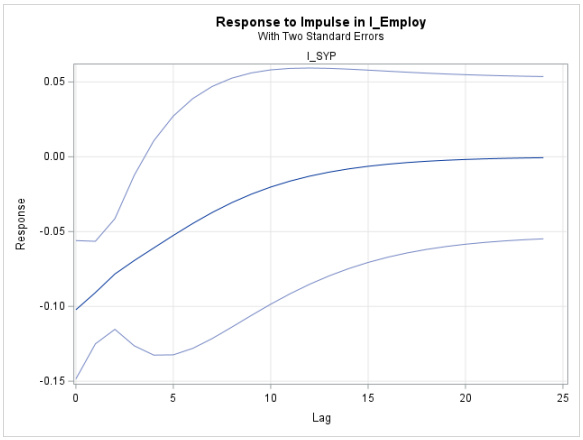
(b) Spruce-Pine-Fir



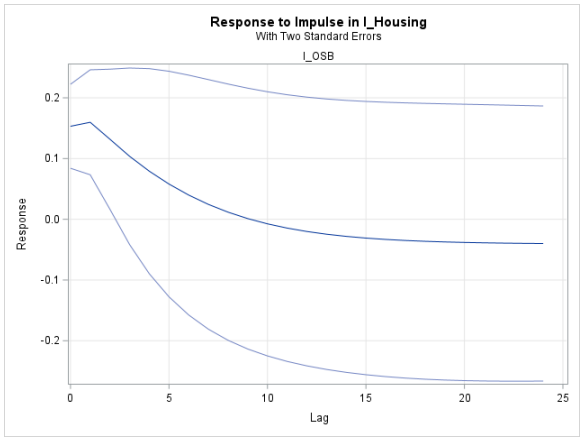
(c) Southern Yellow Pine



(d) Southern Yellow Pine



(e) OSB Sheathing



(f) OSB Sheathing

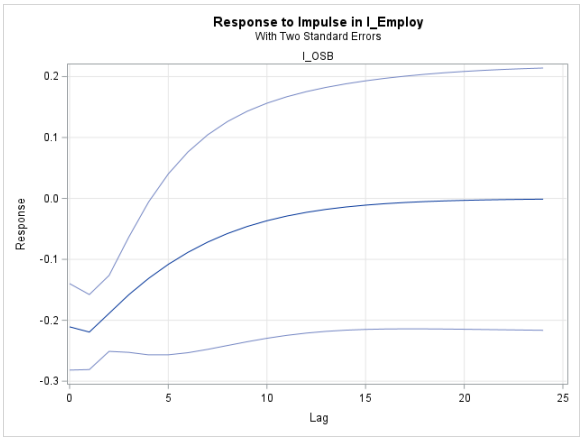
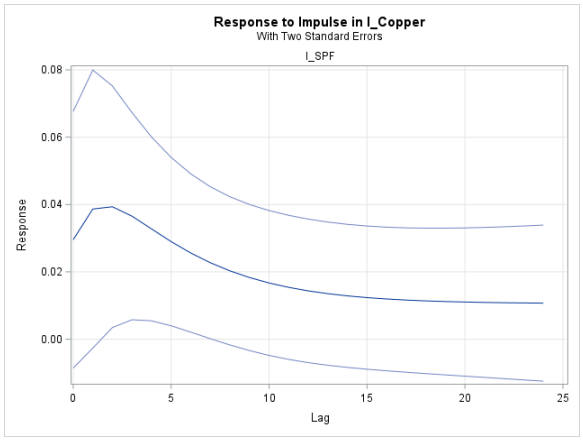
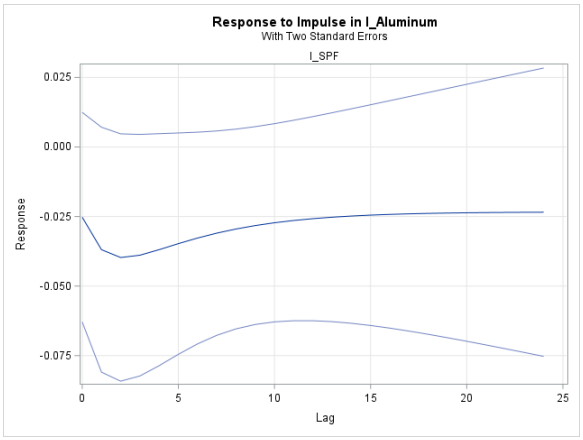


Figure 3.10: Responses to a Shock in Housing Starts (Left) and Employment (Right) (Monthly Data)

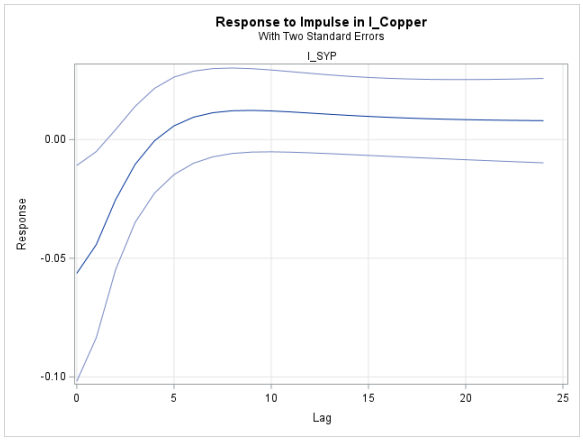
(a) Spruce-Pine-Fir



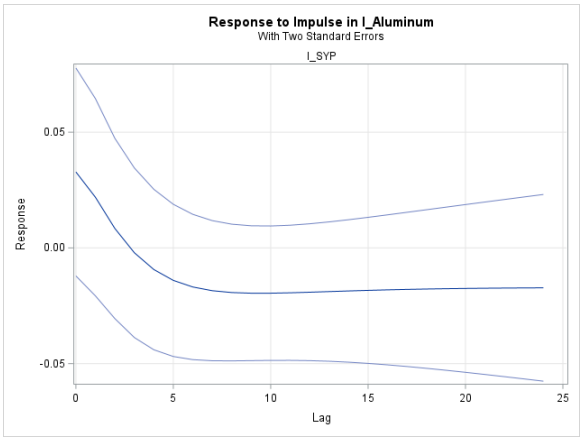
(b) Spruce-Pine-Fir



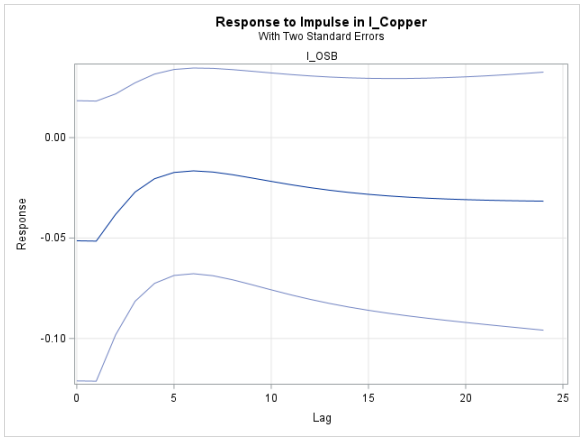
(c) Southern Yellow Pine



(d) Southern Yellow Pine



(e) OSB Sheathing



(f) OSB Sheathing

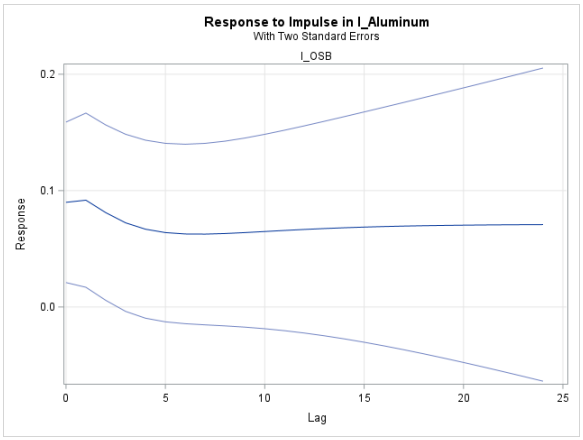
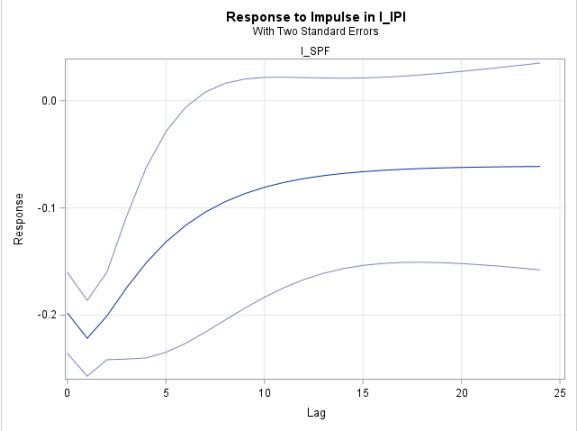
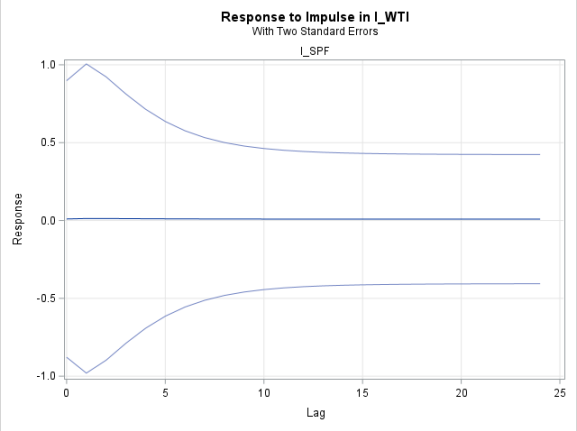


Figure 3.11: Responses to a Shock in Copper Price (Left) and Aluminum Price (Right) (Monthly Data)

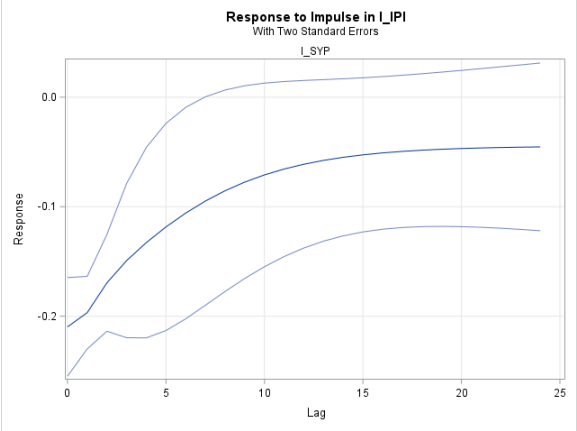
(a) Spruce-Pine-Fir



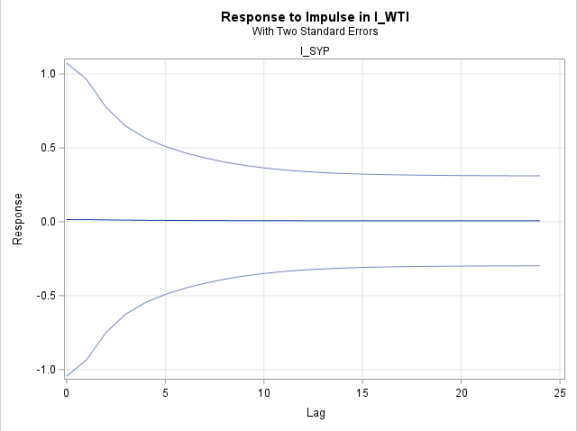
(b) Spruce-Pine-Fir



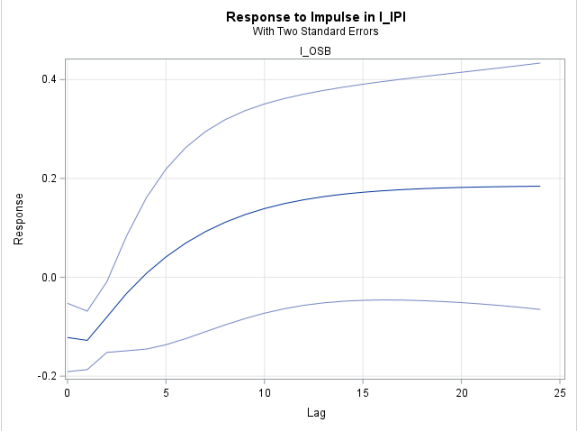
(c) Southern Yellow Pine



(d) Southern Yellow Pine



(e) OSB Sheathing



(f) OSB Sheathing

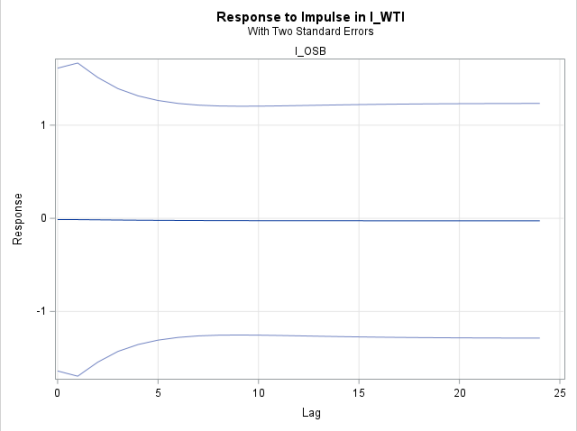


Figure 3.12: Responses to a Shock in IPI (Left) and Crude Oil Price (Right) (Monthly Data)

REFERENCES

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198.

Ahmed, O. and Serra, T. (2014). Economic analysis of the introduction of agricultural revenue insurance contracts in Spain using statistical copulas. *Agricultural Economics*, 46:69–79.

AIFP Lumber (2019). SYP Facts, Grades, Uses & Industry Insights, AIFP Blog.

AIFP Lumber (2020). SPF Wood Facts, Grades, Uses & Industry Insights.

Albanesi, S. and Kim, J. (2021). Effects of the COVID-19 recession on the US labor market: Occupation, family, and gender. *Journal of Economic Perspectives*, 35:3–24.

Alderman, D. (2022). U.S. forest products annual market review and prospects, 2015–2021. General Technical Report FPL-GTR-289. *U.S. Department of Agriculture, Forest Service*.

Anderson, J. D. and Zeuli, K. A. (2001). The revenue risk of value-based pricing for fed cattle: A simulation of grid vs. average pricing. *The International Food and Agribusiness Management Review*, 4:275–286.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61:821.

Babcock, B. A., Hart, C. E., and Hayes, D. J. (2004). Actuarial fairness of crop insurance rates with constant rate relativities. *American Journal of Agricultural Economics*, 86:563–575.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, 38:627.

Bedford, T. and Cooke, R. M. (2002). Vines–A new graphical model for dependent random variables. *The Annals of Statistics*, 30:1031–1068.

Bekkerman, A., Goodwin, B. K., and Piggott, N. E. (2013). A variable threshold band approach to measuring market linkages. *Applied Economics*, 45:2705–2714.

Belasco, E. J., Ghosh, S. K., and Goodwin, B. K. (2009). A multivariate evaluation of ex ante risks associated with fed cattle production. *American Journal of Agricultural Economics*, 91:431–443.

Belasco, E. J., Schroeder, T. C., and Goodwin, B. K. (2010). Quality risk and profitability in cattle production: A multivariate approach. *Journal of Agricultural and Resource Economics*, 35:385–405.

Bergman, J. J., Noble, J. S., McGarvey, R. G., and Bradley, R. L. (2017). A Bayesian approach to demand forecasting for new equipment programs. *Robotics and Computer-integrated Manufacturing*, 47:17–21.

Bessler, D. A. and Kling, J. L. (1986). Forecasting vector autoregressions with Bayesian priors. *American Journal of Agricultural Economics*, 68:144–151.

BLS (2020). Employer-reported workplace injuries and illness – 2019. *Bureau of Labor Statistics*.

BLS (2021). Employer-reported workplace injuries and illness – 2020. *Bureau of Labor Statistics*.

BLS OEWS (2022). Wood product manufacturing – May 2022 OEWS industry-specific occupational employment and wage estimates. *Bureau of Labor Statistics*.

Boyd, M., Pai, J., and Porth, L. (2013). Livestock mortality insurance: Development and challenges. *Agricultural Finance Review*, 73:233–244.

Boykin, C. A., Eastwood, L. C., Harris, M. K., Hale, D. S., Kerth, C. R., Griffin, D. B., Arnold, A. N., Hasty, J. D., Belk, K. E., Woerner, D. R., Delmore, R. J., Martin, J. N., VanOverbeke, D. L., Mafi, G., Pfeiffer, M., Lawrence, T., McEvers, T. J., Schmidt, T. B., Maddock, R. J., Johnson, D. D., Carr, C., Scheffler, J. M., Pringle, T. D., Stelzleni, A. M., Gottlieb, J. S., and Savell, J. W. (2017). National beef quality audit–2016: In-plant survey of carcass characteristics related to quality, quantity, and value of fed steers and heifers. *Journal of Animal Science*, 95:2993–3002.

Buongiorno, J. and Johnston, C. (2018). Potential effects of US protectionism and trade wars on the global forest sector. *Forest Science*, 64:121–128.

Canada Forest Service (2022). Statistical data. *Government of Canada, Natural Resources Canadaa*.

Canada Wood Group (2021). Spruce pine fir. *Canada Wood Group.org*.

Canadian Forest Service (2021). Why are prices so high? *Government of Canada, Natural Resources Canada*.

CBC News (2006). CBC News Indepth: Softwood lumber dispute. *cbc.ca*.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21:520–533.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28:591.

Christiano, L. J. (2012). Christopher a. sims and vector autoregressions. *The Scandinavian Journal of Economics*, 114:1082–1104.

Coble, K. H., Hanson, T. R., Miller, J. C., and Shaik, S. (2003). Agricultural insurance as an environmental policy tool. *Journal of Agricultural and Applied Economics*, 35:391–405.

Coble, K. H., Knight, T. O., Goodwin, B. K., Miller, M. F., and Rejesus, R. M. (2010). A comprehensive review of the RMA APH and COMBO rating methodology final report. *U.S. Department of Agriculture, Risk Management Agency*.

Congressional Research Service (2001). Softwood lumber imports from Canada: History and analysis of the dispute, CRS Report No. RL30826. *Congressional Research Service*.

Corrado, C. J. (2010). Event studies: A methodology review. *Accounting and Finance*, 51:207–234.

Cummings, C., Scott, S., Cohen, C., Brinckhaus, J., Alexander, M., Lee, C., Dempsey, J., and Haines, E. (2023). Softwood lumber products from Canada, Investigation Nos. 701-TA-566 and 731-TA-1342 (Review). *U.S. International Trade Commission*.

Daniels, J. M. (2011). Stumpage market integration in western national forests. *CiteSeer X, Pennsylvania State University*.

Davidson, J., Hendry, D. F., Srba, F., and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *The Economic Journal*, 88:661–661.

Davis, M. A., Ghent, A. C., and Gregory, J. (2021). The work-from-home technology boon and its consequences. *RePEc: Research Papers in Economics*.

Dawson, A. (1981). Sargan's wage equation: A theoretical and empirical reconstruction. *Applied Economics*, 13:351–363.

DeLuca, M. J. and Pinheiro, R. (2023). US labor market after COVID-19: An interim report. *Economic commentary*.

Dempsey, J., Martinez, K., and Haines, E. (2019). Softwood lumber from Canada, Investigation Nos. 701-TA-566 and 731-TA-1342 (Final) (Remand). *United States International Trade Commission*.

Devadoss, S. (2009). Book review: The softwood lumber war: Politics, economics, and the long U.S.–Canadian trade dispute. *The Journal of International Trade & Economic Development*, 18:197–203.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Embrechts, P., Mcneil, A., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. *Risk Management: Value at Risk and Beyond*, pages 176–223.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276.

Fackler, P. L. and Goodwin, B. K. (2001). Chapter 17: Spatial price analysis. *Handbook of Agricultural Economics*, pages 971–1024.

Fanchon, P. and Wendel, J. (1992). Estimating VAR models under non-stationarity and cointegration: Alternative approaches for forecasting cattle prices. *Applied Economics*, 24:207–217.

Fausti, S. W., Wang, Z., Qasmi, B. A., and Diersen, M. A. (2014). Risk and marketing behavior: Pricing fed cattle on a grid. *Agricultural Economics*, 45:601–612.

Federal Reserve Board (2017). Industrial production and capacity utilization – G.17. *Board of Governors of the Federal Reserve System*.

Feng, X. and Hayes, D. (2016). Diversifying systemic risk in agriculture. *Agricultural Finance Review*, 76:512–531.

Feuz, D. M. (1999). Market signals in value-based pricing premiums and discounts. *Journal of Agricultural and Resource Economics*, 24:327–341.

Fischer, B. L. and Outlaw, J. (2023). What's in the farm bill for livestock producers? *Southern Ag Today*.

Foster, A. C. (2008). Summary and explanation of the U.S.-Canada lumber dispute. *Penn State Dickinson School of Law*.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25.

Gagné, G. (2003). The Canada-US softwood lumber dispute: A test case for the development of international trade rules. *International Journal*, 58:335.

Gamber, W., Graham, J., and Yadav, A. (2023). Stuck at home: Housing demand during the COVID-19 pandemic. *Journal of Housing Economics*, 59:101908–101908.

Gan, J. (2006). Substitutability between US domestic and imported forest products: The armington approach. *Forest Science*, 52:1–9.

Gan, J., Tian, N., Choi, J., and Pelkki, M. H. (2022). Synchronized movement between US lumber futures and southern pine sawtimber prices and COVID-19 impacts. *Canadian Journal of Forest Research*, 52:614–621.

GAO (2023). Farm bill: Reducing crop insurance costs could fund other priorities, Report No. GAO-23-106228. *U.S. Government Accountability Office*.

Genest, C. and Rivest, L.-P. (1989). A characterization of Gumbel's family of extreme value distributions. *Statistics & Probability Letters*, 8:207–211.

Gerardi, K., Lambie-Hanson, L., Willen, P., S. Goodman, L., and Wachter, S. (2022). Lessons learned from housing policy during COVID-19. *The Hamilton Project*.

Glauber, J. W. (2022). The US animal insurance program: Rapid expansion at a growing cost to taxpayers. *American Enterprise Institute*.

Global Affairs Canada (2023). Statement by minister Ng on U.S. duty rates on Canadian softwood lumber. *Government of Canada*.

Goodwin, B. K. (1994). Premium rate determination in the federal crop insurance program: What do averages have to say about risk? *Journal of Agricultural and Resource Economics*, 19:382–395.

Goodwin, B. K. (2001). Problems with market insurance in agriculture. *American Journal of Agricultural Economics*, 83:643–649.

Goodwin, B. K. and Holt, M. T. (1999). Price transmission and asymmetric adjustment in the U.S. beef sector. *American Journal of Agricultural Economics*, 81:630–637.

Goodwin, B. K., Holt, M. T., and Prestemon, J. P. (2011). North American oriented strand board markets, arbitrage activity, and market price dynamics: A smooth transition approach. *American Journal of Agricultural Economics*, 93:993–1014.

Goodwin, B. K., Holt, M. T., and Prestemon, J. P. (2021). Semi-parametric models of spatial market integration. *Empirical Economics*, 61:2335–2361.

Goodwin, B. K., Holt, M. T., Önel, G., and Prestemon, J. P. (2017). Copula-based nonlinear modeling of the law of one price for lumber products. *Empirical Economics*, 54:1237–1265.

Goodwin, B. K. and Hungerford, A. (2014). Copula-based models of systemic risk in U.S. agriculture: Implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, 97:879–896.

Goodwin, B. K. and Ker, A. P. (1998). Revenue insurance: A new dimension in risk management. *Choices*, 13:24–27.

Goodwin, B. K. and Mahul, O. (2004). *Risk modeling concepts relating to the design and rating of agricultural insurance contracts*. World Bank Group.

Goodwin, B. K. and Piggott, N. E. (2001). Spatial market integration in the presence of threshold effects. *American Journal of Agricultural Economics*, 83:302–317.

Goodwin, B. K., Ramsey, A. F., and Chvosta, J. (2018). *Applied econometrics with SAS: Modeling demand, supply, and risk*. SAS Institute.

Goodwin, B. K. and Smith, V. H. (2013). What harm is done by subsidizing crop insurance? *American Journal of Agricultural Economics*, 95:489–497.

Goodwin, B. K. and Vado, L. (2007). Public responses to agricultural disasters: Rethinking the role of government. *Canadian Journal of Agricultural Economics-revue Canadienne D Agroeconomie*, 55:399–417.

Gorte, R. W. and Grimmett, J. J. (2006). Softwood lumber imports from Canada: Issues and events, CRS Report RL33752. *Congressional Research Service*.

Green, R., Larsson, K., Lunina, V., and Nilsson, B. (2018). Cross-commodity news transmission and volatility spillovers in the german energy markets. *Journal of Banking and Finance*, 95:231–243.

Hansen, B. (1999). Testing for linearity. *Journal of Economic Surveys*, 13:551–576.

Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics*, 110:293–318.

Hart, C. E., Koontz, S. R., Hoag, D. L., Thilmany, D., Green, J. W., and Grannis, J. L. (2006). The current state of US federally supported livestock insurance. *CABI Publishing eBooks*, pages 164–180.

Haynes, R. W. (1977). A derived demand approach to estimating the linkage between stumpage and lumber markets. *Forest Science*, 23:281–288.

Hogan, R., Anderson, D., and Schroeder, T. (2009). Grid pricing of fed cattle. *AgriLIFE Extension, Texas A&M System*.

Holly Wang, H. and Zhang, H. (2003). On the possibility of a private crop insurance market: A spatial statistics approach. *Journal of Risk & Insurance*, 70:111–124.

Hoover, K. and Fergusson, I. F. (2018). Softwood lumber imports from Canada: Current issues, CRS Report No. R42789.

Hoover, K. and Riddle, A. A. (2021). U.S. forest ownership and management: Background and issues for Congress,CRS Report No. R46976. *Congressional Research Service*.

Howard, J. L. and Liang, S. (2019). U.S. timber production, trade, consumption, and price statistics, 1965-2017. *U.S. Department of Agriculture, Forest Service*.

Igan, D., Kohlscheen, E., and Rungcharoenkitkul, P. (2022). Housing market risks in the wake of the pandemic.

Institute, M. E. (2022). Protectionist barriers to softwood lumber: A costly measure.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Institute of Mathematical Statistics eBooks*, 28:120–141.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Johnston, C. M. and van Kooten, G. C. (2017). Impact of inefficient quota allocation under the Canada-U.S. softwood lumber dispute: A calibrated mixed complementarity approach. *Forest Policy and Economics*, 74:71–80.

Kanieski, B., Schons, S. Z., Cubbage, F. W., and Parajuli, R. (2020). Spatial and cross-product price linkages in the Brazilian pine timber markets. *Forest Policy and Economics*, 117:102186–102186.

Karlsson, S. (2013). Forecasting with Bayesian vector autoregression. *Handbook of economic forecasting*, 2:791–897.

Keenan, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, 72:39.

Ker, A. P. and McGowan, P. (2000). Weather-based adverse selection and the U.S. crop insurance program: The private insurance company perspective. *Journal of Agricultural and Resource Economics*, 25:386–410.

Kojadinovic, I. and Yan, J. (2010). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34.

Koknaroglu, H., Loy, D., Wilson, D., Hoffman, M., and Lawrence, J. (2005). Factors affecting beef cattle performance and profitability. *The Professional Animal Scientist*, 21:286–296.

Koop, G., Pesaran, M., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74:119–147.

Kurowicka, D. and Cooke, R. M. (2006). *Uncertainty analysis with high dimensional dependence modelling*. John Wiley & Sons.

Kuzman, M. K., Oblak, L., Glavonjic, B., Barčić, A. P., Obucina, M., Haviarova, E., and Grošelj, P. (2022). Impact of COVID-19 on wood-based products industry: An exploratory study in Slovenia, Croatia, Serbia, and BiH. *Wood Material Science and Engineering*, 18:1115–1126.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Lee, T.-H., White, H., and Granger, C. W. (1993). Testing for neglected nonlinearity in time series models. *Journal of Econometrics*, 56:269–290.

Lindsey, B., Groombridge, M. A., and Loungani, P. (2000). Nailing the homeowner: The economic impact of trade protection of the softwood lumber industry.

Litterman, R. (1986). Forecasting with Bayesian vector autoregressions: Five years of experience. *Source: Journal of Business & Economic Statistics*, 4:25–38.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin, Heidelberg Springer Berlin Heidelberg.

Lütkepohl, H. and Krätzig, M. (2004). *Applied time series econometrics*. Cambridge University Press.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35:13–39.

Mercier, S. (2016). Federal benefits for livestock and specialty crop producers. *Choices*, 31:1–6.

Messner, Jakob, W., Mayr, Georg, J., and Zeileis, A. (2016). Heteroscedastic censored and truncated regression with crch. *The R Journal*, 8:173.

Meyers, W. H., Devadoss, S., and Helmar, M. D. (1987). Agricultural trade liberalization: Cross-commodity and cross-country impact products. *Journal of Policy Modeling*, 9:455–482.

Miller, D. L. (2023). An introductory guide to event study models. *Journal of Economic Perspectives*, 37:203–230.

Miranda, M. J. and Glauber, J. W. (1997). Systemic risk, reinsurance, and the failure of crop insurance markets. *American Journal of Agricultural Economics*, 79:206–215.

Monke, J. (2023). Farm bill primer: Budget dynamics, CRS Report No. IF12233, Ver. 7. *Congressional Research Service*.

Nagler, T., Schellhase, C., and Czado, C. (2017). Nonparametric estimation of simplified vine copula models: Comparison of methods. *Dependence Modeling*, 5:99–120.

Nagubadi, R. V., Zhang, D., Prestemon, J. P., and Wear, D. N. (2004). Softwood lumber products in the United States: Substitutes, complements, or unrelated? *Forest Science*, 50:416–426.

NAHB (2023). NAHB/Westlake Royal remodeling market index (RMI). *National Association of Home Builders*.

Natural Resources Canada (2022). Canada's softwood lumber industry. *Government of Canada, Natural Resources Canada*.

Nelsen, R. B. (2013). *An introduction to copulas*. Springer Science & Business Media.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69:1519–1554.

Ning, Z. and Sun, C. (2014). Vertical price transmission in timber and lumber markets. *Journal of Forest Economics*, 20:17–32.

Perron, P. and Qu, Z. (2007). A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters*, 94:12–19.

Phillips, A. W. (1957). Stabilisation policy and the time-forms of lagged responses. *The Economic Journal*, 67:265.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. *Economica*, 25:283–299.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53:473–495.

Prestemon, J. P. (2003). Evaluation of U.S. southern pine stumpage market informational efficiency. *Canadian Journal of Forest Research*, 33:561–572.

Prestemon, J. P., Wear, D. N., Abt, K. L., and Abt, R. C. (2017). Projecting housing starts and softwood lumber consumption in the United States. *Forest Science*, 64.

Rahman, S. O. and Devadoss, S. (2002). Economics of the US-Canada softwood lumber dispute: A historical perspective. *Estey Journal of International Law and Trade Policy*, 3.

Ramsey, A. F. and Goodwin, B. K. (2019). Value-at-risk and models of dependence in the U.S. federal crop insurance program. *Journal of Risk and Financial Management*, 12:65.

Ramsey, A. F., Goodwin, B. K., and Ghosh, S. K. (2019). How high the hedge: Relationships between prices and yields in the federal crop insurance program. *Journal of Agricultural and Resource Economics*, 44:227–245.

Ramsey, A. F., Goodwin, B. K., Hahn, W. F., and Holt, M. T. (2021). Impacts of COVID-19 and price transmission in U.S. meat markets. *Agricultural Economics*, 52:441–458.

Rancourt, O. and Giguère, G. (2022). Canadian softwood lumber: A costly dispute for consumers and companies. *Montreal Economic Institute*.

Reuters (2017). U.S. to impose 20 percent duties on Canadian softwood lumber. *Yahoo News*.

Roman, A. H. A., Foster, K. A., and Shook, S. (2006). Evaluation of the substitutability between U.S. and Canadian softwood lumber. *American Agricultural Economics Association 2006 Annual meeting, July 23-26, Long Beach, CA*.

Saadi, H. (2011). Price co-movements in international markets and their impacts on price dynamics. *Springer eBooks*, pages 149–163.

Salmon, M. (1982). Error correction mechanisms. *The Economic Journal*, 92:615.

Schwartz, A. E. and Wachter, S. (2022). COVID-19's impacts on housing markets: Introduction. *Journal of Housing Economics*, 59.

Sedjo, R. A. (2006). Comparative views of different stumpage pricing systems: Canada and the United States. *Forest Science*, 52:446–450.

Shook, S. R., SoriaJ.A, J. A., and Nalle, D. J. (2007). Examination of North American softwood lumber species substitution using discrete choice preferences and disaggregated end-use markets. *Canadian Journal of Forest Research*, 37:2521–2540.

Sims, C., Stock, J., and Watson, M. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58:113–144.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Institut Statistique de l'Université de Paris*, 8:229–231.

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9:449–460.

Sklar, A. (1996). Random variables, distribution functions, and copulas: A personal look backward and forward. *Lecture Notes-Monograph Series*, 28:1–14.

Smeekes, S. and Wilms, I. (2023). bootUR: An R package for bootstrap unit root tests. *Journal of Statistical Software*, 106.

Smith, V. H. (2011). Premium payments why crop insurance costs too much. *American Enterprise Institute*.

Spelter, H. (2006). If America had Canada's stumpage system. *Forest Science*, 52.

Stigler, M. (2020). Nonlinear time series in R: Threshold cointegration with tsDyn. *Handbook of Statistics*, 42:229–264.

Størdal, S., Lien, G., and Trømborg, E. (2021). Impacts of infectious disease outbreaks on firm performance and risk: The forest industries during the COVID-19 pandemic. *Journal of risk and financial management*, 14:318–318.

Suharsono, A., Aziza, A., and Pramesti, W. (2017). Comparison of vector autoregressive (VAR) and vector error correction models (VECM) for index of ASEAN stock price. *International Conference and Workshop on Mathematical Analysis and its Applications*.

Sun, C. and Ning, Z. (2014). Timber restrictions, financial crisis, and price transmission in North American softwood lumber markets. *Land Economics*, 90:306–323.

Sun, Y. and Wang, K. (2022). The implication of copula-based models for crop insurance and reinsurance under systemic risk. *Frontiers in Environmental Science*, 10.

Tang, X. and Laaksonen-Craig, S. (2007). The law of one price in the United States and Canadian newsprint markets. *Canadian Journal of Forest Research*, 37:1495–1504.

Tong, H. (1983). *Threshold models in non-linear time series analysis*. Springer New York, NY.

Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Clarendon Press.

Trivedi, P. K. and Zimmer, D. M. (2006). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1:1–111.

Tsay, R. S. (1986). Nonlinearity tests for time series. *Biometrika*, 73:461–461.

Tsiboe, F. and Turner, D. (2023). Federal Crop Insurance Program (FCIP) participation. *U.S. Department of Agriculture, Economic Research Service*.

Turner, J. A., Buongiorno, J., and Zhu, S. (2005). Effects of the free trade area of the Americas on forest resources. *Agricultural and Resource Economics Review*, 34:104–118.

USDA AMS (2024). Carcass beef grades and standards. *U.S. Department of Agriculture, Agricultural Marketing Service*.

USDA FSA (2024). Disaster assistance: Livestock Indemnity Program. *U.S. Department of Agriculture, Farm Service Agency*.

USDA RMA (2024). Federal crop insurance corporation: Summary of business report. *U.S. Department of Agriculture, Risk Management Agency*.

USITC (2019). Softwood lumber from Canada, Investigation Nos. 701-TA-566 and 731-TA-1342 (Final) (Remand). *U.S. International Trade Commission*.

USITC (2023). Softwood lumber products from Canada, Investigation Nos. 701-TA-566 and 731-TA-1342 (Review). *U.S. International Trade Commission*.

USITC Office of External Relations (2023). USITC makes determinations in five-year (sunset) reviews concerning certain softwood lumber products from Canada. *U.S. International Trade Commission*.

van Kooten, G. C. and Johnston, C. (2014). Global impacts of russian log export restrictions and the Canada–U.S. lumber dispute: Modeling trade in logs and lumber. *Forest Policy and Economics*, 39:54–66.

van Kooten, G. C., Zanello, R., and Schmitz, A. (2022). Explaining post-pandemic lumber price volatility and its welfare effects. *Journal of Agricultural & Food Industrial Organization*, 21:11–19.

von Cramon-Taubadel, S. and Goodwin, B. K. (2021). Price transmission in agricultural markets. *Annual Review of Resource Economics*, 13:65–84.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307.

Ward, C. E., Schroeder, T. C., and Feuz, D. M. (2017). Grid pricing of fed cattle: Base prices and premiums-discounts. *Oklahoma State University Extension*.

WHO (2020). Who director-general's opening remarks at the media briefing on COVID-19 – 11 march 2020. *World Health Organization*.

Xie, W., Nelson, B. L., and Barton, R. R. (2014). A Bayesian framework for quantifying uncertainty in stochastic simulation. *Operations Research*, 62:1439–1452.

Xu, Z., Hennessy, D. A., Sardana, K., and Moschini, G. (2013). The realized yield effect of genetically engineered crops: U.S. maize and soybean. *Crop Science*, 53:735–745.

Yang, L., Yin, Z., Gan, J., and Wang, F. (2020). Asymmetric price transmission of hardwood lumber imported by China after imposition of the comprehensive commercial logging ban in all natural forests. *Forests*, 11:200–200.

Yin, R. and Baek, J. (2004). The US–Canada softwood lumber trade dispute: What we know and what we need to know. *Forest Policy and Economics*, 6:129–143.

Yoders, J. (2021). 2021 4Q Cost Report: Tariffs doubled on Canadian lumber products. *Engineering News-Record*.

Zanello, R., Yin, S., Zeinolebadi, A., and Kooten, v. (2023). COVID-19 and the mystery of lumber price movements. *Forests*, 14:152–152.

Zhang, D. (2007). *The softwood lumber war: Politics, economics, and the long U.S.-Canadian trade dispute*. Earthscan.

Zhang, J., Hu, W., and Zhang, X. (2010). The relative performance of VAR and VECM model. *2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering*.

Zhou, M. and Buongiorno, J. (2005). Price transmission between products at different stages of manufacturing in forest industries. *Journal of Forest Economics*, 11:5–19.

APPENDIX

APPENDIX

A

SUPPLEMENTAL MATERIALS FOR CHAPTER 3

A.1 Supplemental Discussions

Before the introduction of the concept of cointegration, the idea of "error correction models" was used to describe deviation from some equilibrium relation (Lütkepohl 2005). Assume that the equilibrium relation between variable y_{1t} and y_{2t} is linear and described as $y_{1t} = \beta y_{2t}$, and changes in the variables depend on deviations from the previous period as well as previous changes in both variables, then we have,

$$\begin{aligned}\Delta y_{1t} &= \alpha_1(y_{1,t-1} - \beta_1 y_{2,t-1}) + \gamma_{11}\Delta y_{1,t-1} + \gamma_{12}\Delta y_{2,t-1} + \zeta_{1t} \\ \Delta y_{2t} &= \alpha_2(y_{1,t-1} - \beta_2 y_{2,t-1}) + \gamma_{21}\Delta y_{1,t-1} + \gamma_{22}\Delta y_{2,t-1} + \zeta_{2t}\end{aligned} \tag{A.1}$$

where Δ represents the first difference, and the error terms ζ_{1t} and ζ_{2t} are white noise.

This model is closely related to the concept of cointegration. To see this, let us assume both y_{1t} and y_{2t} are $I(1)$, i.e., integrated of order 1, then their first difference is stationary.

This also implies that $\alpha_i(y_{1,t-1} - \beta_i y_{2,t-1})$ terms are stationary. Therefore, in the matrix notion, we have,

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_i \Delta y_{t-1} + u_t \tag{A.2}$$

which is what so called the Vector Error Correction (VEC) model.

It is easy to see that Equation A.4 is just a Vector Autoregressive (VAR) model, rewrite it as the following,

$$y_t - y_{t-1} = \alpha\beta' y_{t-1} + \Gamma(y_{t-1} - y_{t-2}) + u_t \tag{A.3}$$

where[1]

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}, \; u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \; \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \; \beta' = (1, -\beta_1), \; \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

Rearrange the term and we have:

$$y_t = (I_K + \Gamma + \alpha\beta')y_{t-1} + \Gamma y_{t-2} + u_t \tag{A.4}$$

which is a VAR(2) representation.

A.2 Bayesian Approach

Bayesian methods have demonstrated great success in forecasting in many fields (Bessler and Kling 1986; Xie et al. 2014; Bergman et al. 2017). We implement a Bayesian approach to both VAR and VEC models to explore the possibility of better model performance. In this Appendix, we first briefly explain Baye's theorem, which is the fundamental building block for any Bayesian approach. We then present the prior selection process, then calculate the forecasting accuracy metrics for this method. We find that when forecasting manufactured wood products, the Bayesian approach only slightly improves the forecast in some cases. Therefore, considering its calculation complexity and low interpretability, we conclude that the Bayesian approach is not suitable for the context of this study.

[1]Note that β' is normalized in order to obtain a unique estimate. This implies that the cointegration relations are not unique while normalizing ensures a unique estimate.

A.2.1 Bayes' Theorem

Bayes' theorem is the foundation of any Bayesian modeling. In the Bayesian approach, nonsample information, or prior density is given in the form of a density. Here, we provide a detailed discussion of Bayes' Theorem and how prior distribution is selected in the applied empirical analysis.

Given two dependent Events A and B, the probability of Event B occurring conditional on Event A can be written as:

$$p(B|A) = \frac{p(B \cap A)}{p(A)}$$

Similarly, the conditional probability of Event A occurring given Event B is:

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Since $A \cap B = B \cap A$, we can rewrite Eq. 2 as:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

which is the well-known Bayes' rule. This rule can be extended to model estimation. Given dataset x and model parameters θ, the Bayes' rule can be written as:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

which is often simplified to:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

where $\propto$ means proportional to. We can drop the term $p(x)$ since the probability of the dataset does not depend on θ. The term $p(\theta|x)$ is the probability of model parameters computed conditional on the dataset x, in Bayesian terms, the posterior distribution. The term $p(\theta|x)$ represents the likelihood function, which gives the probability of the data conditional on the parameters. Lastly, $p(\theta)$ is the prior distribution, where careful identification is needed based on background knowledge. In summary, the posterior distribution is proportional to the likelihood function multiplied by the prior distribution.

A.2.2 Implementation of BVEC Model

The Bayesian method imposes prior beliefs to achieve a parametrically parsimonious model while still preserving rich patterns between time series, which could be more appropriate when modeling nonstationary data. In the generalized VAR model, the number of coefficients to be estimated is $n^2p + nk$, where n is the number of variables included in the VAR model, p is the lag order, and k is the number of deterministic variables. Even with a modest selection for variable and lag selections, the VAR model is still very likely to be overfitted. The lack of degrees of freedom in a Bayesian framework is no longer an issue since the model would include relevant variables along with the prior information that reflects one's knowledge about the value of the coefficient, which minimizes the mean squared error loss function (Litterman 1986).

The essence of the Bayesian method lies in incorporating prior beliefs to reduce the influence on certain coefficients. Litterman (1986) suggests that instead of just relying on economic theory to extract useful information, one can focus on the prior that gives the best out-of-sample forecast. If some information is useful for the future, it is likely to be useful for the past. Compared to the prior information that only focuses on the existing data, information that is spread across a wider spectrum is better. We adopt the prior that Litterman introduces, namely Minnesota (Litterman) price, for this analysis. Such prior assumes the mean coefficient of 1 for the first own-lag and a prior mean of 0 on all other variables. This implies an assumption of a univariate random walk model for each of the time series. The prior variance is given by

$$v_{ij,L} = \begin{cases} (\frac{\lambda}{L})^2 & \text{if } i = j \\ (\frac{\lambda\theta\sigma_{ii}}{L\sigma_{jj}})^2 & \text{if } i \neq j \end{cases} \tag{A.5}$$

where λ is the standard deviation of the first own-lag, a large λ implies prior information is diffused. θ represents the tightness of other variables, and is bounded between 0 and 1. The larger the θ, the more information is given by the other variables. L is the number of lags, which is 2 selected using Bayesian information criterion. Lastly, σ_{ii} is the i^{th} diagonal element of Σ_u.

We follow Litterman (1986) suggestion to choose the pair of the two hyperparameters, λ and θ, that provides the best out-of-sample forecast. To achieve this, we conduct a grid search across the range from 0 to 1 with a granularity of 0.1 for each hyperparameter. We compare the forecasting performance of a total of 100 possible parameter pairs for BVEC

models. The corresponding accuracy metrics are presented as follows.

A comparison of forecasting accuracy across all three models for each of the wood products is presented in Table A.1. The insignificant differences between the Bayesian approach and the standard models do not provide convincing evidence to conclude that a Bayesian approach with Minnesota priors provides a better forecast. Since priors are often subjective and Bayesian analyses are more complex and computationally challenging, we think the standard VEC model is the best approach to forecast SPF softwood lumber prices given our model setting.

Table A.1: Forecasting Accuracy Comparison with Bayesian Models

Variable	Model	RMSE	MAPE	Theil's U
Spruce-Pine-Fir	VAR(2, 0)	0.475	0.066	0.087
	BVAR(2, 0)	0.469	0.065	0.086
	VECM(2, 0)	0.452	0.061	0.083
	BVEC(2, 0)	0.445	0.060	0.081
Southern Yellow Pine	VAR(2, 0)	0.477	0.063	0.090
	BVAR(2, 0)	0.480	0.063	0.091
	VECM(2, 0)	0.473	0.065	0.089
	BVEC(2, 0)	0.473	0.065	0.089
OSB Sheathing	VAR(2, 0)	0.740	0.108	0.140
	BVAR(2, 0)	0.757	0.110	0.144
	VECM(2, 0)	0.832	0.129	0.158
	BVEC(2, 0)	0.832	0.129	0.158