From Trees to Lumber:
Essays on Forest Management, Logging and Lumber Price

by

Mingtao He

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

Auburn, Alabama
December 11, 2021

Keywords: Economic transformation, logging employment, profitability, Directed Acyclic Graph, Machine Learning, Deep Learning

Approved by

Yaoqi Zhang, Chair, Professor of School of Forestry and Wildlife Sciences
Daowei Zhang, Professor of School of Forestry and Wildlife Sciences
Brian K. Via, Professor of School of Forestry and Wildlife Sciences
Mathew Smidt, Research Forester of Southern Research Station, USDA Forest Service
Wenying Li, Assistant Professor of Department of Agricultural Economics and Rural Sociology, Auburn University

Abstract

The forestry sector includes silviculture, forest management, logging, and wood utilization. Forestry has been experiencing dramatic transformation due to technological advances, changing markets of labor, capital, and goods and service. Recent Covid-19 pandemic is another example of the impact of globalization, including the lumber. This dissertation chose the transformation of forest management in China, the labor market and profitability, and the lumber market in the USA to understand the forestry sector.

The developing countries, for example, China, have experienced an unprecedented transformation of the rural society and livelihoods. Self-subsistence forest management has transitioned to more business-oriented management. Understanding what attributes are driving the transformation of the households and the forest management will help the policy makers to design more specific policies to facilitate the transformation. The Logit model helped to identify factors that are significantly correlated with the transformation of traditional peasant households to three emerging household categories namely Forestry Cooperative (FC), Family Forestry Farm (FFF), and Forestry Specialized Household (FSH), using household survey data from the seven provinces in China in 2016.

In developed countries, for example, the United States, have experienced technological advancements, policy changes, parcelization of forestland, business cycle, and the change of relative costs of factors, which have a significant impact on the forest industry, especially the logging industry. The official databases related to the logging industry aids in the quantitative analysis of employment and profitability in the logging industry in recent decades in the U.S. It was found that employment in the U.S. logging industry has been declining over the past several decades. An investigation of the drivers of employment in the U.S. logging industry from 1997 to 2019, using Directed Acyclic Graph (DAG) and Forecast Error Variance Decomposition (VD), identified the trends in the industry.

Lumber product is one of the primary products coming from the logging industry. Firms engaged in producing, processing, marketing or using lumber and lumber products always take positions in the lumber futures markets. The accurate prediction of future prices can help companies and investors hedge risks and make correct market decisions. Our novel approach utilized the Google Trends Index related to lumber prices as predictors and employed Machine Learning and Deep Learning Models to nowcast lumber futures price, indicating both the methods have higher predictive power.

Acknowledgments

My deep gratitude and appreciation go first to my advisor, Dr. Yaoqi Zhang, for his valuable suggestions and support for my Ph.D. study during the past four years and a half. Dr. Zhang has been supportive since the first day of my journey when I came to Auburn as a visiting student in 2016. He gave me profound guidance on my research and freedom to develop my own ideas; he gave me care and encouragement when I got stuck; and he helped me develop the economic mindset and critical thinking skills, which will go a long way towards my career and life. Without his guidance and persistent help, this dissertation would not have been possible.

My appreciation also extends to my committee members, Dr. Wenying Li, Dr. Mathew Smidt, Dr. Brian K. Via, and Dr. Daowei Zhang, for their encouragement, insightful comments, and suggestions. I am also very grateful for Dr. Samir Huseynov's serve as the University Reader. Also, I would like to thank my consultant at the Miller Writing Center, Ms. Joshita Majumdar, for her help on improving my writing skills and effectively convey my thoughts.

Finally, I would like to thank my family and my best friend, Dr. Yang Hu, for their support and unconditional love.

Auburn Forever! War Eagle!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average Model
ARS	Annual Refiling Survey
BDFT	Board Foot
BIC	Bayesian Information Criterion
CART	Classification and Regression Tree
CLT	Cross-laminated Timber
CNN	Convolutional Neural Network
Conv-1D	1 Dimensional Convolution
DAG	Directed Acyclic Graph
DIY	Do It Yourself
DL	Deep Learning
EIO-LCA	Economic Input-Output Life Cycle Assessment

FC	Forestry Cooperatives
FFF	Family Forestry Farms
FIA	Forest Inventory and Analysis
FORSIM	Forest Simulation Model
FSH	Forestry Specialized Households
HQIC	Hannan-Quinn Information Criterion
H-T	Harris-Tzavalis
I/O	Input-Output
LASSO	Least Absolute Shrinkage and Selection Operator
LSD	Least-Squares Deviation
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
mbf	Thousand Board Feet
MDF	Medium Density Fiberboard
ML	Machine Learning
MSE	Mean Squared Error
MWR	Multiple Worksite Report

NAICS	North American Industrial Classification System
NIPF	Nonindustrial private forest
OES	Occupational Employment Statistics
OR	Odds Ratios
OSB	Oriented Strand Board
PC	Peter-Clark
PCA	Principal Component Analysis
PP	Phillips-Perron
PPI	Producer Price Index
QCEW	Quarterly Census of Employment and Wages
QCR	Quarterly Contributions Report
QWI	Quarterly Workforce Indicators
RNN	Recurrent Neural Network
SARIMA	Seasonal Autoregressive Moving Average Model
SARIMAX	Seasonal Autoregressive Moving Average Model with Exogenous Variables
SMAPE	Symmetric Mean Absolute Percentage Error
SOC	Standard Occupation Code
SVM	Support Vector Machine

TPO	Timber Product Output
UI	Unemployment Insurance
VAR	Vector Autoregression
VD	Variance Decomposition
VMA	Vector Moving Average
XGBoost	Extreme Gradient Boosting

Chapter 1. Introduction

1.1 Background

Forestry sector includes silviculture, forest management, logging and wood utilization. Forestry has been experiencing dramatic transformation due to the technological advances, changing markets of labor, capital, and goods and service. Recent Covid-19 pandemic is another example of the impact of globalization, including the lumber. This dissertation chose the transformation of forest management in China, the labor market and profitability and the lumber market in the USA to understand the forestry sector.

In the past few decades, an unprecedented transformation has taken place in the global rural society, especially in the developing countries, for example, China. China's economic reform was initiated in the rural areas in the late 1970s and then extended to the industries and cities. Over the past 40 years, fast industrialization, rapid growth in income, and significant infrastructure improvements took place that was unprecedented in the Chinese history. One of the effects of the transformation is to enable the rural population, known as ‘peasants’, to change their traditional lifestyles and livelihoods. This transformation also extends to forest management. Self-subsistence forest management has transitioned to more business-oriented management, typically the Forestry Cooperatives (FC), Family Forestry Farms (FFF), and Forestry Specialized Households (FSH) identified in this dissertation. Understanding what attributes drive the transformation of households and forest management will help policymakers design more specific policies to facilitate the transformation.

In developed countries, for example, the United States, with technological advancements, policy changes, parcelization of forestland, economic cycle, and the change of relative costs of factors, the logging industry has undergone dramatic changes in the past few decades. Employment in the logging industry is concentrated in the West and the South, seeing the most significant decline in employment. Since logging wages have increased at about the rate of inflation and the interest rate remains low, logging firms continue to mechanize. On one hand, changes in employment and mechanization affect the cost of logging and the profit, while on the other, firm profitably affects employment and mechanization. It is necessary to provide a quantitative analysis of U.S. logging industry data and present trends in employment and profitability over the last 20 to 30 years. The results will help us better understand the factors affecting the US logging industry.

Employment in the U.S. logging industry has declined in recent years. The West and South, the most concentrated areas of logging, have dropped the most. Employment in the logging industry has fallen by an average of 2% per year since 1997. The labor market in the logging industry depends on both demand and supply. From the demand side, the total removal of forest resources is related to the degree of mechanization and the economic cycle, particularly the housing market and building permits, pulp and paper prices. From the supply side, the most critical variables are relative wages with competitive sectors. However, these economic factors are all interrelated. Therefore, it is necessary to establish a contemporaneous causal relationship among these multiple variables to investigate the employment drivers.

Lumber product is one of the leading products coming from the logging industry. Firms engaged in producing, processing, marketing, or using lumber and lumber products always take positions in the lumber futures markets. Since the COVID-19 pandemic, the lumber futures price has experienced colossal volatility. The average opening price from May 2011 to January 2020 was $337 per thousand board feet (mbf), while the average opening price from February 2020 to May 2021 was $698 mbf. The lumber price reached its highest point in the past ten years on May

7th, 2021, with $1,677 mbf. Therefore, there is an urgent need to find a reliable method to predict the lumber futures price, which would help enterprises and investors hedge risks and make correct decisions in the market.

1.2 Scholarly contributions

This dissertation firstly investigates the direction and magnitude of critical socioeconomic factors affecting the transformation of the households in China using the FC, FFF, and FSH. Secondly, it provides a quantitative analysis of the US logging industry data, and presents employment and profitability trends. Thirdly, this dissertation investigates the drivers of employment in the U.S. logging industry, using DAG and VD. Finally, this dissertation predicts the lumber futures price using different models. In particular, the contributions include:

- Help better understand what economic, social, and livelihood attributes drive the transformation of households and how the traditional forest management shifts to business management and shapes the rural society in China.

- Add both firm-level data from previous research with federal and state-level data, providing additional evidence for industry-level trends and relationships, helping better understand the factors influencing the U.S. logging industry. Also, present a feasibility method for estimating the profit for the logging industry.

- For the first time, the utilization of the DAG approach disclosed the contemporaneous causal relations among employment, wage, mechanization, logging product prices, and production levels, and VD to analyze the dynamic relationship among variables. Thereby, bridge the gap in the literature by exploring contemporaneous causal relationships between employment in the logging industry and different economic variables, and help to understand better the effects of these factors on employment in the short and long run as well.

- Offer a fresh and novel perspective on nowcasting the lumber price using Google Trends Index through Machine Learning Models (SVM, Random Forest, XGBoost, and CART) and Deep Learning Models (ANN, RNN, and CNN).

1.3 Dissertation structure

This dissertation consists of six chapters. Chapter 1, introduced the research background and scholarly contributions. Chapter 2, explored the factors contributing to the transformation of households and traditional forest management using household survey data in seven provinces in China in 2016. The age and education of the household heads, income, the holding areas of cropland, ecological forest, forestland, leasing forestland, and legally contracted forestland, and their located provinces were found to be statistically significant in transforming the household's forest management. The factors that drive the transformation to various ownership types showed some variations as well. The findings can help us better understand not only the transformation of forest management but also the rural economy and society in general. The results have policy implications on how to facilitate forest management transformation further.

Chapter 3, analyzed the logging industry in recent decades in the U.S. based on Occupational Employment Statistics (OES), Quarterly Workforce Indicators (QWI), Quarterly Census of Employment and Wages (QCEW), and Timber Product Output (TPO) reports. The logging industry has been experiencing reduced employment and the implications of the aging workforce. This might be due to increased productivity from the technological advancement of mechanization, and reduced demand for logging. Economic Input-Output Life Cycle Assessment (EIO-LCA) model was applied to estimate the profitability of the entire industry at the state level. It was found that the reduced demand and increased operating costs lead to poor profitability and a wave of closures of logging firms, but also accelerating adjustment in the logging industry. Some severe challenges for the logging industry were identified, including lack of practical monitor tools for the entire sector, structural shortage of labor, and rising operating costs.

Chapter 4, investigated the drivers of employment in the U.S. logging industry from 1997 to 2019, using Directed Acyclic Graph (DAG), which was first applied to disclose the contemporaneous causal relations among employment, wages, mechanization, logging production level and product prices. Based on the Vector Autoregression (VAR) model and DAG, Forecast Error Variance Decomposition (VD) was used to examine their dynamic relationships. The results showed that logging production level affects employment directly and indirectly. An increase (decrease) in the level of logging production directly increases (decreases) wages, followed by an increase (decrease) in employment. The results of VD showed employment in the logging industry is most prominently explained by the production level (highest 52.0% at horizon 1-year), followed by the wage (highest 42.0% at horizon 20-year). In contrast, capital and product price have a limited influence on employment.

Chapter 5, explored whether Internet browsing habits can accurately nowcast the lumber futures price. The predictors were Google Trends Index related to lumber prices. By employing both Machine Learning and Deep Learning methods, it was shown that despite the high predictive power of both methods, on average, Deep Learning Models can better capture trends and provide more accurate predictions than the Machine Learning Models. The Artificial Neural Network model is the most competitive model, followed by the Recurrent Neural Network model. Finally, Chapter 6 concluded the dissertation with some possible future research directions.

Chapter 2. From peasant to farmer: transformation of forest management in China

2.1 Introduction

Chinese economic reform was initiated in the rural areas in the late 1970s, and then extended to industries and cities. Over the past 40 years, fast industrialization, rapid growth in income, and significant improvements in infrastructure have been taking place that are unprecedentedly in Chinese history. The transformation is not only that of a centralized planning economy becoming a market-based economy, but also the economic structure from primarily an agricultural economy to a more industrial and services based economy. However, 40% of the Chinese population still live in rural areas and the long-standing rural–urban inequalities still remain (National Bureau of Statistics of China 2019). Farming is also at a disadvantaged position due to the manipulated low prices of agricultural products in order to support industrialization (Gustafsson and Shi 2002; Long et al. 2011), and small amount of the land that is available for the large population. The increasing labor costs and the evenly distributed land make agriculture and forest management less competitive in recent decades.

The transformation of rural China is essentially from a traditional subsistence economy to market economy described by Schultz (1964) in the middle of last century when many developed countries were in similar transformations. Both the central and local governments have tried hard to improve road access, the quality of life for rural populations and the rural environments in recent decades (Liu et al. 2013, 2016; Long et al. 2010; Wen 2018; Zhou et al. 2018). Financial capital has been mobilized to rural development, such as micro-financing in forestry (Zhou et al. 2016), land transfer is encouraged, and labor is more specialized to boost agricultural and entrepreneurial advancements. All of these changes together with increasing income and information access have dramatically changed the rural lifestyle, promoted technology-driven agriculture, and new economic and community organizations (Chen et al. 2010; Long et al. 2011).

One of the impacts of the rural–urban transformation is enabling the rural population, known as 'peasants', to change their traditional lifestyles and livelihoods. In rural China, 'peasant' is not just an occupation, but also a symbol of a certain social class having a traditional and self-sufficient mode of agricultural production and poorly educated (Cohen 1993; Schneider 2015; Zhang and Donaldson 2010). Because of the existing dual agricultural and non-agricultural registered residence system, peasants can move to cities as itinerant populations, however, they cannot enjoy the same benefits and services as the urban or non-agricultural residents do (Cheng and Selden 1994; Fan 1999). However, the policy allows some peasants to convert to commercial farmers (Zhang and Donaldson 2010). Unlike 'peasant', 'farmer' is more of an occupation or profession, not connoting to any social class, identity, or status. Farmers can be either agricultural or non-agricultural registered residents and their products are more for sale than self-sufficiency. Thus, 'farmer' is a transformed generation of peasant with a higher level of education, technological skills, management ability, income, and social status. Most importantly, their formal occupation is based in agriculture (Wei and Liu 2013).

This transformation also extends to forest management, particularly when most rural households obtained forestland use rights after the second round of collective forest tenure started in 2003 (Wang et al. 2007), and where forestry is important in the rural economy (Wang et al. 2014, 2015). Forest management might experience faster transformation than agriculture as timber is more for sale and timberland is less equally distributed (Huang et al. 2019). Along with issuing forest tenure certificates to households, the government launched other reforms including promotion of new forest management entities, namely Forestry Cooperatives (FC), Family

Forestry Farms (FFF), and Forestry Specialized Households (FSH) (Table 1). We have good reason to believe that they more represent the new types of forest management with more business-orientation.

Table 1. Characteristics of the forestry cooperative, family forestry farm, and forestry specialized household.(State Forestry Administration 2017)

Characteristics	Emerging forestry entities		
	Forestry Cooperative (FC)	Family Forestry Farm (FFF)	Forestry Specialized Household (FSH)
Objective	Better access to market and financial resource; increase market power and returns through pooling land, labor and capital	Increased economy of scale and promote forest management mainly through self-employment	Increased economy of scale, promote rural employment and division of labor
Scale	Large	Medium	Large
Registration	Industrial and commercial department	Industrial and commercial department	Forestry department
Legal status	Specialized cooperative	Individual business	Individual household

The 2008 financial crisis negatively affected non-agricultural employment, causing some peasants not satisfied in cities to return to villages with a broader perspective, knowledge, skills and capital than those who stayed at home. Some of them chose to manage forestry as a business and became farmers who are more capable of accessing markets, financial support from the government, and using machine. Other peasants might be slower in the transformation. The differences are a resulted of entrepreneurship, business involvement, specialization, and scale (Zhu et al. 2013).

The Forestry Cooperative (FC) refers to an economic mutual aid organization, which is voluntarily and jointly organized by households or similar providers of forestry production and management services (Zhang et al. 2014). FC was initiated with the law related with China's

Farmers' Professional Cooperatives in 2007. The law required cooperatives to register at the Industrial and Commercial Department. The objectives of promoting these cooperatives are to strengthen the rural economy, link small households to large markets, and promote better forest management (Bijman and Hu 2011; Deng et al. 2010; Yang et al. 2013). The cooperatives enjoy better access to land pool and financial resources from the government (Zhang and Zhi 2010).

Family Forestry Farms (FFF) have been listed for priority development by the central government since 2008 and were designated as the main agricultural and forestry identity in 2013. After the introduction of supporting policies in Liaoning Province in 2014, the policies supporting FFF were extended nationwide and these entities began to emerge. FFF refers to a market and profit-oriented economic enterprises invested in and operated mainly by family members. The main income of the household should come from the farm, and the scale of the farm should exceed a certain size. FFF has the qualifications of a market entity, such as self-employed and sole proprietorship enterprises (He et al. 2017; Huang and Liang 2018; Shen and Shen 2018).

Forestry Specialized Households (FSH) have been promoted by the central government for more than a decade (Li 2016). While the FC mainly employs the labors from their members and FFF is mainly dependent on their family members, FSH is characterized by the management of large-scale forestland through employed labor.

Such transformations have important implications on forest management, household income, rural employment, and many other socioeconomic attributes. Garnevska et al. (2011) argued that a community's likelihood to form cooperatives is impacted by prevailing legal environment, political leadership, public policy, and the community's rational behavior towards community feelings. Zhang and Donaldson (2010) found that well-developed markets for agricultural products, labor, and land allow rural households to shift from peasant to farmer. In a similar study, Omiti et al. (2009) found that transformation of rural households is augmented by market proximity, product price, and availability of market information. Government land administration and agricultural spending policies have also been shown to strongly contribute to rural household transformation (Sitko and Jayne 2014).

The transformation and the driving socioeconomic factors have not been widely investigated. Our study aimed to investigate the direction and magnitude of important socioeconomic factors affecting the transformation of the households in China using the FC, FFF and FSH, which we identified them as transformed households or farmers. We believe the results of the study will help us better understand what economic, social, and livelihood attributes are driving the transformation of households and how the traditional forest management shifts to business management and shapes the rural society in China. The results will help the government to design more specific policies and programs to facilitate the transformation. The results can be extended into many other countries experiencing similar transformations.

2.2 Empirical model

The transformation for the households and their management types can be explained, at least in part, through the households' economic and social attributes, as well as their land characteristics. In order to identify the characteristics of traditional households (peasant) contributing to their transformation to the new forms (farmers), we adopted a logistic regression framework. We assumed that the individual attributes of the 'peasants', including age and education of household head, and household level attributes including total income, family size, cropland area, the composition of the forestland and land tenure, as well as location are likely to affect the transformation likelihood.

The response variable in our study was whether a 'peasant' household transformed to a 'farmer' household: $Y = 1$ if a household is a FC, FFF or FSH and $Y = 0$ if a household is a traditional peasant household. The explanatory variables are represented by vectors consisting of characteristics of the household and household head.

The probability that a household would transform from a traditional household to a new type of management is given by Eq. (1):

$$\mathrm{P}_{ik}(Y_k = 1 \text{ if transformed, 0 otherwise} | X_{ik}) = \frac{e^{\beta_{0k}+\beta_{ik}x_{ik}}}{1+e^{\beta_{0k}+\beta_{ik}x_{ik}}} \quad (1)$$

where Y_k is the dependent variable with a value of 1 for the household transforming to the k^{th} type of transformed household. Three regressions were made when $k = 1, 2, 3$ to understand household becoming a FC, FFF and FSH, respectively, and another regression combining all transformed household when $k = 4$) and 0 for nontransformation. x_{ik} is the i^{th} explanatory variable that relates to the k^{th} type of transformed household. β_{ik} is the i^{th} estimated parameter that relates to the k^{th} type of transformed household. The parameters were estimated using maximum likelihood. (Cramer and Ridder 1988; Sheikh et al. 2003).

Eq. (1) can be solved for β_{ik}, to derive Eq. (2)

$$\beta_{ik} = log\left(\frac{P_{ik}(Y=1)}{1-P_{ik}(Y=1)}\right) - log\left(\frac{P_{ik}(Y=0)}{1-P_{ik}(Y=0)}\right) \quad (2)$$

And the Odds Ratio (OR_{ik}) can be defined as

$$OR_{ik} = \frac{\frac{P_{ik}(Y=1)}{1-P_{ik}(Y=1)}}{\frac{P_{ik}(Y=0)}{1-P_{ik}(Y=0)}} \quad (3)$$

Thus, Eq. (2) reduces to Eq. (4):

$$e^{\beta_{ik}} = OR_{ik} \tag{4}$$

Following Xie et al. (2014), we included the square of AGE to check the monotonicity of the transformation probability with the household's age. Some physical characteristics including cropland area, the composition of the forestland and the land tenure are likely to have an impact on the transformation. The composition of forestland can be classified by economic forest, timber forest and bamboo forest in China. Households manage their timber or bamboo forest less intensively than economic forests, which is a forest with main purpose of producing fruit, edible oil, industrial raw materials and medicinal materials, including fruit, tea and mulberry trees. A larger proportion of timber forest (the type of forest mainly for timber production) may slow down the transformation. The composition of forestland can be also classified by ecological forest and commercial forest. More ecological forest may have a negative impact on the transformation.

There are significant differences in the physiographic and policy attributes among the Chinese provinces. For example, Hunan is located in central China and has high forest coverage; Shaanxi is located in north China with only 43% forest coverage; Gansu is located in northwest China and is one of the least developed provinces and part of it is desert area, but Gansu introduced a series of policies to promote FC, FFF and FSH. Hence, we inserted province-level variables in our model.

2.3 Data collection

The socioeconomic data from 2009 to 2017 for this study were collected from and with the help of the State Forestry Administration of the following provinces: Liaoning, Fujian, Jiangxi, Hunan, Yunnan, Shaanxi and Gansu. We applied stratified sampling method to randomly select seven provinces in seven regions in China (Figure 1).

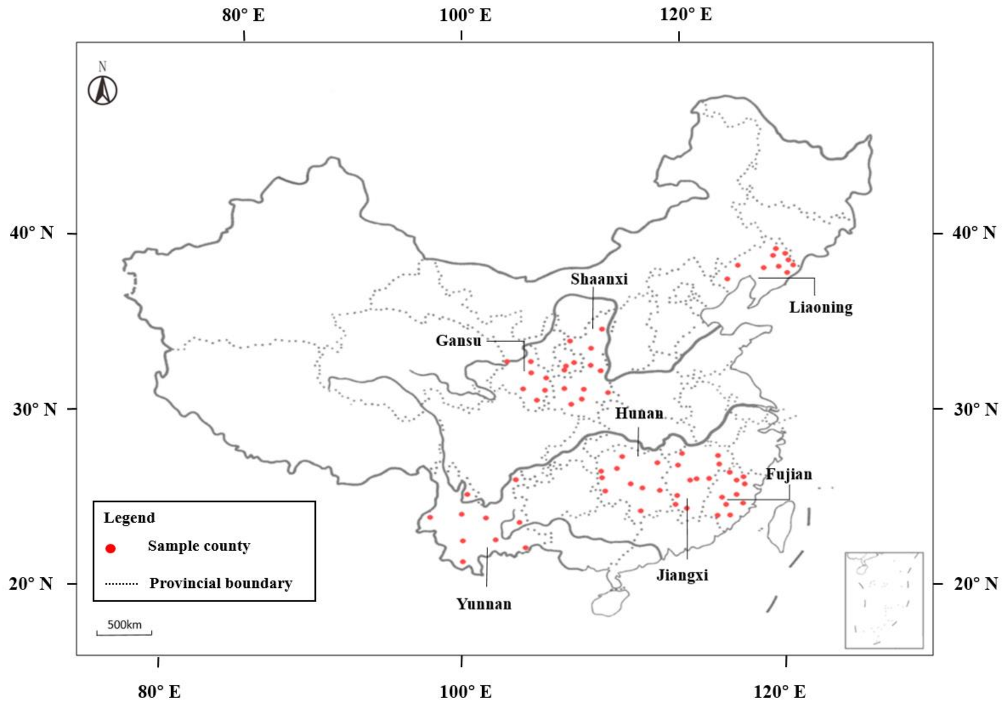

Figure 1. Distribution of sample counties across the Liaoning, Fujian, Jiangxi, Hunan, Yunnan, Shaanxi and Gansu Provinces in China.(State Forestry Administration 2012)

Each province was divided into three regions. Then stratified sampling was applied to select sample counties, based on the status of forest resources and socioeconomic conditions. 10 counties were selected from each province. In each sample county, a symmetric equidistance method was applied to select sample townships and towns, according to economic conditions. Five townships were selected from each county, and a village was randomly selected from each township. In each village, the symmetric equidistance method was applied to select 10 households based on the household registration list (State Forestry Administration 2010). Altogether, these surveys were designed to cover 3500 households in 350 villages from 70 counties of 7 provinces. Field surveys were conducted with the same households every year since 2009. In this study, we only use the 2016 dataset. We omitted households with missing or incomplete information. Thus, our sample size was 3487 instead of 3500.

The summary of the socioeconomic variables attributes of the households is presented in Table 2. Among the 3487 households, there were 180 (5%), 26 (1%), 146 (4%) and 3135 (90%) households who joined FCs, established FFFs, became FSHs and did not change, respectively. Thus, there were 352 transformed households in 2016 in our sample. We separately estimated the probability that a household would transform from a traditional household to a transformed household. Therefore, in the logit model of FCs, FFFs households, FSHs and transformed households, the sample sizes are 180, 26, and 146 and 352, respectively. Jiangxi had the largest number of households participating in FCs and FSHs among these seven provinces in our sample, while Liaoning has the largest number of FFFs. Jiangxi also had the largest number of transformed households. Gansu, the least developed province of the seven provinces, has the largest number of traditional peasants (Figure 2).

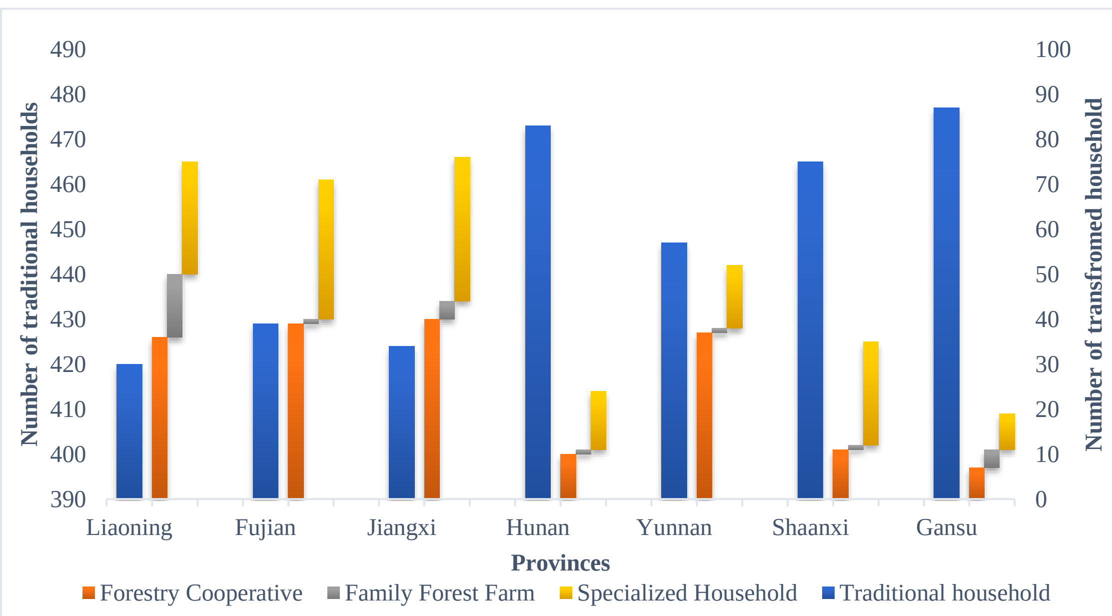

Figure 2. Distribution of traditional and transformed households among seven different provinces, China.

Table 2. Descriptive statistics of household characteristic variables.

Variables	Definition	Total households	Traditional households	Transformed household			
				Forestry cooperative	Family forestry Farm	Forestry specialized household	Total
–	Sampled households #	3487	3135	180	26	146	352
AGE	Household head age (years)	55.28 (10.85)	55.36 (10.94)	53.12 (9.76)	56 (11.32)	56.08 (9.84)	54.56 (9.99)
EDU	Education of household head (Elementary & less = 1; Junior high = 2; Senior high/Technical = 3; Junior college or above = 4)	1.77 (0.76)	1.76 (0.75)	1.92 (0.75)	2.12 (0.86)	1.90 (0.87)	1.93 (0.81)
INC	Total household income (1000 Yuan*)	82.63 (417.8)	70.53 (370.2)	153.7 (444.6)	94.40 (72.26)	252.3 (974.7)	190.2 (704.7)
POP	Family size (#)	4.67 (2.03)	4.63 (2.04)	4.92 (1.80)	5.08 (1.50)	5.06 (2.08)	4.99 (1.91)
CRP	Cropland area (ha)	0.50 (1.03)	0.48 (0.74)	0.79 (3.27)	0.73 (0.73)	0.48 (0.54)	0.65 (2.37)
ECO	Ecological forestland/total forestland (%)	0.48 (0.47)	0.50 (0.47)	0.27 (0.38)	0.42 (0.46)	0.35 (0.43)	0.32 (0.41)
TIM	Timber forestland & bamboo/total forestland (%)	0.55 (0.47)	0.54 (0.47)	0.57 (0.45)	0.45 (0.49)	0.65 (0.44)	0.59 (0.45)
FOR	Total forestland (ha)	6.02 (18.80)	4.49 (9.88)	17.24 (60.23)	10.09 (16.36)	24.11 (36.34)	19.56 (49.32)
TRS	Forestland transfer (ha)	0.69 (8.39)	0.28 (2.08)	4.74 (31.97)	0.83 (4.22)	4.42 (17.22)	4.32 (25.42)
CTS	Legally contracted forestland/total forestland (%)	0.50 (0.49)	0.49 (0.49)	0.50 (0.468)	0.81 (0.40)	0.62 (0.46)	0.57 (0.47)

Notes: Each of the values in the table is the mean of 3487 observations; *1Chinese Yuan = 0.144 US $, as of Dec 31, 2016; **15mu=1ha

The household characteristics are also summarized in Table 2. The average age of the household heads was 55 years old. Households that joined FC were statistically significant younger. Most of the household heads had finished elementary school, 16% of them had finished high school or higher education, and only 70 of them, accounting for 2%, had finished university. The transformed households were more educated: 20% of them finished high school or higher education and 5% of them had education in university, while for the traditional counterparts, these were 15% and 2%, respectively. The average family size was 4.67. The average income of traditional households was only 69,950 yuan (1 Chinese Yuan = $0.144 US as of Dec 31, 2016) and 16,096 yuan per capita, which was lower than that of transformed households (190,200 yuan and 49,985 yuan per capita) and FSH (252,300 yuan and 74,410 per capita) (Table 2).

On average, households surveyed had 0.50 ha cropland, and 0.11 ha cropland per capita, which is slightly higher than the national average of 0.095 ha per person. In our sample, ecological forests accounted for 47% of the total forest area. Specifically, ecological forests, which refer to forests with ecological and social benefits as the main functions and are delineated according to the regulations of the central or the provincial governments, accounted for only 27%, 42% and 35% of the total forest area in the samples of FC, FFF and FSH, respectively. The proportion of ecological forests among transformed households was 32%, which was lower than that of traditional households (50%). Classified by forest uses, 55% of the forest area was comprised of timber forests, forests with the main purpose of cultivating and providing timber or bamboo, including arboreal forests (45%) and bamboo forests (10%). Notably, FSHs have the highest proportion of timber forest area (65%), and the proportions of timber forest area in households joining FC and FFF were 57% and 45%, respectively.

Among the households surveyed, they had 6.02 ha of forestland on average. The transformed households had 19.56 ha of forestland, which was much greater than that of the traditional households with 4.50 ha on average: FSHs and FCs were larger scale with 17.25 ha and 24.11 ha, respectively, and FFFs had 10.09 ha forestland on average. The average area of forestland transfer was 0.69 ha. Notably, for households joining FCs and the FSHs, their areas of transfer of forestland were 4.74 ha and 4.42 ha, respectively, suggesting that FC and FSH had much larger scales of forestland as they had leased in some forestland from others. Also, in our sample, 50% of the forestland had signed their legal contracts with the village, which could make them have more security of forestland tenure.

2.4 Results and discussion

The likelihood ratio Chi square of the fitted models were 149.24, 51.56, 212.93, and 314.01, for FC, FF, and FSHs, and the total households, respectively, all of which were strongly significant ($p < 0.01$) (Table 3). It turns out that the models fitted well with the data. The Pseudo -R2 were 0.107, 0.171, 0.178 and 0.138, respectively. The McKelvey and Zavoina's -R2 were 0.262, 0.383, 0.260 and 0.347, respectively. The variability of dependent variable explained in these models were not high, perhaps because a few potential factors were missing, such as the level of households' forestry skills, the quality of forestlands, local climatic conditions and the frequency of disasters. Of the 16 variables in the study, we found nine were statistically significant affecting the likelihoods of transforming to FC, six to FFF, and nine to FSHs. Overall, 12 factors were statistically significant in facilitating households to become FC or FFF, or FSH.

Table 3. Results of the logit model on transformed households.

Variables	Forestry cooperative	Odds ratios	Family forestry farm	Odds ratios	Specialized household	Odds ratios	Total transformed household	Odds ratios
AGE	0.093	1.10	− 0.047	0.95	0.13*	1.14	0.086*	1.09
	(0.065)		(0.13)		(0.08)		(0.049)	
AGE^2	− 0.001	1.00	0.0005	1.00	− 0.001	1.00	− 0.0008*	1.00
	(0.0006)		(0.0012)		(0.001)		(0.0004)	
EDU	0.13	1.14	0.51**	1.67	0.15	1.16	0.16**	1.17
	(0.11)		(0.25)		(0.12)		(0.08)	
INC	0.00	1.00	0.0001	1.00	0.0002**	1.0002	0.0001	1.00
	(0.0002)		(0.0005)		(0.001)		(0.0009)	
POP	0.055*	1.06	0.15**	1.16	0.08**	1.08	0.076***	1.08
	(0.033)		(0.06)		(0.035)		(0.028)	
CRP	0.007**	1.01	0.0001	1.00	− 0.011	0.99	0.006**	1.01
	(0.003)		(0.014)		(0.008)		(0.003)	
ECO	− 0.67***	0.51	− 0.85	0.43	− 0.77***	0.46	− 0.75***	0.47
	(0.23)		(0.54)		(0.27)		(0.17)	
TIM	− 0.23	0.80	− 0.55	0.58	0.057	1.06	− 0.13	0.88
	(0.18)		(0.49)		(0.23)		(0.14)	
FOR	0.001***	1.00	0.002***	1.00	0.003***	1.00	0.003***	1.00
	(0.0004)		(0.0007)		(0.0004)		(0.0003)	
TRS	0.003*	1.00	− 0.0005	1.00	0.004***	1.00	0.003***	1.00
	(0.0015)		(0.003)		(0.001)		(0.001)	
CTS	0.24	1.28	1.39**	4.01	0.78***	2.19	0.50***	1.65
	(0.20)		(0.68)		(0.24)		(0.16)	
LN	1.68***	5.37	1.38**	3.97	0.95**	2.57	1.26***	3.51
	(0.45)		(0.68)		(0.46)		(0.30)	
FJ	1.53***	4.63	− 1.30	0.27	1.06**	2.88	1.09***	2.96
	(0.49)		(1.35)		(0.52)		(0.33)	

Variables	Forestry cooperative	Odds ratios	Family forestry farm	Odds ratios	Specialized household	Odds ratios	Total transformed household	Odds ratios
JX	1.58***	4.86	0.54	1.72	1.00*	2.72	1.10***	3.02
	(0.50)		(1.00)		(0.53)		(0.34)	
HN	0.31	1.36	− 0.59	0.56	0.44	1.55	0.22	1.25
	(0.56)		(1.28)		(0.55)		(0.37)	
YN	1.41***	4.10	− 1.35	0.26	0.22	1.25	0.75**	2.11
	(0.47)		(1.22)		(0.52)		(0.33)	
SX	0.31	1.37	− 1.72	0.18	0.20	1.22	0.09	1.09
	(0.52)		(1.43)		(0.51)		(0.34)	
Constant	− 6.65***	0.001	− 6.04	0.0024	− 9.00***	0.0001	− 6.22***	0.002
	(1.84)		(3.90)		(2.35)		(1.40)	
Observations	3318		3164		3284		3487	
Log-likelihood	− 624.96		− 124.95		− 490.77		− 983.79	
Pseudo R^2	0.107		0.171		0.178		0.138	
McKelvey and Zavoina's R^2	0.262		0.383		0.260		0.347	
LR chi^2	149.24		51.56		212.93		314.01	
Prob > chi^2	0.0000		0.0000		0.0000		0.0000	

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; Values in the parentheses indicate standard error of the coefficients; Liaoning (LN), Fujian (Fj), Jiangxi (JX), Hunan (HN), Yunnan (YN) and Shaanxi (SX)

The OR of the variable AGE was 1.09, indicating that with an increase of individuals' age increases their likelihood to convert to a transformed household by 1.09 folds. Also, age has a positive impact on the FSH transformed (OR = 1.14), but it seemingly does not have significant impacts on switching to FC or FFFs, it had a significant impact on transformation to one of the three new types (OR = 1.09) (Table 3). However, the parameter estimated of AGE and AGE^2 were 0.086 and − 0.0008 ($p < 0.1$), respectively, suggesting age of household heads increase would increase the likelihood of transformation before reaching age of 56 years, and continues to get older would decrease the likelihood after 56 years old. Likely younger households tend to be risk takers and are more open to adopt new technologies and new organizational forms (Sheikh et al. 2003). With age increase, the household heads gain additional skills and experiences and are more inclined to transformation (Xie et al. 2014), while they would not try a newer technology or organizational structure. These two effects add up so that the likelihood of transformation begins to decrease after they get older than a certain age.

Education (EDU) had positive impacts on the total households transformed ($OR = 1.17$, $p < 0.05$). The respondents' education was also found to significantly affect their likelihood to transform to FFF in particular ($OR = 1.67$, $p < 0.05$). While household income (INC) was more likely to positively affect its likelihood to transform to FSH ($OR = 1.0002$, $p < 0.05$) (Table 3). Our finding align with the findings of Wei and Liu (2013) who also found traditional households with higher incomes and education had greater likelihoods of transformation. McCullough et al. (2008) also suggested that under the consumption upgrade and restructuring of supply chains, the traditional households with capital and knowledge intensive production technologies are more transformative. They also claimed that higher incomes, more skilled labor and more educated households are more likely to be the transformed households. Although education has a positive relationship to transformation according to the results and these researches, it probably became negative when the household heads received more education than a certain level that would allow them to choose an alternative career.

Population (POP) and total forestland area (FOR) are highly correlated with the traditional households converting to FC, FFF and FSH, and all household types combined (Table 3). Household family population relates to the labor a family can supply. Since FFF and FSH require more labor, family size being positively linked to the probability of a household to transform is logical. McCullough et al. (2008) found that the traditional households with larger land areas are more likely to transform to new forms of management. Total cropland area (CRP) OR is 1.01 ($p < 0.05$), which means increasing one ha of cropland would increase a household's likelihood to join FC increases by 1.01 times, suggesting the complementary nature of farming and forest management. Households with large cropland strengthen the rural economy and consolidate business in rural area, including higher intentions for forest management (Xie et al. 2014; Romm et al. 1987).

Unlike all other variables, proportion of a household's ecological forest area (ECO) had a strong negative impact on its likelihood to transform to FC ($OR = 0.51$, $p < 0.01$), FSH (0.46, $p < 0.01$) or all of the house types combined (0.47, $p < 0.01$) (Table 3). If households' forests were designated as ecological forests, the households can receive government subsidies and would be restrained from logging and other removal activities. Given this, the households might intensify farming or search for off-farm jobs (Liu et al. 2010).

The positive and significant signs of area of forestland transfer (TRS) on transformation of FCs, FSHs, and the three types combined (Table 3), indicate that the probability of transformation is higher when households have transferred more forestland from others. Ceng and Xia (2014) found empirical evidence that the transfer of land has a positive effect on transformation from peasant to farmer. The forestland transfer means the households rent forestland from other households and thus they had larger forestland area and would pay more attention to forestry. As a result, these households are more likely to join FCs, establish FFF or be FSH. The signing of a legal contract (CTS) is also critical for the transformation. In our results, the proportion of legally contracted forestland positively affects the transformation to new farmers ($p < 0.01$), which implies legal contracts are important for the successful transformation of households. The share of legally contracted forestland in total forestland area measures the security of forestland tenure. The households that signed forestland contracts with their villages could have more security of forestland tenure, and thus invest more in their forestland.

In addition to respondents' socioeconomic attributes, their geographical distributions also statistically significant affected their decision to switch. The largest possibility of such transformation was in Liaoning (LN) province, among the seven provinces in this study. Peasants in this province were more likely to switch to each of the three household types and all the types combined. Liaoning is the first province having FFF supporting policies, and it also has some early policies to support FC and FSH. Fujian (FJ) and Jiangxi (JX) provinces exhibited similar trends. Participants in these provinces transformed to all the other household types except FFF. Fujian is a well-developed province with a highly developed market economy and the first province that started collective forestland tenure reform. Both Fujian and Jiangxi have high forest coverage and policies supporting new types of farmers. Peasants in Yunnan (YN) province were not likely to switch to any new household type except FC. Yunnan, located in southwest China, is one of China's most undeveloped provinces and is situated in a mountainous area. A large percentage of the households in this province are ethnic minorities and they are traditionally used to helping each other, thus local households might prefer to work together to manage their forestland. Respondents in Hunan (HN) and Shanxi (SX) provinces were significantly less likely to switch to any of FC, FFF and FSH, because of lack of policy support and underdevelopment of commercial markets. Households in Gansu (GS) province were the least likely to be transformed among the seven provinces. This is perhaps caused by the fact that Gansu is most underdeveloped among these provinces with a poorly developed market economy, and part of the areas is deserts, which is not easy to develop as forestry.

2.5 Conclusions and recommendation

China is transforming from an agrarian society to an industrial society and more and more people are migrating from the countryside to urban areas. Currently, even the most isolated communities have noticeable exposure to commercial markets. This is particularly true for the communities dependent on forest products and non-wood forest products for sale. However, the transformation in terms of forest management practices and aligned market development still seems slow and exhibits significant regional or provincial variation. This study has identified factors that are significantly correlated with the transformation of traditional peasant households to three emerging household types namely FC, FFF, and FSH.

The contributing factors identified in this study include age and education level of household head, total income of household, areas of cropland and forestland, forestland transfer, legally contracted and proportion of ecological forest area, as well as province. Our findings reveal that the age pattern is a quadratic curve. An increase in household head age increases the likelihood of transforming to the new categories before the age of approximately 56 years. The age becomes negative after getting older than 56 years. Households with well-educated heads or high income showed increased probabilities to transform their structure. Family size has appeared to be one of the strongest driving forces for traditional households to transform. Our results further reveal that a household's share of cropland area has important positive impacts on potential household transformation, which may be attributed to its financial strength. We also conclude that household transformation potentials have significant variations among provinces. Households in Liaoning, Fujian, and Jiangxi have higher probabilities of transformation owing to the size and strength of their well-developed economies. Households in Liaoning are more likely to build up FFF since this is China's first province to introduce policies supporting FFF. In contrast, traditional households in Hunan and Shanxi provinces had the least likely to opt for household transformation.

Our analysis shows that the household's share of forestland area and the area of forestland transfer have positive impacts on potential household transformation. These results suggest that policy makers should improve the incentives to engage the households that have moved to cities or are not well versed in forestry practices, to transfer their forestland use rights. This would help consolidate forestland to fewer households. Our findings also reveal that the legal contract of land use rights is an important factor for household transformation, indicating importance of security of forestland tenure. As a result, the forestry authority should continue to issue the government sanctioned forest tenure certificates and promote the collective adoption of the legal contracts. This might expedite the household transformation process with the prospect of better economies of scale across the country.

Rural households are gradually transforming around the world, so too is forest management (Keskitalo 2017). China's households may have made substantial transformations from rural households in a relatively short period, but still are not the point of post-industrial societies in which forest management is becoming to aim more one amenities and less timber production (Zhang et al. 2005, 2009). We only identified FC, FFF and FSH as transformed households, and may have omitted many other households who have carried out transformation but not identified and categorized as the three types in this study. A more complete picture of household transformation will help the policymakers plan for more comprehensive and future oriented policies.

2.6 References

Bijman J, Hu D (2011) The rise of new farmer cooperatives in China: Evidence from Hubei Province. Journal of Rural Cooperation 39:99-113

Ceng F, Xia Y (2014) Agricultural Land Transfer and Cultivation of New Farmers - An Empirical Analysis Based on Polynomial Distribution Lags Model. Agricultural Technology Economy 6:14-21

Chen Y, Liu Y, Xu K (2010) Characteristics and mechanism of agricultural transformation in typical rural areas of eastern China: A case study of Yucheng City, Shandong Province. Chinese Geographical Science 20:545-553 doi:https://doi.org/10.1007/s11769-010-0430-4

Cheng T, Selden M (1994) The origins and social consequences of China's hukou system. The China Quarterly 139:644-668 doi:https://doi.org/10.1017/S0305741000043083

Cohen ML (1993) Cultural and political inventions in modern China: the case of the Chinese "peasant". Daedalus:151-170

Conteh AM, Moiwo JP, Yan X (2016) Using a logistic regression model to analyze alley farming adoption factors in Sierra Leone. Small-scale Forestry 15:109-125 doi:https://doi.org/10.1007/s11842-015-9311-0

Cramer JS, Ridder G (1988) The logit model in economics. Statistica Neerlandica 42:297-314 doi:https://doi.org/10.1111/j.1467-9574.1988.tb01241

Deng H, Huang J, Xu Z, Rozelle S (2010) Policy support and emerging farmer professional cooperatives in rural China. China Economic Review 21:495-507 doi:https://doi.org/10.1016/j.chieco.2010.04.009

Fan CC (1999) Migration in a socialist transitional economy: heterogeneity, socioeconomic and spatial characteristics of migrants in China and Guangdong Province. International Migration Review 33:954-987 doi:https://doi.org/10.1177/019791839903300406

Garnevska E, Liu G, Shadbolt NM (2011) Factors for successful development of farmer cooperatives in Northwest China. International Food Agribusiness Management Review 14:69

Gustafsson B, Shi L (2002) Income inequality within and across counties in rural China 1988 and 1995. Journal of Development Economics 69:179-204 doi:https://doi.org/10.1016/S0304-3878(02)00058-5

He M, Wen C, Li Y (2017) Analysis on the Operational Efficiency and Its Influencing Factors of the Family Forest Farm in China. Forestry Economics 3:27-34

Huang Z, Liang Q (2018) Agricultural organizations and the role of farmer cooperatives in China since 1978: past and future. China Agricultural Economic Review 10:48-64 doi:https://doi.org/10.1108/CAER-10-2017-0189

Islam KN, Rahman MM, Jashimuddin M, Hossain MM, Islam K, Al Faroque M (2019) Analyzing multi-temporal satellite imagery and stakeholders' perceptions to have an insight into how forest co-management is changing the protected area landscapes in Bangladesh. Forest Policy Economics 101:70-80 doi:https://doi.org/10.1016/j.forpol.2019.01.011

Li K (2016) Report on the Work of the Government. Delivered at the Fourth Session of the Twelfth National People's Congress of the People's Republic of China on March 5, 2016. Beijing

Liu C, Lu J, Yin R (2010) An estimation of the effects of China's priority forestry programs on farmers' income. Environmental Management 45:526-540 doi:https://doi.org/10.1007/s00267-010-9433-2

Liu Y et al. (2016) Progress of research on urban-rural transformation and rural development in China in the past decade and future prospects. Journal of Geographical Sciences 26:1117-1132 doi:https://doi.org/10.1007/s11442-016-1318-8

Liu Y, Lu S, Chen Y (2013) Spatio-temporal change of urban–rural equalized development patterns in China and its driving factors. Journal of Rural Studies 32:320-330 doi:https://doi.org/10.1016/j.jrurstud.2013.08.004

Long H, Liu Y, Li X, Chen Y (2010) Building new countryside in China: A geographical perspective. Land Use Policy 27:457-470 doi:https://doi.org/10.1016/j.landusepol.2009.06.006

Long H, Zou J, Pykett J, Li Y (2011) Analysis of rural transformation development in China since the turn of the new millennium. Applied Geography 31:1094-1105 doi:https://doi.org/10.1016/j.apgeog.2011.02.006

McCullough EB, Pingali PL, Stamoulis KG (2008) The transformation of agri-food systems: globalization, supply chains and smallholder farmers. Food & Agriculture Org., London, VA

National Bureau of Statistics of China (2019) Preliminary Accounting Results of GDP for the Fourth Quarter and the Whole Year of 2018 vol 2019. Beijing

Nkamleu GB, Manyong VM (2005) Factors affecting the adoption of agroforestry practices by farmers in Cameroon. Small-scale Forest Economics, Management Policy 4:135-148 doi:https://doi.org/10.1007/s11842-005-0009-6

Omiti J, Otieno D, Nyanamba T, McCullough E (2009) Factors influencing the intensity of market participation by smallholder farmers: A case study of rural and peri-urban areas of Kenya. African Journal of Agricultural Resource Economics 3:57-82

Schneider M (2015) What, then, is a Chinese peasant? Nongmin discourses and agroindustrialization in contemporary China. Agriculture Human Values 32:331-346 doi:https://doi.org/10.1007/s10460-014-9559-6

Sheikh A, Rehman T, Yates C (2003) Logit models for identifying the factors that influence the uptake of new 'no-tillage'technologies by farmers in the rice–wheat and the cotton–wheat farming systems of Pakistan's Punjab. Agricultural Systems 75:79-95 doi:https://doi.org/10.1016/S0308-521X(02)00014-8

Shen M, Shen J (2018) Evaluating the cooperative and family farm programs in China: A rural governance perspective. Land Use Policy 79:240-250 doi:https://doi.org/10.1016/j.landusepol.2018.08.006

Sitko NJ, Jayne TS (2014) Structural transformation or elite land capture? The growth of "emergent" farmers in Zambia. Food Policy 48:194-202 doi:https://doi.org/10.1016/j.foodpol.2014.05.006

State Forestry Administration (2017) Guiding Opinions of the State Forestry Administration on Accelerating the Cultivation of New Forestry Management Subjects. State Forestry Administration, Beijing

Wang G, Innes JL, Lei J, Dai S, Wu SW (2007) China's forestry reforms. Science 318:1556-1557 doi:https://doi.org/10.1126/science.1147247

Wei X, Liu W (2013) New Professional Farmers: Connotation, Characteristics and Cultivation Mechanism. Agricultural Economy 7:73-75

Wen W (2018) Focusing on Poverty Alleviation to Secure a Decisive Victory in Building a Moderately Prosperous Society in all Respects: Based on Targeted Poverty Alleviation in Hunan Province. Paper presented at the 2018 3rd International Conference on Politics, Economics and Law (ICPEL 2018), October 09-11, 2018

Xie Y, Gong P, Han X, Wen Y (2014) The effect of collective forestland tenure reform in China: Does land parcelization reduce forest management intensity? Journal of Forest Economics 20:126-140 doi:https://doi.org/10.1016/j.jfe.2014.03.001

Yang L, Wen Y, Aguilar F (2013) Nonindustrial family forest landowners' stated willingness-to-participate in forest cooperatives in Southern China. International Journal of Forestry Research 2013:1-12 doi:http://dx.doi.org/10.1155/2013/983168

Zhang J, Zhi L (2010) Current situation and prospect of forestry cooperative economic organization. World Forestry Research 23:65-68

Zhang QF, Donaldson JA (2010) From peasants to farmers: Peasant differentiation, labor regimes, and land-rights institutions in China's agrarian transition. Politics Society 38:458-489 doi:https://doi.org/10.1177/0032329210381236

Zhou Y, Guo Y, Liu Y, Wu W, Li Y (2018) Targeted poverty alleviation and land policy innovation: Some practice and policy implications from China. Land Use Policy 74:53-65 doi:https://doi.org/10.1016/j.landusepol.2017.04.037

Zhu Q, Mi S, Yang L, Huang L, Lu Y (2013) Occupation farmers and analysis of affecting factors of the industry development: Taking Zhejing Province as a case study. Bulletin Of Science And Technology 11:218-222

Chapter 3. Logging industry in the United States: employment and profitability

3.1 Introduction

With technological advancements, policy changes, parcelization of forestland, business cycle, and the change of relative costs of factors, the labor force of the logging industry in the U.S. has undergone dramatic changes in the past few decades (Conrad IV et al., 2018a). Employment in the logging industry is concentrated in the West and the South, which experienced large declines in employment. The employment decline was likely related to the change in the age distribution of the loggers (Baker and Greene, 2008; Rum-mer, 1994). Survey results from various studies have indicated an increase in the mean and median age of logging business owners and employees (Allen et al., 2008; Bolding et al., 2010; Broussard Allred, 2009; Egan, 2011; Greene et al., 2013; Grushecky et al., 2006; LeBel and Stuart, 1998; Leon and Benjamin, 2012). These authors discuss the importance of the age imbalance in logging, and many U.S. industries face an aging workforce as the Baby-boomer generation nears retirement age (Butler, 2008; Dorr and Feuerhelm, 2021; Grice et al., 2011; Schwatka et al., 2012).

Since logging wages have increased at about the rate of inflation and the interest rate remains low, logging firms continue to mechanize. Capital investments were mostly used for harvesting and transportation equipment, in addition to buying stumpage (Shivan et al., 2020). Early in the 2000s, the initial investment of logging firms was between \$0.4 million to \$1.5 million (Rickenbach and Steele, 2005). By the mid-2010s, the average equipment investment of Georgia logging firms was \$1.97 million, and \$2.23 million for South Carolina firms (Conrad IV et al., 2018b). However, demand for logging workers continues despite mechanization (Abt, 2013). The logging firm owners have indicated that recruitment of qualified employees was a challenging aspect of managing their business (Bolding et al., 2010; Egan and Taggart, 2009).

On one hand, changes in employment and mechanization affect the cost of logging and profit, while on the other, firm profitably affects employment and mechanization. In some firm-level surveys, the profits of logging firms were decreasing (Baker et al., 2014; Blinn et al., 2015; Broussard Allred, 2009), while other surveys in other regions showed the profits still increasing, or at least stable (Egan, 2011; Milauskas and Wang, 2006; Ricken-bach and Steele, 2005). However, the acquisition of the profit data is solely based on the questionnaires issued to the selected firms. Due to time and budget constraints, only a few surveys are conducted, and the accuracy of self-reported firm financial data is unknown. The limits of survey data and the lack of industry-level data obscure our understanding of logging industry profitability. Therefore, a simple and cost-effective economic model to estimate the profit for the industry would be valuable to investigate the factors affecting profit and how changes in profit affect the operational behavior of logging firms and would provide information for the policymakers.

Previous research data from logging firms were collected selectively and focused on aspects of the logging industry like demography, employment, harvesting systems, pro-duction level, operational costs, and/or profitability. The analyses were beneficial for the research in the logging industry. However, they failed to identify the overall situation and trends. Conflicting conclusions drawn from these analyses might be traced to sample size effects or regional differences. This paper adds both firm-level data from previous research with federal and state-level data, providing additional evidence for industry-level trends and relationships.

This study focuses on the two major core factors related to business in the logging industry: labor and firms. The objective of this paper is to provide a quantitative analysis of U.S. logging industry data and present trends in employment and profitability over the last 20 to 30 years. The results of the study will help us better understand the factors influencing the U.S. logging industry.

3.2 Data and methodology

3.2.1 Study Area

The study area included regional and state data from New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont), Mid-Atlantic (Delaware, Maryland, New Jersey, New York, and Pennsylvania), the South (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia), Mid-West (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin), the Southwest (Arizona, New Mexico, Oklahoma, and Texas), and the West (Alaska, Colorado, California, Hawaii, Idaho, Montana, Nevada, Oregon, Utah, Washing-ton, and Wyoming).

The main interest of this paper is to study the industry-level situation and trends in the logging industry. Therefore, we selected indicators for the logging industry using official estimates of employment, wage, number of establishments, production level, and production price of logging industry (Code 1133 North American Industry Classification System, NAICS) at the state-level from U.S. Bureau of Labor Statistics, U.S. Census Bureau and U.S. Department of Agriculture, Forest Service. The logging industry (NAICS 1133) comprises firms primarily engaged in cutting timber; producing rough, round, hewn, or riven primary wood; cutting and transporting timber; and producing wood chips in the field (Office of Management and Budget, 2017).

3.2.2 Data Sources

Specifically, we extracted the data for the logging industry from Occupational Employment Statistics (OES), Quarterly Workforce Indicators (QWI), Quarterly Census of Employment and

Wages (QCEW), Timber Product Output (TPO) Reports, and Producer Price Index (PPI). Due to the availability of data, the analysis presented in this paper focuses on the period from 1997-2019.

OES program collects data on wage and salary workers in nonfarm establishments for about 800 occupations, including national and state annual employment, hourly wage, and annual wage data. The OES survey is a semi-annual mail survey of nonfarm establishments (Bureau of Labor Statistics, 2020c). The data is classified by the Standard Occupation Code (SOC) and NAICS Code. The data extracted from OES included state-level employment, mean hourly wage, and annual mean wage data of Fallers, Log-ging Equipment Operators, Log Graders and Scalers, and Logging Workers, All Others from 1997 to 2019 (Bureau of Labor Statistics, U.S. Department of Labor, 2020).

The QWI has a set of 32 economic indicators, including employment, job creation/ destruction, wages, hires, and other measures of employment flows. The QWI data is based on the administrative records on employment collected by the states, social security data, federal tax records, and other census and survey data. QWI data set includes quarterly national and state employment and wage data for most industries. Application Programming Interfaces (APIs) can be applied to extract QWI data via Python, R or Excel. Package “CenPy” was used to run APIs from QWI on Python and Library “tidyqwi” on R. The data extracted from QWI included information like state-level employment data from 1997 to 2017, state-level employment by age classes in 1997, 2007, and 2017, state-level monthly earnings of newly stable employees in the logging industry and all industries in the U.S. from 1997 to 2019, and state-level total quarterly payroll from 2001 to 2019 (U.S. Census Bureau, 2019).

The QCEW publishes a quarterly count of employment and wages reported by employers covering more than 95 percent of the U.S. jobs available at the county, state, and national levels by detailed industry. The QCEW data is collected from the unemployment insurance (UI) accounting system, Quarterly Contributions Report (QCR), Report of Feder-al Employment and Wages, Annual Refiling Survey (ARS), and Multiple Worksite Report (MWR). However, QCEW excludes sole proprietors, the unincorporated self-employed, unpaid family members, specific farm and domestic workers from having to report employment data, which likely reduces the representativeness of the data set (Bureau of Labor Statistics, 2020a). Data from QCEW can be extracted through the One-Screen Data Search (Bureau of Labor Statistics, 2020f). The State-level Number of Establishments in the logging industry (NAICS 1133) from 2001 to 2015 were also extracted from QCEW (Bureau of Labor Statistics, 2020e).

TPO is conducted by Forest Inventory and Analysis (FIA) to estimate timber products at the state-level. Primary wood-using mills were sampled, by state, to estimate round-wood production (U.S. Department of Agriculture, Forest Service, 2020). However, TPO only covers some states and some years. The state-level total volume of roundwood products from 1997 to 2018 was extracted from the TPO data set (U.S. Department of Agriculture, Forest Service, 2019). PPI data by NAICS Industry can be extracted at the one screen tool (Bureau of Labor Statistics, 2020d). However, the PPI data set does not have any state-level data. The national PPI for NAICS 1133 from 1997 to 2018 was extracted from PPI by Industry.

3.2.3 Methodology

We used OES data to calculate the compound annual rate of growth with regards to the total hours worked in the logging industry at the state-level and used TPO data to calculate the compound annual rate of growth with regards to the volume of output in the logging industry at the state-level. Considering that the volume of output cannot reflect the improvement of product quality, we introduced the price factor, PPI in logging industry with an index base set at 1981 = 100 (Harrison and Sharpe, 2009). The volume output multiplied by PPI expressed the value of output, which was used to estimate the labor productivity (Moulton, 2018). The rate of growth in labor productivity based on volume of output is equal to the compound annual rate of growth in the volume of logging production minus the compound annual rate of growth in hours worked. And the rate of growth in labor productivity based on value of output is equal to the compound annual rate of growth in the economic value of logging production minus the compound annual rate of growth in hours worked.

Considering the difficulty of collecting profit data and the lack of profit statistics at the federal or state-level, we applied a new method to simply estimate the profit of logging firms which was the Economic Input-Output Life Cycle Assessment (EIO-LCA) Model. EIO-LCA models are based on the environmental Input-Output (I/O) modeling approach (Leontief, 1986) and are developed by Carnegie Mellon University (Norman et al., 2007). EIO-LCA models were applied to estimate the materials and energy resources required for the supply chain, environmental emissions, and economic values. In EIO-LCA models, the output of the first tier of suppliers, X_1, is given by Eq.(5):

$$X_1 = (I + A)y \tag{5}$$

where I is the identity matrix, y is final demand, and A is the matrix of intermediate input coefficients. Eq. (5) means the sector and all other sectors need to produce $I \times y$ and $A \times y$ units of production, respectively, to meet the demand (Carnegie Mellon University, 2021b).

The output of the first tier of suppliers also creates a demand for output from their direct suppliers, the second tier of suppliers. The final demand of the second tier of suppliers is $A \times A \times y$.

Consequently, the final demand of the third tier of suppliers is $A \times A \times A \times y$, and so on. Thus, the total output can be written as:

$$X = (I+A+AA+AAA+\cdots)y \tag{6}$$

where X is a vector of total output.

The intermediate input coefficients a_{ij} can be calculated by Eq. (7):

$$a_{ij} = X_{ij}/X_j{}^{-1} \tag{7}$$

where X_{ij} is the intermediate transaction from industry i to industry j, X_j is the total input of industry j.

The value-added for any industry is the difference between its total input and the total cost of intermediate transactions:

$$V_j = X_j - \sum_{i=1}^{n} X_{ij} \tag{8}$$

where V_j is the value-added for industry j (Khongprom et al., 2020), and it is the sum of compensation of employees, taxes, and profits in industry j (Carnegie Mellon University, 2021a).

3.3 Results and discussions

3.3.1 Employment

Declining employment is a problem endemic to the logging industry and experienced in all industrialized countries (Goldstein et al., 2005), such as Canada (Dodson et al., 2015) and Europe (Spinelli et al., 2013) with a similar situation in the U.S. The regional employment declined from 1997 to 2017 (Table 4), and employment in the whole country fell at an annual rate of 2.0% (U.S.

Census Bureau, 2019). The Southwest region experienced the fastest decline of all six regions with a decrease of 3.9%. The Mid-Atlantic had the lowest employment and the second-fastest regional decline. The South, which had the highest employment, had a decline near the U.S. total at -1.8%.

Table 4. Annual growth of employment in logging industry from 1997 to 2017.(U.S. Census Bureau, 2019)

Region	Total Employment, 1997*	Total Employment, 2017	Total Growth of Employment (%)	Annual Growth of Employment (%)
New England	3283	2750	-16.2	-0.9
Mid-Atlantic	2284	1327	-41.9	-2.7
The South	36761	25575	-30.4	-1.8
The Southwest	3154	1430	-54.7	-3.9
Mid-West	5206	4176	-19.8	-1.1
The West	20853	12553	-39.8	-2.5

* For states that had no employment data for 1997, the closest available year was used.

The lack of newly hired workers is one of the main reasons presented for declining employment (Greene et al., 2013). One proposed cause of the decline in the number of younger loggers has been the relatively low appeal of logging employment and business creation (Broussard Allred, 2009). The surveys attribute difficulties in recruiting to uncertainty and instability concerning business outlook and seasonal (Blinn et al., 2015; Egan and Taggart, 2009; Shivan et al., 2020). Logging jobs are physically demanding, mostly outdoors, and require work in poor weather and isolated areas (Bureau of Labor Statistics, 2020b). Higher compensation may be required to attract new employees. However, over the years, the wages of newly hired employees in the logging industry have been almost the same as the average for all industries in the U.S., and in some years even lower (Figure 3), while other industries can offer higher pay, better benefits, and more steady work (Blinn et al., 2015).

The age class of the employees reflected the expansion of the industry into the mid-1990s followed by a steady decline. The mode age class shifted from 35-44 in 1997 to 45-54 (28%) by 2007. By 2017 age classes 45-54 (25%) and 55-64 (22%) had similar em-ployment. The 55-64 age class showed relatively large growth from 2007 to 2017, increasing from 10% to 22% (Figure 4). The population in the U.S. had a similar trend and those aged 55 and older accounted for 26.7% and 36.7% in 1999 and 2019, respectively (Bureau of Labor Statistics, U.S. Department of Labor, 2020). The shift in age has coincided with the decline in employment, so the shift may result from the aging of current workers and the limited demand for new labor.

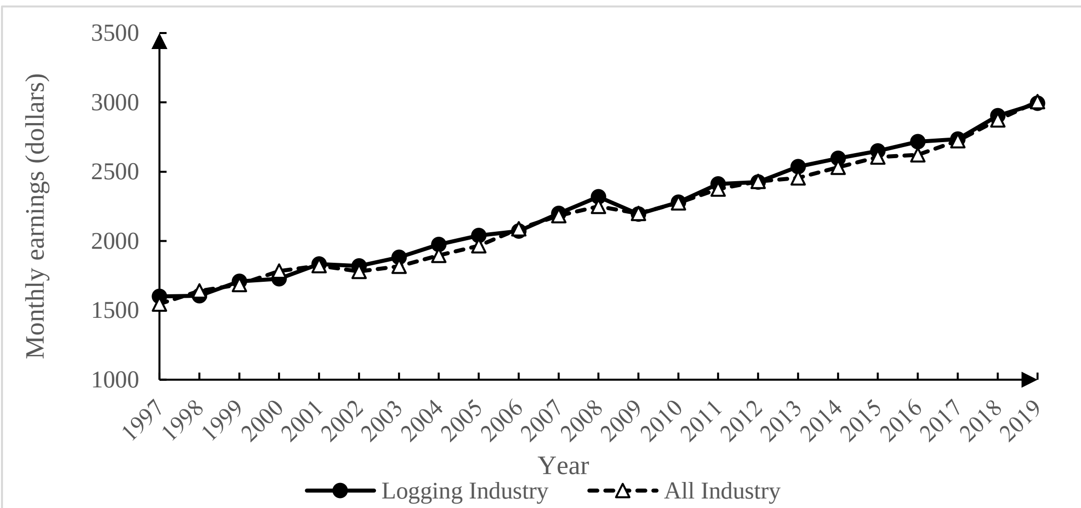

Figure 3. Average monthly earnings of newly stable employees in logging industry and all industries, US, from 1997 to 2019.(U.S. Census Bureau, 2019)

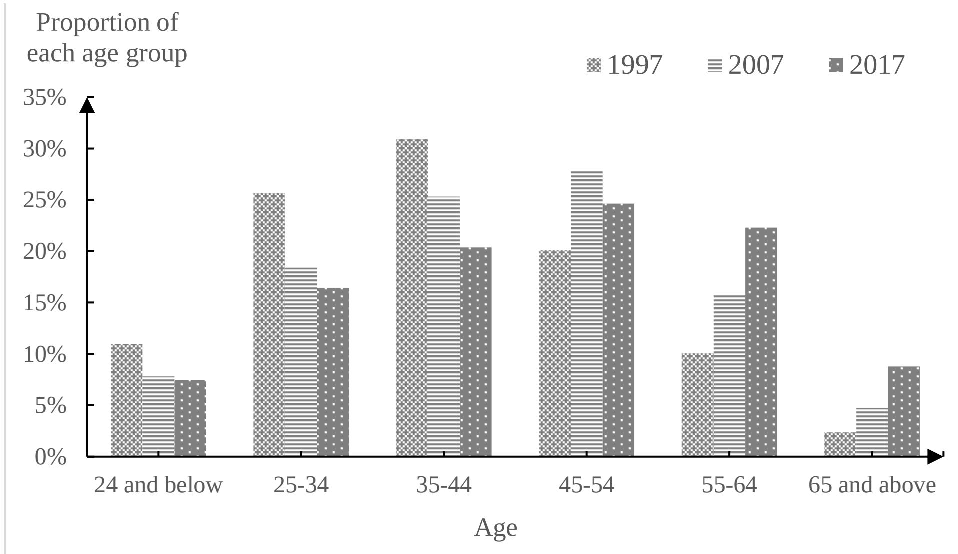

Figure 4. Age distribution of employees in logging industry, US, 1997, 2007 and 2017.(U.S. Census Bureau, 2019)

Technical advancement is another main reason for declining employment. We estimated the rate of growth in labor productivity of some states in the South, Southwest and the West (Table 5). The Volume of Output column and the Value of Output column are the rate of growth in the volume of logging production and the economic value of logging production, respectively. The Hours Worked column is the rate of growth in hours worked. Except for Tennessee and Oregon, both rates of growth in labor productivity were positive. The output and hours worked have declined in most states since 1997, but output declined at a slower pace than hours worked, which led to labor productivity growth.

Although employment in the logging industry declined, the number of logging equipment operators remained stable, from 22,690 in 2002 to 21,110 in 2019. The fallers declined, from 8410 in 2002 to 3180 in 2019. According to SOC codes, fallers (45-2021) use motor-manual methods (chainsaws) to fall trees (Bureau of Labor Statistics, 2019a). The change could indicate an increase in mechanization, increased productivity caused by equipment upgrades, or be related to changes in the terrain and forest types where timber harvesting occurred. The two states with declining productivity in Table 5Table 5 also might be characterized as somewhat dependent on motor-manual felling. However, similar states like Washington and Kentucky had growth. Although mechanization can replace labor, the logging industry might also face a structural shortage of labor. This industry did not lack workers, but the availability of skilled and technical workers may be limited. For example, it may take a new worker a year to master forwarder operation, including time on simulators (Wilson, 2017). With the progress of mechanization, logging firms increasingly needed equipment operators but faced difficulties in recruiting qualified employees (Bolding et al., 2010).

As the data in Table 5 indicate, the production level decreased in most states across this time. Most of the major states for logging, such as most southern states, had varying degrees of decline in output and corresponding declines in employment. A small number of states, Florida, South Carolina, and Oregon, had increased harvest level, but employment declined.

Table 5. The rate of growth in output, hours worked, and labor productivity, U.S. (%). (Bureau of Labor Statistics, U.S. Department of Labor, 2020; U.S. Department of Agriculture, Forest Service, 2019)

Region	State	Report year	Volume of output	Value of output	Hours worked	Labor productivity (volume)	Labor productivity (value)
The South	Alabama	1997-2015	-1.5	-1.3	-2.3	0.8	1.0
	Arkansas	1997-2015	-1.4	-1.2	-4.6	3.4	3.6
	Florida	1997-2015	0.3	0.5	-1.5	1.8	2.1
	Georgia	1997-2015	-0.1	0.2	-2.2	2.1	2.4
	Kentucky	1997-2015	-0.3	0.0	-2.4	2.1	2.4
	Louisiana	1997-2015	-0.3	0.0	-2.0	1.8	2.0
	Mississippi	1997-2015	-1.3	-1.1	-2.5	1.2	1.5
	N. Carolina	1997-2015	-1.7	-1.4	-1.7	0.1	0.3
	S. Carolina	1997-2015	0.8	1.0	-0.7	1.4	1.7
	Tennessee	1997-2015	-2.5	-2.2	-0.9	-1.6	-1.3
	Virginia	1997-2015	0.0	0.3	-0.3	0.3	0.6
The Southwest	Oklahoma	1997-2015	-1.4	-1.1	-5.1	4.0	4.2
	Texas	1997-2013	-2.5	-2.2	-3.2	0.8	1.1
The West	California	2000-2016	-1.8	-1.1	-4.7	3.0	3.7
	Colorado	2002-2016	4.5	5.8	3.7	0.7	2.0
	Idaho	2001-2015	0.9	2.2	-2.4	3.3	4.7
	Oregon	2003-2017	0.1	1.3	3.0	-2.8	-1.6
	Washington	2002-2016	-1.6	-0.3	-1.7	0.1	1.5
	Wyoming	2000-2018	3.7	4.6	-3.2	7.2	8.1
Average			-0.3	0.2	-1.8	1.6	2.1

3.3.2 Profitability

Figure 5 (a) and (b) are the results of the logging EIO-LCA models with the infla-tion-adjusted profits of logging firms and profits per logging firm (2019 Constant-dollar) in several states. From 1995 to 2009, the real profit and weighted average profit made by logging firms in these states continued to decline, reaching the lowest point in 2009 because of the economic recession.

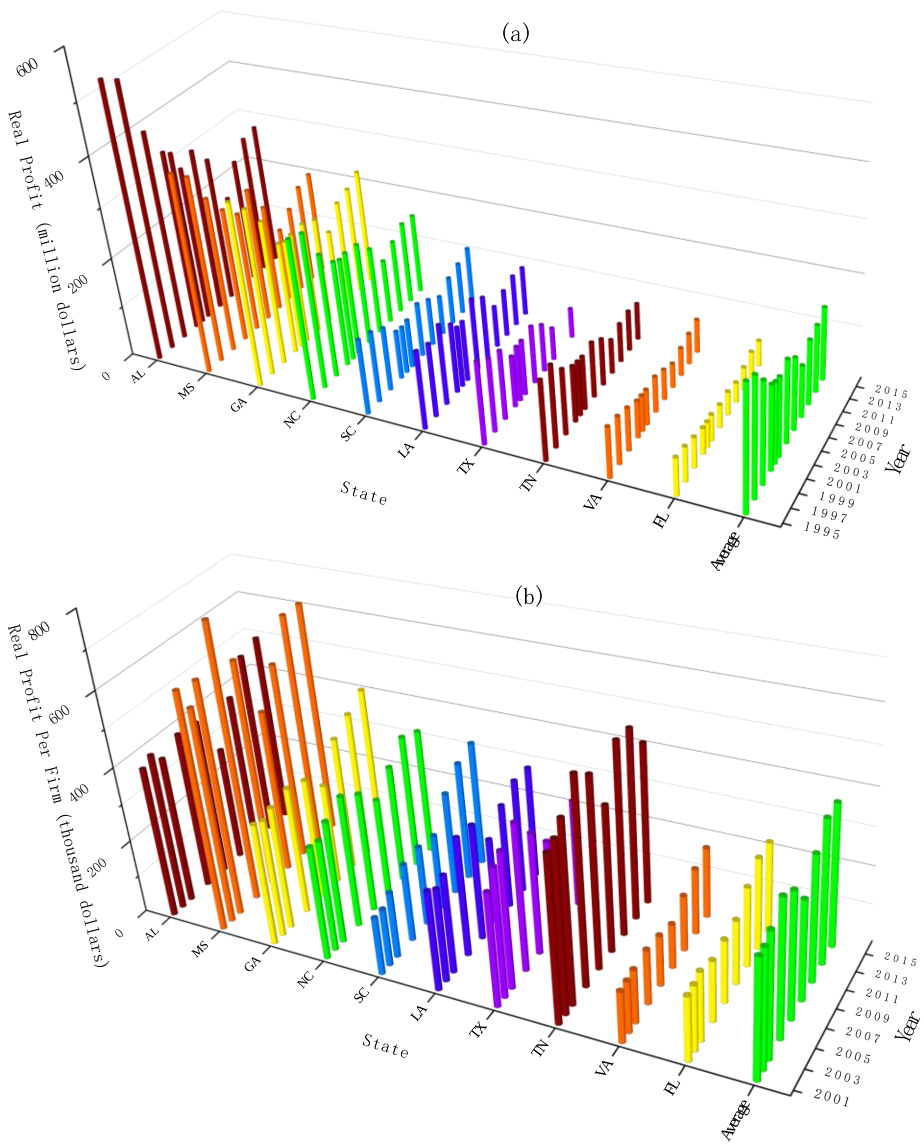

Figure 5. Inflation-adjusted (a) net profits of the logging industry at state-level and (b) net profits per logging firm in some states from 1995 to 2015, based on the EIO-LCA models (2019 Constant-dollar). (Carnegie Mellon University, 2020; U.S. Census Bureau, 2019)

Notes: AL, MS, GA, NC, SC, LA, TX, TN, VA, and FL signify Alabama, Mississippi, Georgia, North Carolina, South Carolina, Louisiana, Texas, Tennessee, Virginia, and Florida, respectively.

Demand for logging services is highly dependent on the economic cycle. The amount of timber harvested has largely been impacted by the demand for wood-frame housing (Drapala, 2009), the pulp and paper and furniture industries in the past decades (Abbas et al., 2014; Grushecky et al., 2006). When the economy is in boom periods, demand for construction, house renovation and furniture stimulate the demand for logging production, which in turn promotes the increase in prices and then the logging output. During an economic recession, logging suffers a sharp drop in profits due to the decline in demand and then prices.

Figure 6 demonstrates the relationship between the annual new private-owned housing units started in the U.S. and profit per logging firms. The new housing started can be applied as an indicator of demand for logging production. It can be observed that there is a positive correlation between profit and the new housing, as depicted in Figure 6. The new housing units started to reach their lowest point at 6648 in 2009 because of the economic crisis of 2008. Meanwhile, the profit also reached its lowest point. Following the economic crisis, the demand for housing began to increase, which was reflected in logging profit. Although the profit and new housing units had the same tendency after 2009, the profit decreased more than the new housing units during the recession and did not return to the level before the recession, making them not less correlated after 2009. This may result from the increasing operating costs, which squeezed the profit.

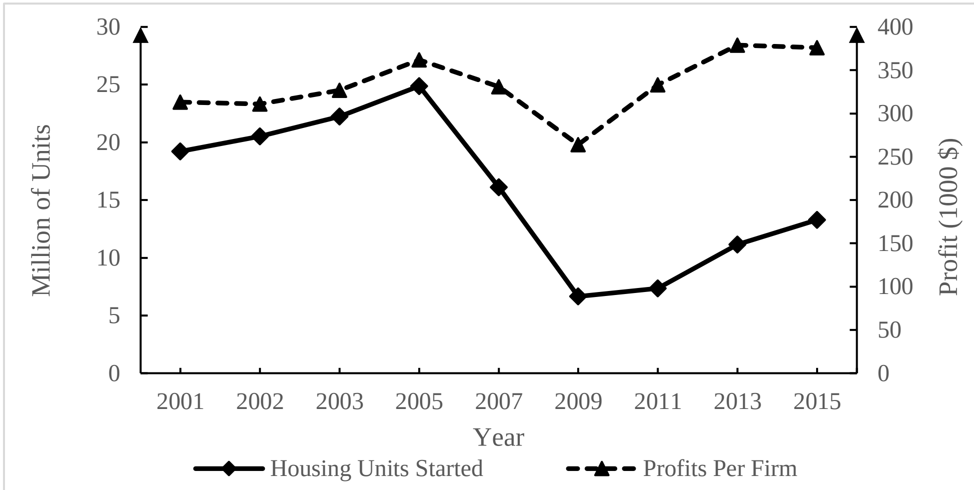

Figure 6. Inflation-adjusted profit per firm and annual new privately-owned housing units started (2019 Constant-dollar), 2001 to 2015. (Federal Reserve, 2021; U.S. Census Bureau, 2019)

The data show that the production level of all the states in 2009 had dropped sharply (Figure 7 (a)), and the Annual PPI of Logging also showed a decrease (Figure 8) due to the 2008 economic recession. Revenues from the logging industry in several states fell sharply in 2009 and then began to rebound (Figure 7(b)), indicating that the revenue of logging services had been seriously affected by the economic recession. From the comparison of these indicators, we can confer that the economic cycle had an impact on demand for logging production and then affected the price and production level of firms, finally influencing both the revenue and the profit.

Increasing operating costs (e.g., insurance premiums, wages, logging equipment and fuel costs) also reduced profit (Baker et al., 2014; Jacobson et al., 2009). Logging firms have operating costs as a combination of internal costs with labor, capital and operating cost components, which can be observed in Figure 9(a) and (b).

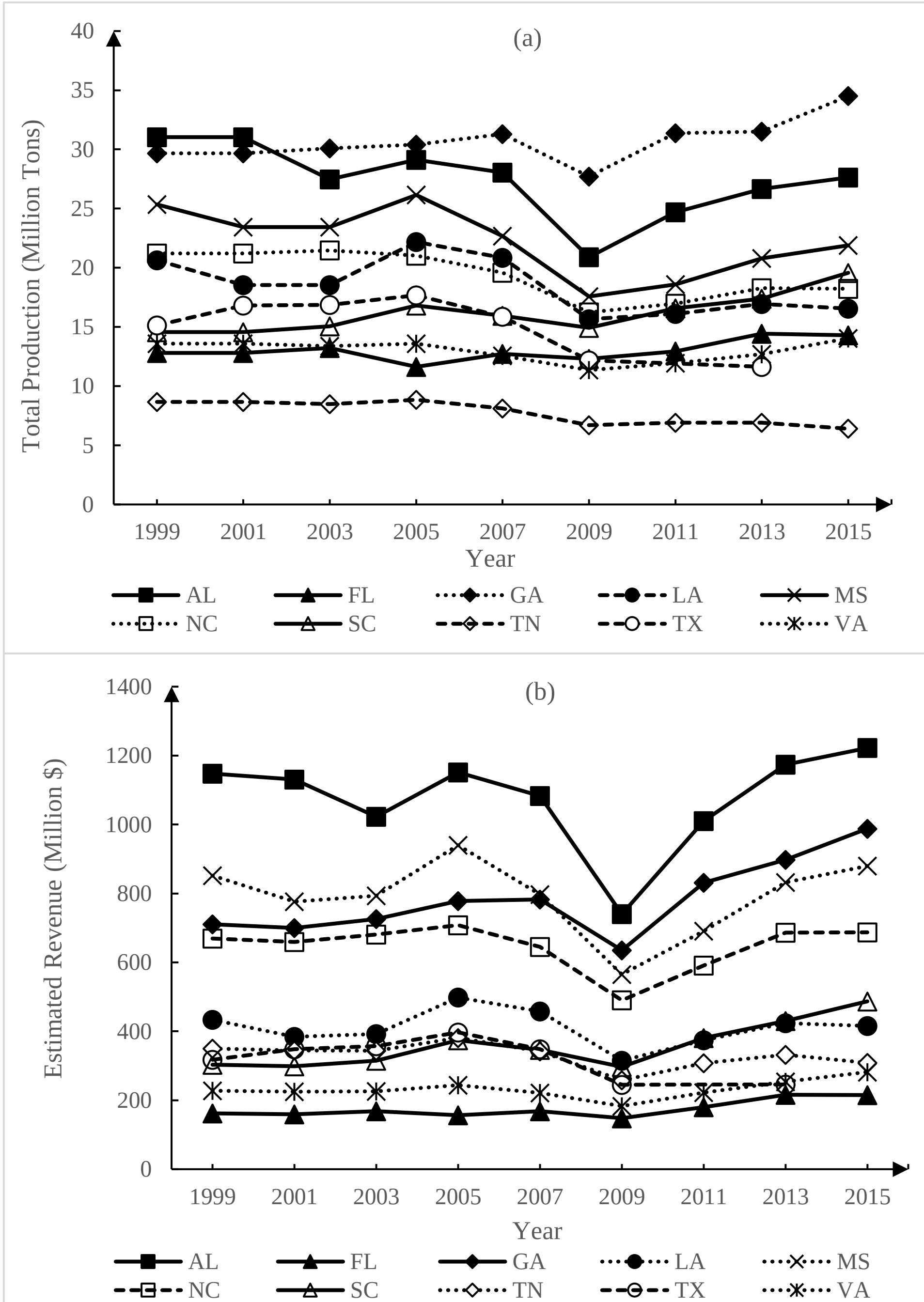

Figure 7. Logging industry (a) production level and (b) estimated revenue in some states, 1999 to 2015. (Bureau of Labor Statistics, U.S. Department of Labor, 2020; Timber Update, 2020; U.S. Department of Agriculture, Forest Service, 2019)

Notes: AL, FL, GA, LA, MS, NC, SC, TN, TX, and VA signify Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia, respectively.

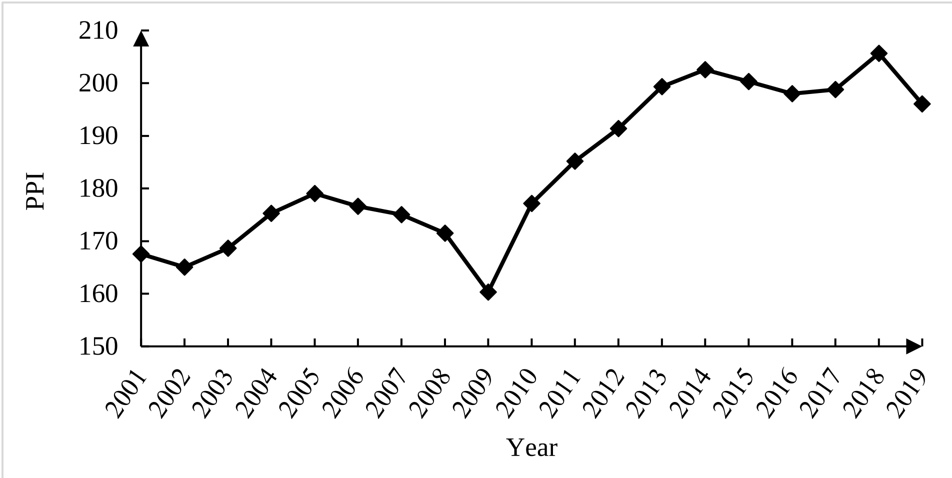

Figure 8. Annual Producer Price Index of logging (Index Dec 1981=100), US, from 2001 to 2019. (U.S. Census Bureau, 2019)

Notes: PPI signifies Producer Price Index.

We projected that wage and profits had a relatively negative correlation, as in Figure 9(a). For example, from 2002 to 2005, the real wages have fallen, while profits have risen. When wages reached a low point in 2005 and 2013, profits reached high points, which indicated logging firms lacked profit-sharing distributions with their employees.

Figure 9(a) also showed that wage accounted for a larger proportion of total costs. In the South, the wage accounted for more than 30% of the total costs (Xu et al., 2014). Nominal wages in the logging industry have increased by 3.73 times since 1977, with an average annual increase of 3.18% (Bureau of Labor Statistics, 2019b).

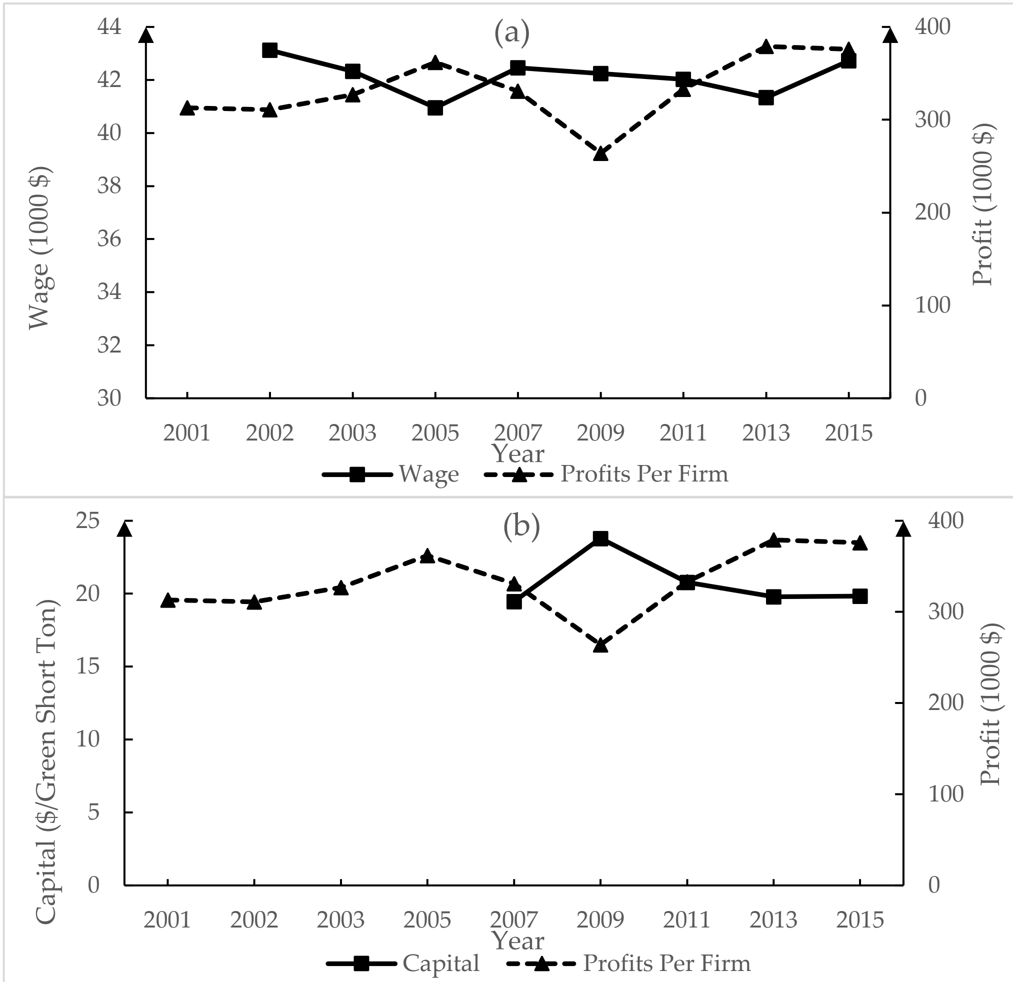

Figure 9. Inflation-adjusted profit per firm (a) annual wage, and (b) capital stock per ton production (2019 Constant-dollar), 2001 to 2015. (Barynin et al., 2013; U.S. Census Bureau, 2019)

Notes: Green Short Ton= 2,000 pounds of fresh-cut woody material at a "green" moisture content (Shelly, 2007).

We selected the states with the largest employment in the six regions and estimated the annual inflation-adjusted payroll (2019 Constant-dollar) paid by per logging firms in these six states from 2001 to 2019 (Figure 10). Real payroll per firm in these states had risen from 2001 to 2016, except for New York, which remained stable. Unlike the demand for logging production and prices, real payroll per firm in these states was not greatly affected by the economic recession. After 2016, real payroll per firm in Oregon and Texas fell sharply, but the average of these six states had increased steadily from 2001 to 2019. The increase in payroll was coincidental with the change in labor productivity. The average growth in labor productivity (volume) of sample states was 1.6%, while the labor productivity (volume) of Texas alone was 0.8%, and that of Oregon was -2.8% (Table 5).

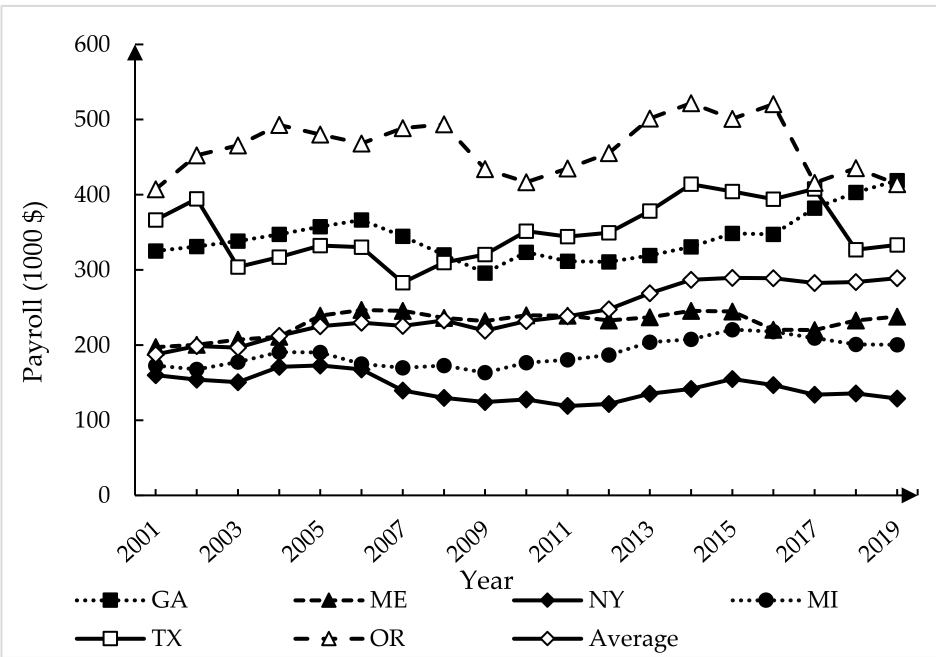

Figure 10. Inflation-adjusted payroll per logging firms (2019 Constant-dollar) in some states from 2001 to 2019.(Bureau of Labor Statistics, 2020d; U.S. Census Bureau, 2019)

Notes: GA, ME, NY, MI, TX, and OR signify Georgia, Maine, New York, Michigan, Texas, and Oregon, respectively.

Figure 9(b) depicts a similar relationship between capital and profit. Capital stock per ton production increased sharply in 2009. The decline in production level, which resulted from economic recession, led to an increase in unit capital stock. This also implies an increase in unit capital expenditure because some capital expenditures are fixed costs and will not decrease as the production level decreases. This also shows that the sharp decline in profits in 2009 was not only due to changes on the demand side but also due to constant labor costs and capital expenditures, which led to an increase in unit costs and squeezed profits. Transportation costs have also risen over the period, including driver wages, log truck insurance and transportation rates (Conrad IV et al., 2018b; Costello and Suarez, 2015).

After 2009, real profit rose (Figure 5(a)). Low profit, resulting from low demand and rising costs, was previously an important reason for firm owners to leave the log-ging industry (Egan and Taggart, 2004). The number of logging firms had been de-creasing, and the recession starting in 2008 accelerated this process. In the short run, logging firms would run without making a profit if they could cover variable costs to meet cash flow demands. In the long term, logging firms would use unprofitable jobs to bridge the gap between profitable jobs, especially when idling the business would result in greater losses (Regula et al., 2018).

With the reduction of profit, the impact of the 2008 economic recession, logging firms closed. However, the economic recovery increased demand while the number of logging firms still fell, leading to the rising market prices. The average firm profits have also been restored to the pre-recession level (Figure 5(b)). Thus, despite rising costs, increased profit began to attract new firms to enter the logging industry. The number of logging firms in some states stopped decreasing in 2015 and started to increase slightly (Bureau of Labor Statistics, 2020e).

3.4 Conclusion

This chapter analyzed the logging industry in the U.S. in recent decades from federal and state-level data, including OES, QWI, QCEW and TPO. These data sets contain many state-level indicators from multiple states over long periods, which can provide valuable information to investigate the facts and trends in the logging industry. The firm-level surveys also have value, such as those from Georgia and South Carolina (Conrad IV et al., 2018b), Michigan (Abbas et al., 2014), Maine (Taggart and Egan, 2011), and the South (Abt, 2013; Baker et al., 2012; Greene et al., 2013). These surveys can provide important information about the logging industry at the firm level, but those surveys were not conducted every year and suffered from small sample size and the associated bias. As a result, the multi-year, industry-level data can better demonstrate changes across the U.S. or regions and serve for comparisons between states and/or regions. The U.S. Bureau of Labor Statistics and the USDA Forest Service should pay attention to the construction of these databases, ensure the completeness and validity of the data, and strive to include data from more states. Thus, it would be easier for policymakers and industry practitioners to effectively monitor the entire logging industry.

This paper presents a feasibility method for estimating the profit for the logging industry. Due to time and budget constraints and a lack of profit statistics at the industry–level, it is valuable to estimate profit by the EIO-LCA model, which is simple and cost-effective. It can provide information to support the logging business operation and policymaking.

While it appears the declining employment resulted from the unavailability of newly hired workers (Greene et al., 2013), the more fundamental cause can be technological advancement represented by mechanization and the decline in the demand for logging production. We believe the decline in employment due to the inadequacy of newly hired workers is a short-term issue, whereas, mechanization tends to be a long-term one. This is consistent with other studies (Allen et al., 2008; Baker and Greene, 2008; Broussard Allred, 2009; Leon and Benjamin, 2012). Additionally, the inadequacy of newly hired workers results in the continuing aging workforce, which Canada is also facing (Allen et al., 2008). Future research needs to apply econometric modeling to analyze the contemporaneous causal relations among employment, wage, mechanization, production price and other factors in the logging industry, and to investigate the dynamic relationship among employment and other factors in short and long terms.

The labor shortage in the logging industry is a structural shortage, not a lack of labor, but a lack of skilled and technical workers. The structural shortage of labor will be a serious challenge for the logging industry. Logging firms are making up for the labor shortage through mechanization. However, mechanization also means increasing the qualifications of required workers. But based on the reality of a continuing aging workforce and the decline of younger workers entering, finding qualified workers could be difficult (Leon and Benjamin, 2012; Xu et al., 2014). A similar situation has also been reported in other countries, for example, New Zealand (Bayne and Parker, 2012; Kirk et al., 1997). Employers may need some assistance in on-the-job or off-the-job training to increase the number of qualified loggers. Skills certification and occupational licensing have been used to provide mobility for workers and may help employers easily identify qualified workers.

Since the early 1970s, employment in the logging industry in the U.S. has been steadily declining, while the production level has increased significantly, mainly due to technological advancements. As a result, the overall productivity of the logging industry has increased. There are significant differences in the logging industry productivity among different regions and states. Georgia and South Carolina have the highest logging labor productivity. Rising productivity levels in the southern U.S. have been described by Conrad IV et al. (2018b). High productivity in the

South is largely the result of the mechanized harvest system and its compatibility with planted pine forests. Mechanized logging operations accounted for more than 70% of the logging firms in Georgia since 1987, and it has accounted for more than 80% since 1992 (Baker and Greene, 2008). These studies are consistent with the results of our study. However, an opposite situation has been identified and the logging employment in Montana, Canada, decreased by 44% since 1993, while production level and revenue decreased more (64% and 71%, respectively), which indicated productivity of the logging industry might decrease (Morgan, T.A., Keegan, C.E., Hayes, S.W., Sorenson, C.B., 2013). Another study from Alberta, Canada, found that the rate of technical change and total factor productivity growth was negative because of stringent forest management regulations (Wang and An, 2019).

The logging industry is an important part of the timber supply chain and has an important impact on sustainable forest management. Therefore, logging firms with high production efficiency will determine the future of forestry in the U.S. (Conrad IV et al., 2018a; Duc et al., 2009). Future research is needed to measure the capital productivity and total factor productivity of logging firms at the industry level and study the influencing factors.

The prosperity of the logging industry is highly dependent on the economic conditions with an impact on both demand and price. Other studies based on firm-level data have reached similar conclusions: the economic recession that began in 2008 had severely affected the business environment in which they operated and that their profits were not sufficient to sustain their operations (Baker et al., 2012; Jacobson et al., 2009). Coupled with the continuous increase in operating costs, these two factors together have led to a wave of closures of logging firms. Similar situations were found in Canada (Cubbage et al., 1988; Dodson et al., 2015) and Europe (Spinelli et al., 2017). But it also accelerated the adjustment of the logging industry. With the decline in the number of logging firms, the adjustment of business strategies and the recovery of the economy, the profits of logging firms have risen again. With the outbreak of COVID-19, the U.S. economy fell into a recession again. Subsequently, the federal government launched multiple rounds of economic stimulus policies, which not only stimulated the economic recovery but also promoted the prosperity of the real estate market. Future research can focus on the impact of COVID-19 as a natural experiment to study the consequences of the economic cycle on the logging industry or the impact of the real estate market on the logging industry.

3.5 References

Abbas, D., Handler, R., Hartsough, B., Dykstra, D., Lautala, P., Hembroff, L., 2014. A survey analysis of forest harvesting and transportation operations in Michigan. Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering 35 (2), 179–192.

Abt, K.L., 2013. Employment and income trends and projections for forest-based sectors in the US. In: Wear, David N; Greis, John G., eds. 2013. The Southern Forest Futures Project: technical report. Gen. Tech. Rep. SRS-GTR-178. Asheville, NC: USDA-Forest Service, Southern Research Station. 293-308. 178, 293–308.

Allen, T.T., Han, H., Shook, S.R., 2008. A structural assessment of the contract logging sector in the Inland Northwest. Forest prod. j. 58 (5), 27.

Baker, S., Dodson, B., Greene, D., Hayes, S. (Eds.), 2014. Evaluation of an Online National Survey of Timber Harvesting Contractors.

Baker, S., Greene, D., Harris, T., 2012. Impact of timber sale characteristics on harvesting costs. Proc. South. For. Econ. Work (Charlotte, NC: Mississippi State University) pp, 94–105.

Baker, S.A., Greene, W.D., 2008. Changes in Georgia’s logging workforce, 1987-2007. Southern Journal of Applied Forestry 32 (2), 60–68.

Barynin, P., Taylor, D., Ghouri, W., Warth, C., 2013. Wood Supply Chain Analysis: Special Market Analysis Study. RISI.

Bayne, K.M., Parker, R.J., 2012. The introduction of robotics for New Zealand forestry operations: Forest sector employee perceptions and implications. Technology in Society 34 (2), 138–148.

Blinn, C.R., O'Hara, T.J., Chura, D.T., Russell, M.B., 2015. Minnesota's Logging Businesses: An Assessment of the Health and Viability of the Sector. Forest Science 61 (2), 381–387.

Bolding, M.C., Barrett, S.M., Munsell, J.F., Groover, M.C., 2010. Characteristics of Virginia's logging businesses in a changing timber market. Forest products journal 60 (1), 86–93.

Broussard Allred, S., 2009. Logging firm succession and retention. Forest products journal 59 (6), 31–36.

Bureau of Labor Statistics, 2019a. Occupational Employment and Wages: 45-4021 Fallers. https://www.bls.gov/oes/current/oes454021.htm#st. Accessed 2020.

Bureau of Labor Statistics, 2019b. National Industry-Specific Occupational Employment and Wage Estimates: NAICS 113300 - Logging.

Bureau of Labor Statistics, 2020a. Handbook of Methods: Quarterly Census of Employment and Wages. https://www.bls.gov/opub/hom/cew/home.htm. Accessed 20 January 2021.

Bureau of Labor Statistics, 2020b. Occupational Outlook Handbook: Logging Workers. https://www.bls.gov/ooh/farming-fishing-and-forestry/logging-workers.htm. Accessed 8 December 2020.

Bureau of Labor Statistics, 2020c. OES Data Overview. https://www.bls.gov/oes/oes_ques.htm#. Accessed 20 January 2021.

Bureau of Labor Statistics, 2020d. Producer Price Index by NAICS Industry. https://data.bls.gov/PDQWeb/pc. Accessed 20 December 2020.

Bureau of Labor Statistics, 2020e. Quarterly Census of Employment and Wages. https://www.bls.gov/cew/. Accessed 23 December 2020.

Bureau of Labor Statistics, 2020f. Quarterly Census of Employment and Wages:One-Screen Data Search. https://data.bls.gov/PDQWeb/en. Accessed 15 December 2020.

Bureau of Labor Statistics, U.S. Department of Labor, 2020. Occupational Employment Statistics. www.bls.gov/oes/. Accessed 26 December 2020.

Butler, B.J., 2008. Family forest owners of the United States, 2006. Gen. Tech. Rep. NRS-27. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station. 72 p. 27.

Carnegie Mellon University, 2020. Economic Input-Output Life Cycle Assessment Model. Green Design Institute, Carnegie Mellon University. http://www.eiolca.net. Accessed 1 November 2020.

Carnegie Mellon University, 2021a. Economic Results: Value Added Effects in Monetary Units. Carnegie Mellon University. http://www.eiolca.net/Method/interp-results/econ_direct_percent.html. Accessed 6 January 2021.

Carnegie Mellon University, 2021b. Theory and Method behind EIO-LCA. Carnegie Mellon University. http://www.eiolca.net/Method/eio-lca-method.html. Accessed 6 January 2021.

Conrad IV, J.L., Greene, W.D., Hiesl, P., 2018a. A Review of Changes in US Logging Businesses 1980s–Present. Journal of Forestry 116 (3), 291–303.

Conrad IV, J.L., Greene, W.D., Hiesl, P., 2018b. The evolution of logging businesses in Georgia 1987-2017 and South Carolina 2012-2017. Forest Science 64 (6), 671–681.

Costello, B., Suarez, R., 2015. Truck driver shortage analysis 2015. Arlington, VA: The American Trucking Associations.

Cubbage, F.W., Stokes, B.J., Granskog, J.E., 1988. Trends in southern forest harvesting equipment and logging costs. Forest Products Journal Vol. 32 (2): 6-10.

Dodson, E., Hayes, S., Meek, J., Keyes, C.R., 2015. Montana logging machine rates. International Journal of Forest Engineering 26 (2), 85–95.

Dorr, B.M., Feuerhelm, S.L., 2021. Addressing the silver tsunami in the accounting industry. Journal of Work-Applied Management.

Drapala, P., 2009. Decline in Housing Market Hits Forestry Industry Hard. Mississippi Agricultural News, December 17.

Duc, N.M., Shen, Y., Zhang, Y., Smidt, M., 2009. Logging productivity and production function in Alabama, 1995 to 2000. Forest products journal 59.

Egan, A., 2011. Characteristics of and challenges faced by logging business owners in Southern New England. Northern Journal of Applied Forestry 28 (4), 180–185.

Egan, A., Taggart, D., 2004. Who will log? Occupational choice and prestige in New England's north woods. Journal of Forestry 102 (1), 20–25.

Egan, A., Taggart, D., 2009. Public perceptions of the logging profession in Maine and implications for logger recruitment. Northern Journal of Applied Forestry 26 (3), 93–98.

Federal Reserve, Federal Reserve Economic Data: New Privately-Owned Housing Units Started, 2021. fred.stlouisfed.org/series/HOUST#. Accessed 29 September 2021.

Goldstein, J.P., Irland, L.C., Senick, J.A., Bassett, E.W., 2005. The Intergenerational Supply of Loggers Under Conditions of Declining Economic Well-Being. Industrial Relations: a Journal of Economy and Society 44 (2), 331–340.

Greene, W.D., Marchman, S.C., Baker, S.A., 2013. Changes in logging firm demographics and logging capacity in the US South, in: Proceedings of the 36th Annual Council on Forest Engineering Meeting, p. 7.

Grice, A., Peer, J.M., Morris, G.T. (Eds.), 2011. Today's aging workforce—Who will fill their shoes? IEEE, 483-491.

Grushecky, S.T., McGill, D.W., Anderson, R.B., 2006. Inventory of wood residues in southern West Virginia. Northern Journal of Applied Forestry 23 (1), 47–52.

Harrison, P., Sharpe, A., 2009. A Detailed Analysis of the Productivity Performance of the Canadian Forest Products Sector Since 2000. Centre for the Study of Living Standards 2009-09.

Jacobson, M., Finley, J., Schmid, C., 2009. Factors and Trends in Pennsylvania's Logging Industry.

Khongprom, P., Champanoi, S., Suwanmanee, U., 2020. An Input-Output Approach for Environmental Life Cycle: Assessment of Cement Production. CHEMICAL ENGINEERING TRANSACTIONS 81, 1345-1350.

Kirk, P.M., Byers, J.S., Parker, R.J., Sullman, M.J., 1997. Mechanisation developments within the New Zealand forest industry: the human factors. Journal of Forest Engineering 8 (1), 75–80.

LeBel, L.G., Stuart, W.B., 1998. Technical Efficiency Evaluation of Logging Contractors Using a Nonparametric Model 9 (2), 15–24.

Leon, B.H., Benjamin, J.G., 2012. A survey of business attributes, harvest capacity and equipment infrastructure of logging businesses in the northern forest. School of Forest Resources, University of Maine, Orono.

Leontief, W., 1986. Input-output economics. Oxford University Press.

Milauskas, S.J., Wang, J., 2006. West Virginia logger characteristics. Forest products journal 56 (2), 19.

Morgan, T.A., Keegan, C.E., Hayes, S.W., Sorenson, C.B., 2013. Montana's forest products industry: 2013 Outlook. Bureau of Business and Economic Research. https://www.bber.umt.edu/pubs/forest/Outlook/forestproducts2015.pdf. Accessed 14 November 2021.

Moulton, B.R., 2018. The Measurement of Output, Prices, and Productivity: What's Changed Since the Boskin Commission? https://www.brookings.edu/research/the-measurement-of-output-prices-and-productivity. Accessed 23 December 2020.

Norman, J., Charpentier, A.D., MacLean, H.L., 2007. Economic input-output life-cycle assessment of trade between Canada and the United States. Environmental science & technology 41 (5), 1523–1532.

Office of Management and Budget, 2017. The North American Industry Classification System (NAICS) revision for 2017. Office of Management and Budget. https://www.census.gov/naics/reference_files_tools/2017_NAICS_Manual.pdf. Accessed 12 November 2021.

Regula, J., Germain, R., Bick, S., Zhang, L., 2018. Assessing the Economic Viability of Loggers Operating Tree-Length Harvest Systems in the Northeast. Journal of Forestry 116 (4), 347–356.

Rickenbach, M., Steele, T.W., 2005. Comparing mechanized and non-mechanized logging firms in Wisconsin: Implications for a dynamic ownership and policy environment. Forest products journal 55 (11), 21.

Rummer, R.B., 1994. Labor for Forestry Operations–Issues for the 1990s. Transactions of the ASAE 37 (2), 639–645.

Schwatka, N.V., Butler, L.M., Rosecrance, J.R., 2012. An aging workforce and injury in the construction industry. Epidemiologic reviews 34 (1), 156–167.

Shelly, J.R., 2007. Woody biomass definitions and conversion factors. University of California, Berkeley. https://ucanr.edu/sites/WoodyBiomass/newsletters/IG003_-_Woody_Biomass_Definitions_and_Conversions_Factors31510.pdf. Accessed 14 November 2021.

Shivan, G.C., Potter-Witter, K., Blinn, C.R., Rickenbach, M., 2020. The Logging Sector in the Lake States of Michigan, Minnesota, and Wisconsin: Status, Issues, and Opportunities. Journal of Forestry.

Spinelli, R., Magagnotti, N., Facchinetti, D., 2013. Logging companies in the European mountains: an example from the Italian Alps. International Journal of Forest Engineering 24 (2), 109–120.

Spinelli, R., Magagnotti, N., Jessup, E., Soucy, M., 2017. Perspectives and challenges of logging enterprises in the Italian Alps. Forest Policy and Economics 80, 44–51.

Taggart, D., Egan, A., 2011. Logging across borders and cultures: An example in northern Maine. Forest products journal 61 (7), 561–569.

Timber Update, 2020. All Timber Prices. https://timberupdate.com/timber-prices/. Accessed 15 November 2020.

U.S. Census Bureau, 2019. Quarterly Workforce Indicators. https://qwiexplorer.ces.census.gov/static/explore.html#x=0&g=0.

U.S. Department of Agriculture, Forest Service, 2019. Timber Product Output (TPO) Reports, Knoxville, TN. http://srsfia2.fs.fed.us/php/tpo_2009/tpo_rpa_int1.php. Accessed 19 September 2020.

U.S. Department of Agriculture, Forest Service, 2020. Timber Products Output Studies. https://www.fia.fs.fed.us/program-features/tpo/. Accessed 20 January 2021.

Wang, S., An, H., 2019. Technical change and productivity growth in the Alberta logging industry. Forest Policy and Economics 102, 130–137.

Wilson, J., 2017. The Job No One Wants: Why Won't young People Work in Logging? https://www.theguardian.com/us-news/2017/aug/23/logging-industry-work-employment-oregon. Accessed 12 May 2021.

Xu, Y., Smidt, M., Zhang, Y., 2014. Logging worker wage, performance, and experience. Forest products journal 64 (5-6), 210–216.

Chapter 4. What drives the change in employment in the U.S. logging industry? -A Directed Acyclic Graph approach

4.1 Introduction

Forestry plays an essential role in the U.S. rural economy. For example, the forestry industry in Virginia generated more than $23 billion in industry output and employed nearly 145,000 people in 2006 (Bolding et al., 2010; Rephann, 2008). The logging industry is an integral part of the forest industry, providing raw materials (for example, sawn wood and wood chips) to the wood processing industry. It was estimated that the logging industry contributed $36.2 billion to the economy and created 488 thousand jobs in the U.S. (Jolley et al., 2020). In the Northern Forest region, the logging industry employed approximately 11 thousand employees and provided valuable jobs for rural communities, where no other jobs were available (Leon and Benjamin, 2012). For example, the logging industry in Maine generates good job opportunities in rural areas where employment opportunities are limited (Taggart and Egan, 2011). It was estimated that logging and trucking industries in Maine contributed $619 million in output, 9 thousand jobs, and $342 million in labor income in 2017 (Bailey et al., 2020).

However, employment in the logging industry has declined across the U.S. in recent years. The West and South, where the most concentrated areas of logging, have declined the most. Employment in the logging industry has fallen by an average of 2% per year since 1997 (Figure 11). Oregon provides the most prominent employment opportunity in the West and even the whole of the U.S. The logging employment in Oregon dropped from 7,727 in 1997 to 7,408 in 2002 and further declined from 6,631 in 2007 to 5,262 in 2017. In 2002, Alabama had the most significant logging employment in the South at 5,133, while Georgia came in second at 4,968. Georgia surpassed Alabama as the state with the most logging jobs in 2017, but employment in both states fell to 3,994 and 3,772, respectively.

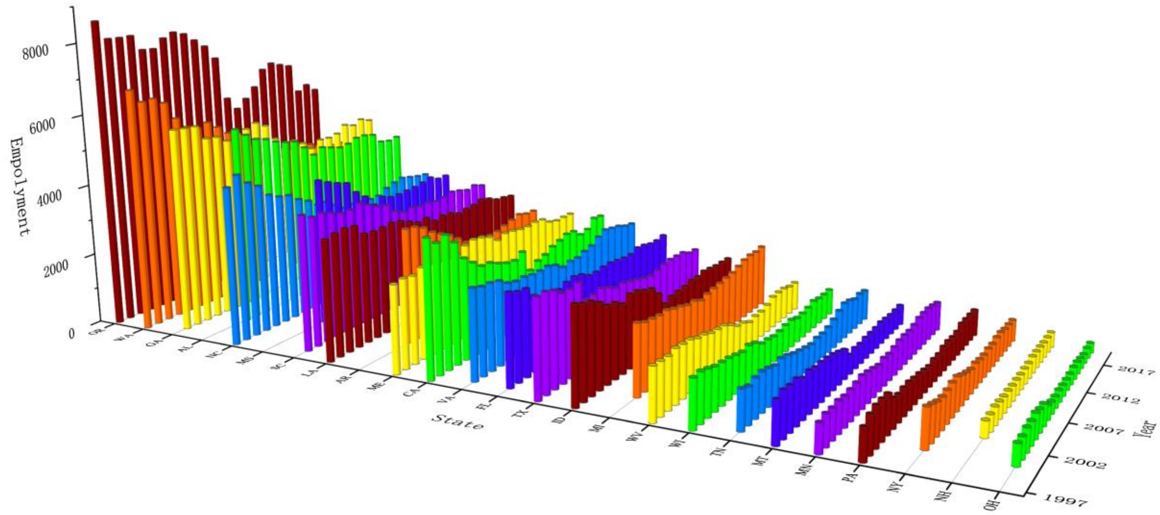

Figure 11. Employment in Logging Industry across the U.S., 1997-2019.

This study applied the Directed Acyclic Graph (DAG) to investigate the drivers of employment in the U.S. logging industry from 2007 to 2017. It attempted to bridge this gap by exploring contemporaneous causal relationships among employment, wages, mechanization, logging product level and product prices. The labor market in the logging industry depends on both demand and supply. From the demand side, the total removal of forest resources is associated with the degree of mechanization and the business cycle, particularly the housing market and building permits, pulp and paper prices. From the supply side, the most critical variables are relative wages with competitive sectors. Then, this study employed the Forecast Error Variance Decomposition (VD), and tried to understand better the effects of these factors on employment in the short and long-run.

Previous research mainly focused on one single factor in the logging industry, such as demography, employment, harvesting systems, and/ or production level, and only pairwise directional connectedness between two variables was identified. For example, Abbas et al. (2014) analyzed the employment and mechanization in Michigan and Wisconsin logging industry by statistical analysis and found that the decreased production level resulted from the shutdown of the pulp and paper industries leading to the logging equipment operators leaving the industry. Jacobson et al. (2009) collected data via focus groups and a survey questionnaire in Pennsylvania and found that the production partly affected employment through statistical analysis. Lee and Eckert (2002) had a similar conclusion based on statistical analysis to study the logging industry in the states of Washington and Oregon in the U.S. and Japan. Shivan et al. (2020) investigated the status of the logging industry in Michigan, Minnesota, and Wisconsin via descriptive and inferential statistical techniques and found that higher wages and benefits can increase employment. Duc et al. (2009) used the data from a mail survey in Alabama to regress production functions and found the elasticity between labor and machine was unitary. Allred et al. (2011) surveyed the midwest logging firm and applied Principal Component Analysis (PCA) and Analysis of variance (ANOVA) to investigate the influence of mechanization on the cost and profitability of the logging industry. Baker and Greene (2008) conducted a survey in Georgia and found that mechanization increases production per person-hour and the efficiency of the human capital using statistical analysis, which indicates that an increase in the capital can reduce costs, which in turn lowers product prices.

However, employment, wage, mechanization, logging product prices, and production levels are interrelated and, therefore, a contemporaneous causal relationship needs to be established between these multiple variables. Some comprehensive studies systematically describe various aspects of the logging industry (Boltz et al., 2003; Conrad and Greene, 2017; Conrad IV et al., 2018; Moskalik et al., 2017), but most are literature reviews that do not focus on the reasons behind these facts.

4.2 Data and empirical approach

4.2.1 Data sources

We extracted the data for the logging industry (Code 1133 North American Industrial Classification System, NAICS) to construct the time- series from Quarterly Workforce Indicators (QWI), Timber Product Output (TPO) Reports, Wood Supply Chain Analysis, and Timber Update (Table 6). Due to the limited availability of data, we are only able to work on a smaller dataset from 11 states (Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia) for six years (2007, 2009, 2011, 2013, 2015 and 2017). There were a few missing data in 2008, 2010, 2012, 2014, and 2016. As a result, we applied the average interpolation method for the missing data in between two data points. Table 6 shows the list of variables used in this study. To reduce the skewness of the raw data, the Emp, W, and Q were transformed into their respective logarithmic forms.

Table 6. List of variables.

Variables	Description	Data Sources	Unit
Emp	Employment in the logging industry	QWI	#
W	Average monthly earnings of logging workers	QWI	$
K	Capital stock per ton production	Wood Supply Chain Analysis	$/Green Short Ton
P	Logging product price (stumpage price)	Timber Update	$
Q	Logging production level	TPO	1000 Cubic Feet

Notes: QWI (Quarterly Workforce Indicators): https://qwiexplorer.ces.census.gov/static/explore.html#x=0&g=0; Wood Supply Chain Analysis: https://www.forestresources.org/resources/research/item/1532-wood-supply-chain-analysis-special-market-analysis-study; Timber Update: https://timberupdate.com/timber-prices/; TPO (Timber Product Output Reports): http://srsfia2.fs.fed.us/php/tpo_2009/tpo_rpa_int1.php.

4.2.2 Empirical approach

To investigate the causal relationships among employment, wage, mechanization, logging product prices, and logging production levels in the logging industry, the following steps were identified. We first applied a multivariate time-series model, the Vector Autoregression (VAR) model. Secondly, we employed a graphical modeling analysis, the Directed Acyclic Graph (DAG) approach to capture the dynamic relationships between variables and determine their contemporaneous causal relationships. Finally, we conducted a structural analysis, Forecast Error Variance Decomposition (VD), to prove the importance of shocks in explaining changes in each variable in the VAR and show how the importance of shocks changes over time.

4.2.2.1 Vector Autoregression

The Vector Autoregression (VAR) model was first proposed by Sims (1980), which can provide a framework for understanding the causal relationships of multivariate time-series data. This paper employed the VAR model to capture the dynamic interdependence between employment, wage, mechanization, logging product price, and logging production level. The VAR model with n variables can be written as:

$$Y_t = \text{c} + \sum_{l=1}^{p} B_l Y_{t-l} + \varepsilon_t \tag{9}$$

where Y_t is an ($n \times 1$) vector of the intended variable; c is an ($n \times 1$) vector of constants; p represents the lag order of the model; B_l is an ($n \times n$) matrix of autoregressive coefficients to be estimated for lagged l period; ε_t is an ($n \times 1$) vector of uncorrected random errors. In this study, Y_t is a (5×1) vector including the variables *Emp*, *W*, *P*, *Q* and *K* in period *t*; *c* is a (5×1) vector; B_l is a (5×5) matrix of coefficients; ε_t is a (5×1) vector.

However, the VAR model cannot explain the contemporaneous relationships between the variables since the correlation is hidden in the error term of the VAR model (Haigh and Bessler, 2004; Ji et al., 2018). Additionally, it is hard to economically explain the coefficients of the VAR model (Sims, 1980). As a result, Directed Acyclic Graph and Forecast Error Variance Decomposition were widely used based on the VAR model.

4.2.2.2 Directed Acyclic Graphs

DAG approach was pioneered by Pearl (1995) and Spirtes et al. (2000) to explore the contemporaneous casual relationships and identify the causal patterns. DAG approach was first used to determine the causal flows based on the residual of the VAR model by Swanson and

Granger (1997). The residual correlation coefficient of the VAR model can be applied to build upon the contemporaneous causal flows by the DAG approach. In this paper, we used the DAG approach to explore the contemporaneous relationships of economic factors in the logging labor industry and identify the causal patterns among them.

The basic idea behind the DAG is to depict the causal link (cause → effect) between two variables to represent the contemporaneous causal flow. If these two variables, for example, *X* and *Y*, are linked by an arrow, it signifies they are adjacent. The arrow represents the causal relationship between *X* and *Y*. If the arrow is from *X* to *Y* ($X \rightarrow Y$), *X* is referred to as the parent of *Y* and *Y* is *X*'s child (Chen et al., 2021), which suggests *X* results in *Y*. Therefore, if there is no edge between *X* and *Y* (*X* *Y*), it means there is no causal relationship between *X* and *Y*. If there is a non-directed edge (*X*-*Y*), it means the direction of the causal relationship between *X* and *Y* is unknown. Also, the bidirected edges ($X \leftrightarrow Y$) indicate a bidirectional causality relationship between *X* and *Y* (Pan et al., 2019). However, the bidirectional edges do not exist in DAG (Chen et al., 2021).

In this study, we applied Peter-Clark (PC) algorithm to identify the edges and direction of the causal relationship among the variables. The PC algorithm is divided into two steps:

Firstly, a complete undirected graph is built up. In this graph, all the variables have an edge linking to every other variable. Then, the unconditional correlation test between any pairs of variables is carried out. If the correlation is not statistically different from zero, the edge between these two variables would be eliminated.

Secondly, conditional correlation is tested for the remaining edges. The remaining edges are checked for the first-order conditional correlation, given any third variable. If the correlation is not statistically different from zero, the edge would be deleted. Then the edges which survive the first-order conditional correlation test are checked for second-order partial correlation and so on. The algorithm continues to check the conditional correlation test for *N* variables until $(N-2)^{th}$ order (Spirtes et al., 2000).

To test whether the unconditional correlations and conditional correlations are statistically different from zero, Fisher's z-statistic was applied in this study:

$$z[\rho(\mathrm{i},\mathrm{j}|k)\ n] = \left[\frac{1}{2}\sqrt{(n-|k|-3}\right] \times ln\left\{\frac{|1+\rho(\mathrm{i},\mathrm{j}|k)|}{1-\rho(\mathrm{i},\mathrm{j}|k)}\right\} \tag{10}$$

where *n* is the number of observations which are applied to calculate the correlations, $\rho(i,j|k)$ is the population conditional correlation coefficient between series *i* and *j*, which is conditional on series *k* (Bessler and Yang, 2003).

4.2.2.3 Forecast Error Variance Decomposition

To analyze the dynamic structure of the VAR model, Forecast Error Variance Decomposition (VD) is applied to simulate how much of the forecast error variance of variables can be explained by exogenous shocks to the other variables and endogenous shocks by themselves (Bernanke and Gertler, 1995).

A VAR model can be expressed as a Vector Moving Average (VMA) (Enders, 2008). Therefore, Eq.(11) can be iterated backward infinite times to obtain a moving average order:

$$Y_t = \mu + \sum_{l=0}^{\infty} B_l \varepsilon_{t-l} \tag{11}$$

where $\mu = (\mathrm{I} + B_1 + B_2 + B_3 + \cdots)B_0$ is an unconditional mean of Y_t (Alsaedi and Tularam, 2020; Sheng and Tu, 2000). Thus, the m^{th} horizon forecast error is

$$Y_{t+m} - E_t(Y_{t+m}) = \sum_{l=0}^{m-1} B_l \varepsilon_{t+m-l} \tag{12}$$

And the m^{th} horizon forecast error variance of $y_{1,t}$ is

$$var(y_{1,t+m}) = \varepsilon_1{}^2 \sum_{l=1}^{m-1} \theta_{1,2}(l)^2 + \varepsilon_2{}^2 \sum_{l=1}^{m-1} \theta_{1,3}(l)^2 + \ldots + \varepsilon_n{}^2 \sum_{l=1}^{m-1} \theta_{1,n}(l)^2 \quad (13)$$

where θ is the impulse response function. Therefore, the ratio of relative variance contribution can be represented as:

$$R_{1,n}(m) = \frac{\varepsilon_n{}^2 \sum_{l=1}^{m-1} \theta_{1,n}(l)^2}{var(y_{1,t+m})} \quad (14)$$

where $R_{1,n}(m)$ is represents how much of the change in Variable 1 is caused by the shock of Variable *n* at the m^{th} horizon (Enders, 2008).

Because of the contemporaneous correlation among the errors of the VAR model, Cholesky decomposition is used to orthogonalize the covariance matrix of the residuals (Sims, 1980). However, the input order of the variable would be essential to the VD (Swanson and Granger, 1997) because different input order leads to varying results of VD. The previous research confirms the input order based on their subjective causal assumptions and analyses (Alsaedi and Tularam, 2020; Esmaeili and Rafei, 2021; McKenzie et al., 2009; Omisakin and Olusegun, 2008). The DAG approach identifies the causal patterns based on the data without any subjective assumptions and analyses, which can be used to confirm the input order of VD.

Due to the panel structure of the dataset, it is necessary to test the stationarity of each panel series to avoid spurious regression and ensure the validity of the results. Harris-Tzavalis, Breitung, and PP-Fisher tests were applied to examine the dataset that shows stationarity. If all variables were not stationary at their level, the Johansen-Fisher panel cointegration test was performed on the dataset. The Johansen-Fisher test is a non-parametric test that does not assume homogeneity in the coefficients. After testing the stationary and cointegration of our data, a VAR model of employment and the influencing factors were established. Subsequently, the DAG approach was applied to identify the causal relationship among the variables based on the results of the VAR model. Finally, the VD was utilized to investigate the dynamic relationship among the variables in the short- and long- run.

4.3 Results

The preceding section is divided into three sections: VAR, DAG and VD, which discuss the analysis results. The VAR section focuses on the tools used for performing the aforementioned analyses, including Panel Data Unit Root Test, Panel Data Cointegration Test, and VAR. The DAG section is subdivided to focus on each of the paths among employment, production level, wage, capital, and product price. The VD section focuses on the dynamic relationship among these variables in the short- and long- run.

Due to the panel structure of the dataset, it is necessary to test the stationarity of each panel series to avoid spurious regression and ensure the validity of the results. Harris-Tzavalis, Breitung, and PP-Fisher tests were applied to examine the stationarity of the dataset. If all variables were not stationary at their level, the Johansen-Fisher panel cointegration test was performed on the dataset. The Johansen-Fisher test is a non-parametric test that does not assume homogeneity in the coefficients. After testing the stationary and cointegration of our data, a VAR model of employment and the influencing factors were established. Subsequently, the DAG approach was applied to identify the causal relationship among the variables based on the results of the VAR model. Finally, the forecast error variance decomposition was utilized to investigate the dynamic relationship among the variables.

4.3.1 Panel data unit root test

For time-series data analysis, we need to confirm the stationarity of the data series and avoid the potential spuriousness (Kao, 1999; Olagunju et al., 2021). As a result, applying the panel unit root tests is necessary before analyzing the panel data (Table 7). In our study, the Harris-Tzavalis (Harris and Tzavalis, 1999), Breitung (Breitung, 2002), and PP-Fisher (Maddala and Wu, 1999) tests were explicitly employed.

The results show that none of the level tests on the W, P and Q rejected the null hypothesis of non-stationarity/existence of a unit root, but all the first difference tests on them reject the original hypothesis, indicating that the W, P and Q are first-different stationary and integrated of order one. Only the level test of the PP-Fisher on Emp and K rejected the null hypothesis and not the Harris-Tzavalis and Breitung test. However, all of the first difference tests on them rejected the null hypothesis significantly. Therefore, Emp and K can be regarded as first-different stationary and integrated of order one.

Table 7. Results of panel unit root test.

Variables	Method	H-T	Breitung	PP
Level	*Emp*	0.998	0.475	93.215***
	W	1.003	6.670	1.817
	P	1.012	-0.147	4.881
	Q	0.997	-0.582	29.581
	K	1.022	1.618	76.207***
First Difference	*Emp*	0.317***	-2.684***	60.332***
	W	0.7228***	-1.537*	73.476***
	P	0.178***	-4.870***	49.106***
	Q	0.4834***	-1.531*	31.323*
	K	0.272***	-2.934***	44.824***

Notes: ***, **, and * denote rejection of the null hypothesis at the 1%, 5%, and 10% levels, respectively.
H-T and P.P. signify Harris-Tzavalis and Phillips-Perron-Fisher, respectively.

4.3.2 Panel data cointegration test

As all variables are not stationary at their level but integrated under order one, it is necessary to employ the panel cointegration test before further econometric analysis. To make sure the results are robust, Pedroni Test (Pedroni, 2004) and Kao Test (McCoskey and Kao, 1998) were applied (Table 8). The results show that all the cointegration tests significantly reject the null hypothesis (no cointegration). Hence, there is strong evidence indicating that all the five variables have a long-run stable equilibrium relationship.

Table 8. Results of panel cointegration test.

	Pedroni Test		Kao Test
Modified PP	3.126***	ADF	-1.865**
PP	-5.858***		
ADF	-6.256***		

Notes: ***, **, and * denote rejection of the null hypothesis at the 1%, 5%, and 10% levels, respectively. PP and ADF signify Phillips-Perron and Augmented Dickey-Fuller, respectively.

4.3.3 Vector Autoregression

After testing the stationarity and cointegration of our data, we build up a VAR model of employment and the influencing factors. Table 9 shows the results of the five-variable VAR residual correlation matrix. The DAG approach was applied to analyze the VAR residual correlation matrix of these five variables to get contemporaneous causal patterns among employment, wage, capital, product price, and production level.

Table 9. Residual Correlation Coefficient Matrix of VAR

	Emp	W	P	Q	K
Emp	1				
W	0.7996	1			
P	0.6773	0.6941	1		
Q	0.7381	0.7630	0.4336	1	
K	-0.8987	-0.9223	-0.7380	-0.7065	1

4.3.4 Directed Acyclic Graph results

After testing the stationarity of each panel series and their cointegration, we carried out a VAR model. Following that, we applied the PC algorithm in Tetrad 6.8 to analyze the residual correlation coefficient matrix of the VAR model to obtain the DAG. The DAG, in turn, was used to disclose the contemporaneous causal structure. Figure 12 represents the Complete Undirected Graph on the variables. Figure 13 presents the DAG on these variables at the 20% significance level, which allows us to more accurately identify a contemporaneous causal relationship in a small sample (Spirtes et al., 2000).

In Figure 12, the Complete Undirected Graph shows the undirected path connecting each variable with every other variable. It reveals that these five economic variables were interrelated with each other.

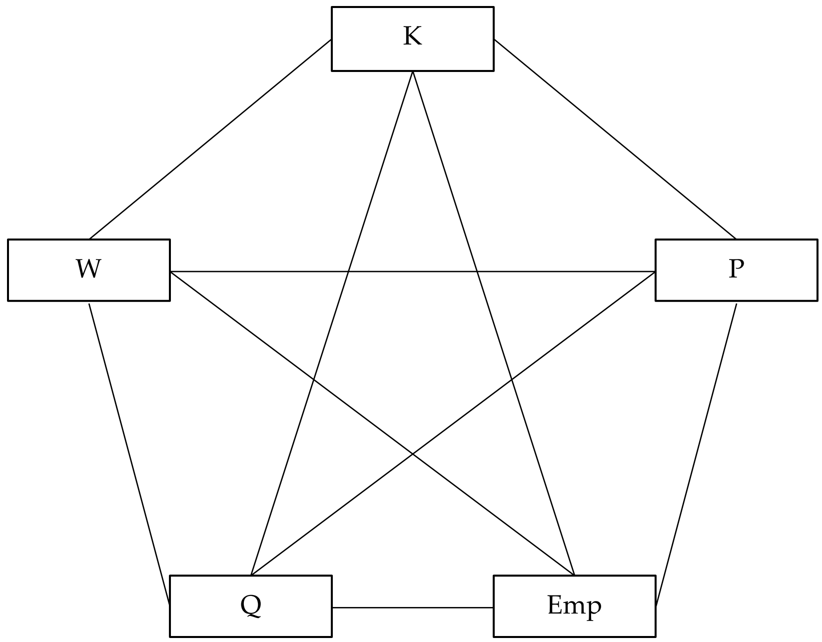

Figure 12. Complete Undirected Graph on employment in the logging industry.

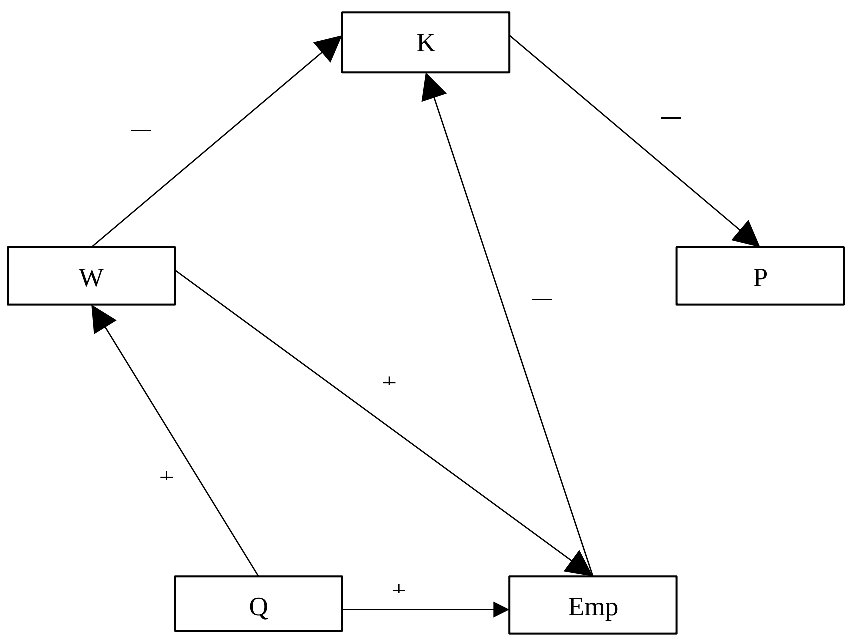

Figure 13. Directed Acyclic Graph on employment in logging industry (Significant Level at 20%).

Figure 13 shows the causal relationships paths among employment, production level, wage, capital, and product price. There are six paths in the graph: $Q \overset{+}{\rightarrow} Emp$, $Q \overset{+}{\rightarrow} W$, $W \overset{+}{\rightarrow} Emp$, $Q \overset{+}{\rightarrow} Emp$, $Emp \overset{-}{\rightarrow} K$, and $K \overset{-}{\rightarrow} P$.

The paths among employment and its influencing factors can be refined into the following two paths: $\mathrm{Q} \overset{+}{\rightarrow} \mathrm{Emp}$, and $\mathrm{Q} \overset{+}{\rightarrow} \mathrm{W} \overset{+}{\rightarrow} \mathrm{Emp}$.

- $\mathrm{Q} \overset{+}{\rightarrow} \mathrm{Emp}$: The results show that the level of production can directly promote employment. Higher production levels mean that logging firms need to hire more employees to increase production. In other words, a decline in employment may be caused by a decrease in product output.
- $\mathrm{Q} \overset{+}{\rightarrow} \mathrm{W} \overset{+}{\rightarrow} \mathrm{Emp}$: The results show that production level has a positive impact on wage, and wage directly impacts employment.

Figure 13 also shows the relationship between employment, wage, capital, and product prices, which discussed here further：

- $\mathrm{Q} \overset{+}{\rightarrow} \mathrm{Emp} \overset{-}{\rightarrow} \mathrm{K}$: The production level affects capital through both employment and wage and indirectly affects the capital. Additionally, capital is negatively affected by employment, showing a substitution effect between capital and labor.
- $\mathrm{Q} \overset{+}{\rightarrow} \mathrm{W} \overset{-}{\rightarrow} \mathrm{K}$: The production level has a direct positive effect on wage. And wage negatively affected capital.
- $\mathrm{K} \overset{-}{\rightarrow} \mathrm{P}$: The product price has a negative influence on capital.

4.3.5 Forecast Error Variance Decomposition results

After analyzing the contemporaneous causal relationship via DAG, we then investigated how much of the change in employment across time is caused by endogenous shocks by itself and how much by exogenous shocks, as well as other variables. The Forecast Error Variance Decomposition (VD) was utilized to investigate this dynamic relationship among variables (Table 10).

Table 10 reports the decomposition at horizons 1- to 5-, 10-, 15- and 20-year. It shows that employment in the logging industry is most prominently explained by the production level, followed by the wage. The promotion effect of production level on employment gradually decreases over time, and the percentage contribution decreases from 52.0% at horizon 1-year to 37.2% at horizon 20-year. The wage has the opposite effect. The wage plays an increasing role in the variation of employment, with 13.6% at horizon 1-year and 42.0% at horizon 20-year. Capital and product price only have limited influence on employment over time. These results are also consistent with that of the DAG. In addition, employment is also influenced by itself with a percentage contribution of 34% at horizon 1-year, and then the influence decreases step-wise. The influence is reduced to 4% at the horizon 20-year.

Forecast error variance for the wage is most prominently explained by the production level (55%), followed by the endogenous shock of the wage (45%). The shocks from employment, capital, and price are relatively small compared to that of production level and wage itself, which are consistent with those of the DAG. In addition, wage and employment play an important role in driving the capital. The wage keeps the stable influence on capital. On the other hand, employment plays a decreasing role with the extension of the forecast period.

Table 10. Results of Forecast Error Variance Decomposition.

VD of Variable	Horizon	Q	W	Emp	K	P
Q	1	1	0	0	0	0
	2	0.813	0.149	0.012	0.017	0.009
	3	0.633	0.267	0.023	0.035	0.043
	4	0.526	0.289	0.024	0.043	0.118
	5	0.592	0.216	0.019	0.028	0.145
	10	0.445	0.346	0.034	0.048	0.128
	15	0.415	0.369	0.036	0.052	0.128
	20	0.392	0.390	0.038	0.055	0.125
W	1	0.582	0.418	0	0	0
	2	0.486	0.468	0.013	0.021	0.013
	3	0.380	0.509	0.021	0.039	0.051
	4	0.319	0.484	0.021	0.045	0.131
	5	0.498	0.322	0.018	0.027	0.134
	10	0.468	0.332	0.032	0.044	0.124
	15	0.440	0.350	0.034	0.048	0.127
	20	0.413	0.371	0.036	0.052	0.128
Emp	1	0.543	0.128	0.329	0	0
	2	0.496	0.287	0.186	0.018	0.013
	3	0.397	0.372	0.146	0.035	0.051
	4	0.318	0.373	0.134	0.042	0.133
	5	0.469	0.262	0.085	0.027	0.158
	10	0.437	0.351	0.036	0.048	0.128
	15	0.412	0.372	0.036	0.052	0.128
	20	0.390	0.392	0.038	0.056	0.124
K	1	0.498	0.341	0.089	0.072	0
	2	0.420	0.404	0.084	0.077	0.015
	3	0.337	0.432	0.084	0.086	0.061
	4	0.391	0.351	0.068	0.072	0.118
	5	0.559	0.292	0.041	0.040	0.068
	10	0.557	0.297	0.029	0.033	0.084
	15	0.541	0.299	0.029	0.035	0.096
	20	0.520	0.306	0.030	0.037	0.107
P	1	0.188	0.315	0.071	0.006	0.419
	2	0.263	0.156	0.037	0.003	0.541
	3	0.583	0.199	0.022	0.016	0.180
	4	0.571	0.318	0.031	0.033	0.048
	5	0.492	0.397	0.038	0.047	0.026
	10	0.510	0.388	0.038	0.043	0.022
	15	0.522	0.378	0.037	0.041	0.022
	20	0.534	0.368	0.036	0.040	0.023

4.4 Discussion

4.4.1 Directed Acyclic Graph Analysis

4.4.1.1 Q $\overset{+}{\rightarrow}$ Emp

The shrinking logging industry has accelerated the decline in logging employment. The amount of timber harvested in the U.S. has largely been impacted by a sharp decline in demand for wood-frame housing (Drapala, 2009; Yin, 2001), and a systematic reduction in the pulp and paper and furniture industries in the past decades (Abbas et al., 2014; Grushecky et al., 2006). In most major logging states, such as most southern states, demand for timber has declined to varying degrees (Figure 1), thereby lowering employment.

The decline in demand for logs has caused them not to operate at full capacity. For example, it was estimated only 60% of their total operating capacity in Minnesota and 73% in Michigan (Abbas et al., 2014; Blinn et al., 2015). The decrease in production volume led to the increase in unit cost because for a given size and production level, and the reduced production volume resulted in logging firms not operating at the lowest point of the cost curve, leading to increased operating costs (LeBel and Stuart, 1998). In response to declining production levels, logging firms need to adjust their operating capacity to return to the lowest point of the cost curve. As a result, these firms have to reduce the number of employees accordingly.

4.4.1.2 Q $\xrightarrow{+}$ W $\xrightarrow{+}$ Emp

The effect of production level on wage is easy to understand because the wage is based on the production, at least partly (Xu et al., 2014). When the production level increases, it requires logging firms to recruit more employees, thereby driving up wages. In addition, existing employees may have to increase their working hours to match the higher production level, so wages also rise, and vice versa. Thirdly, the firms may offer higher wages to hire trained employees to increase productivity. In the context of expected experience, the wage increase for the performance step is reasonable (Xu et al., 2014).

Wage has an effect on employment for two reasons. First, wages can provide market signals for logging firms and workers. Logging firms can raise wages to recruit more employees. Second, higher wages also signal opportunities and career prospects, attracting additional employees.

4.4.1.3 Q $\xrightarrow{+}$ Emp $\xrightarrow{-}$ K

Employment has a direct negative effect on capital. On the one hand, reduced employment can lead to an increase in capital. The U.S. logging industry faces a severe aging workforce and declining recruitment of the upcoming generations(He et al., 2021). As more and more employees retire, to recruit new employees is challenging for low profit margins and full of uncertainty, instability, and seasonal operations (Egan and Taggart, 2009; Shivan et al., 2020). Logging is physically demanding and the fatal civilian occupation in the U.S. because logging workers must spend all their time outdoors, sometimes in poor weather and often in isolated areas (Bureau of Labor Statistics, 2020; Scott et al., 2020). Logging firms cannot offer attractive salaries to attract new employees (He et al., 2021). As a result, the young will not enter the industry to replace those who leave (Baker and Greene, 2008). Thus, the logging firms had to replace jobs through the substitution of capital for labor.

On the other hand, increased employment can result in a decrease in capital. The logging firms, especially small-scale firms, tend to employ more staff to exploit the internal economies instead of investing in mechanized systems. Small-scale logging firms are widespread in the U.S. and have been remarkably tenacious. The small-scale logging firms are developed with a long history in the U.S. (Conrad IV et al., 2018). One of the main reasons for the presence of many small logging firms is parcelization (Milauskas and Wang, 2006; Yin et al., 1998). The number of forestland owners has increased rapidly in the last decade, resulting in a decline in the average size of forest ownerships, most of which are nonindustrial private forest (NIPF) owners (Rickenbach and Steele, 2006). However, the large logging firms with mechanized harvest systems may not match this small-scale forestland as well as the low logging volume (Greene et al., 1998). In contrast, small logging firms have advantages in harvesting on small size private tracts (Blinn et al., 2015). Unlike large firms, small logging firms do not own mechanized equipment and employ much staff to take advantage of the internal economies of scale (Shivan et al., 2020). Compared with large firms, small firms are more inclined to operate seasonally and reduce capital expenses to maintain efficiency. Small firms are also less likely to afford the cost of equipment repair and maintenance (LeBel and Stuart, 1998). As a result, small logging firms are 'hand crews', they hire employees to harvest timber by chain saw instead of harvesting systems (Egan, 2011).

4.4.1.4 $Q \xrightarrow{+} W \xrightarrow{-} K$

One of the possible reasons for this situation is that although high wages provide incentives for logging firms to replace labor with machines, high wages have made it difficult for firms to accumulate capital. Logging firms must remain profitable to remain in business and continue investing in their businesses, but the high wage reduces their profitability (Jacobson et al., 2009). The largest contribution to costs is due to the wage. Wages account for an average of about 30% of the total costs in the South (Xu et al., 2014).

As a result, confronting high wages, the logging firms, especially the small logging firms, cannot afford the wage costs of a large team, and they would outsource some production processes instead of investing in mechanized systems, which can cope with the low production level (Stuart et al., 2010). Logging firms tend to contract labor-intensive activities because they can alleviate their workload and keep a small staff crew (Wang, 1999), thus saving the salary costs. Trucking is a large cost for logging firms (Yin and Caulfield, 2002) and most small firms cannot afford large investments and expenses of trucking, so they would choose to contract out trucking, which can be much cheaper than operating their fleets and focus on the harvesting business (Hamsley et al., 2007; Shivan et al., 2020).

4.4.1.5 $K \xrightarrow{-} P$

In fact, logging firms can benefit from mechanization by using more equipment and technology to reduce product prices and become more cost-efficient (Mac Donagh et al., 2017). In the past two decades, logging firms have been promoting the mechanization of harvested systems. The proportion of loggers using the capital-intensive mechanized harvesting systems has increased over time, making logging a much more capital-intensive industry (Kollberg, 2005).

The mechanization process enables logging firms to obtain higher productivity. The productivity of the logging firms in most states has increased since 1997 because of the widespread use of mechanized harvested systems (He et al., 2021). The mechanized firms in Wisconsin produced 0.73 million cubic feet per year on an average, while the non-mechanized firms produced only 0.23 million cubic feet per year (Rickenbach and Steele, 2005). Previous research shows that due to significant capital investments, production increased from 3.4 to 5.5 tons per person-hour between 1987 and 2012 in the South (Greene et al., 2013). According to a survey conducted in the northeastern U.S., the average unit cost of fully mechanized crews was 0.795 dollars/ cubic foot, while that of the hand-felling crews was 0.947 dollars/ cubic foot (Kelly et al., 2017). Therefore, the mechanized harvested systems contributed to increasing productivity, decreasing the average per-unit cost (Cubbage and Carter, 1994), and thus the per-unit logging product price decreased.

4.4.2 Forecast Error Variance Decomposition Analysis

The wage keeps the stable influence on capital. This result might be explained by the fact that the increase in wage has been long-term and stable in promoting logging firms to choose to outsource some of their businesses instead of mechanization. On the other hand, employment plays a decreasing role in driving capital with the extension of the forecast period. This result may be explained by the fact that the logging firms need to consider costs, markets, profitability, and ease of obtaining loans to decide whether to purchase machines in the long-run. Therefore, the impact of employment on capital, in the long-run, is getting lower and lower. The product price is determined mainly by itself in the short-run, while this endogenous shock gradually decreases in the long-run. This relationship may partly be explained by price rigidities, such as menu costs and other frictions in adjusting prices (Angeletos and Jennifer, 2009), or imperfect information (Lucas, JR et al., 2012; Mankiw and Reis, 2002).

4.5 Conclusion

This study focuses on investigating the driving factors for employment in the logging industry in the U.S. from 2007 to 2017. A DAG approach was applied to study the contemporaneous causal relations among employment, wage, mechanization, logging product prices, and production levels. The VD was then employed to analyze the dynamic relationship among variables.

The results of the DAG analysis show that there are two conduct paths affecting employment in the logging industry. First, the production level directly impacts employment and has a positive impact on employment. Second, the production level drives wages, and wages promote employment. In addition, the production level affects mechanization through its impact on employment and wage, followed by the influence of mechanization on the product price. The VD results based on the DAG and VAR model verify that employment is most prominently explained by the production level, followed by the wage. The wage is not only influenced by itself but also by the production level. The wage and employment influence mechanization largely. The product price is mainly influenced by itself in the short-run, while this endogenous shock gradually decreases in the long-run.

Based on the previous empirical results, we put forth the following policy implications and suggestions. First, if the policy goal is to promote employment and maintain employment stability, then increasing logging production supplies from small and medium-sized logging firms would be helpful since they are more inclined to hire employees to reinforce the competitive position of large firms rather than mechanization. A broadly similar point has also been made by Lee and Eckert (2002). Second, although mechanization can solve the shortage of employment and reduce product prices, many logging firms still hire employees instead of purchasing machines because consideration of several conditions was needed to achieve mechanization: low loan interest, ease of obtaining loans, efficient equipment maintenance, reasonable operating costs, and production level compatible with mechanized harvesting systems (Cook et al., 2021; He et al., 2021; Mac Donagh et al., 2017). As a result, if the policy goal is to promote the mechanization of the logging industry, for example, to reduce the price of timber products, then policymakers at least need to address these obstacles, providing tax breaks, loan concessions, and fiscal subsidies for those firms which are going to purchase mechanized logging system. Third, with the advancement of mechanization, the logging workers also need to keep up with technological progress. The policymakers can offer some skills training programs to increase the number of qualified logging machine operators, for example, assisting the 'hand crews' transform into mechanized crews.

4.6 References

Abbas, D., Handler, R., Hartsough, B., Dykstra, D., Lautala, P., Hembroff, L., 2014. A survey analysis of forest harvesting and transportation operations in Michigan. Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering 35, 179–192.

Allred, S., Michler, C., Mycroft, C., 2011. Midwest logging firm perspectives: Harvesting on increasingly parcelized forestlands. International Journal of Forestry Research 2011. https://doi.org/10.1155/2011/320170.

Alsaedi, Y.H., Tularam, G.A., 2020. The relationship between electricity consumption, peak load and GDP in Saudi Arabia: A VAR analysis. Mathematics and Computers in Simulation 175, 164–178. https://doi.org/10.1016/j.matcom.2019.06.012.

Angeletos, G.-M., Jennifer, L., 2009. Incomplete information, higher-order beliefs and price inertia. Journal of Monetary Economics 56, S19-S37. https://doi.org/10.1016/j.jmoneco.2009.07.001.

Bailey, M.R., Crandall, M.S., Kizha, A.R., Green, S., 2020. The Economic Contribution of Logging and Trucking in Maine.

Baker, S.A., Greene, W.D., 2008. Changes in Georgia's logging workforce, 1987-2007. Southern Journal of Applied Forestry 32, 60–68. https://doi.org/10.1093/sjaf/32.2.60.

Bernanke, B.S., Gertler, M., 1995. Inside the black box: the credit channel of monetary policy transmission. Journal of Economic perspectives 9, 27–48. https://doi.org/10.1257/jep.9.4.27.

Bessler, D.A., Yang, J., 2003. The structure of interdependence in international stock markets. Journal of international money and finance 22, 261–287. https://doi.org/10.1016/S0261-5606(02)00076-1.

Blinn, C.R., O'Hara, T.J., Chura, D.T., Russell, M.B., 2015. Minnesota's Logging Businesses: An Assessment of the Health and Viability of the Sector. Forest Science 61, 381–387. https://doi.org/10.5849/forsci.14-013.

Bolding, M.C., Barrett, S.M., Munsell, J.F., Groover, M.C., 2010. Characteristics of Virginia's logging businesses in a changing timber market. Forest products journal 60, 86–93. https://doi.org/10.13073/0015-7473-60.1.86.

Boltz, F., Holmes, T.P., Carter, D.R., 2003. Economic and environmental impacts of conventional and reduced-impact logging in Tropical South America: a comparative review. Forest Policy and Economics 5, 69–81. https://doi.org/10.1016/S1389-9341(01)00075-2.

Breitung, J., 2002. Nonparametric tests for unit roots and cointegration. Journal of econometrics 108, 343–363. https://doi.org/10.1016/S0304-4076(01)00139-7.

Bureau of Labor Statistics, 2020. Occupational Outlook Handbook: Logging Workers. https://www.bls.gov/ooh/farming-fishing-and-forestry/logging-workers.htm (accessed 8 December 2020).

Chen, P., He, L., Yang, X., 2021. On interdependence structure of China's commodity market. Resources Policy 74, 102256. https://doi.org/10.1016/j.resourpol.2021.102256.

Conrad, J., Greene, D. (Eds.), 2017. A Review of Logging Business Characteristics: Comparisons Across Time and Between Regions.

Conrad IV, J.L., Greene, W.D., Hiesl, P., 2018. A Review of Changes in US Logging Businesses 1980s–Present. Journal of Forestry 116, 291–303. https://doi.org/10.1093/jofore/fvx014.

Cook, E.B., Bolding, M.C., Visser, R., Barrett, S.M., O'Neal, B.S., 2021. Assessing the characteristics and supporting factors that lead to logging machine value loss in the Southeastern United States. International Journal of Forest Engineering, 1–10. https://doi.org/10.1080/14942119.2021.1971145.

Cubbage, F., Carter, D., 1994. Productivity and cost changes in southern pulpwood harvesting, 1979 to 1987. Southern Journal of Applied Forestry 18, 83–90. https://doi.org/10.1093/sjaf/18.2.83.

Drapala, P., 2009. Decline in Housing Market Hits Forestry Industry Hard. Mississippi Agricultural News.

Duc, N.M., Shen, Y., Zhang, Y., Smidt, M., 2009. Logging productivity and production function in Alabama, 1995 to 2000. Forest products journal 59. https://doi.org/10.1093/njaf/28.4.180.

Egan, A., 2011. Characteristics of and challenges faced by logging business owners in Southern New England. Northern Journal of Applied Forestry 28, 180–185.

Egan, A., Taggart, D., 2009. Public perceptions of the logging profession in Maine and implications for logger recruitment. Northern Journal of Applied Forestry 26, 93–98. https://doi.org/10.1093/njaf/26.3.93.

Enders, W., 2008. Applied econometric time series. John Wiley & Sons.

Esmaeili, P., Rafei, M., 2021. Dynamics analysis of factors affecting electricity consumption fluctuations based on economic conditions: Application of SVAR and TVP-VAR models. Energy 226, 120340.

Greene, W.D., Jackson, B.D., Woodruff, D.C., 1998. Characteristics of logging contractors and their employees in Georgia. Forest products journal 48, 47.

Greene, W.D., Marchman, S.C., Baker, S.A., 2013. Changes in logging firm demographics and logging capacity in the US South, in: Proceedings of the 36th Annual Council on Forest Engineering Meeting, p. 7.

Grushecky, S.T., McGill, D.W., Anderson, R.B., 2006. Inventory of wood residues in southern West Virginia. Northern Journal of Applied Forestry 23, 47–52. https://doi.org/10.1093/njaf/23.1.47.

Haigh, M.S., Bessler, D.A., 2004. Causality and price discovery: An application of directed acyclic graphs. The Journal of Business 77, 1099–1121. https://doi.org/10.1086/422632.

Hamsley, A.K., Greene, W.D., Siry, J.P., Mendell, B.C., 2007. Improving timber trucking performance by reducing variability of log truck weights. Southern Journal of Applied Forestry 31, 12–16. https://doi.org/10.1093/sjaf/31.1.12.

Harris, R.D.F., Tzavalis, E., 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. Journal of econometrics 91, 201–226. https://doi.org/10.1016/S0304-4076(98)00076-1.

He, M., Smidt, S., Li, W., Zhang, Y., 2021. Logging Industry in the United States: Employment and Profitability. Forests 12.

Jacobson, M., Finley, J., Schmid, C., 2009. Factors and Trends in Pennsylvania's Logging Industry.

Ji, Q., Zhang, H.-Y., Geng, J.-B., 2018. What drives natural gas prices in the United States? – A directed acyclic graph approach. Energy Economics 69, 79–88. https://doi.org/10.1016/j.eneco.2017.11.002.

Jolley, G.J., Khalaf, C., Michaud, G.L., Belleville, D., 2020. The economic contribution of logging, forestry, pulp & paper mills, and paper products: A 50-state analysis. Forest Policy and Economics 115, 102140. https://doi.org/10.1016/j.forpol.2020.102140.

Kao, C., 1999. Spurious regression and residual-based tests for cointegration in panel data. Journal of econometrics 90, 1–44. https://doi.org/10.1016/S0304-4076(98)00023-2.

Kelly, M.C., Germain, R.H., Bick, S., 2017. Impacts of forestry best management practices on logging costs and productivity in the northeastern USA. Journal of Forestry 115, 503–512. https://doi.org/10.5849/JOF.2016-031R1.

Kollberg, M., 2005. Beyond IT and Productivity: Effects of Digitized Information Flows in the Logging Industry. Linköping, Sweden, 197 pp.

LeBel, L.G., Stuart, W.B., 1998. Technical Efficiency Evaluation of Logging Contractors Using a Nonparametric Model 9, 15–24. https://doi.org/10.1080/08435243.1998.10702714.

Lee, R.G., Eckert, P.J., 2002. Establishment size and employment stability in logging and sawmilling: a comparative analysis. Can. J. For. Res. 32, 67–80. https://doi.org/10.1139/x01-146.

Leon, B.H., Benjamin, J.G., 2012. A survey of business attributes, harvest capacity and equipment infrastructure of logging businesses in the northern forest. School of Forest Resources, University of Maine, Orono.

Lucas, R.E., JR, Lucas, R.E., Gillman, M., 2012. 1 Expectations and the Neutrality of Money, in: Collected Papers on Monetary Theory. Harvard University Press, pp. 1–24.

Mac Donagh, P., Botta, G., Schlichter, T., Cubbage, F., 2017. Harvesting contractor production and costs in forest plantations of Argentina, Brazil, and Uruguay. International Journal of Forest Engineering 28, 157–168. https://doi.org/10.1080/14942119.2017.1360657.

Maddala, G.S., Wu, S., 1999. A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and statistics 61, 631–652. https://doi.org/10.1111/1468-0084.0610s1631.

Mankiw, N.G., Reis, R., 2002. Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. The Quarterly Journal of Economics 117, 1295–1328. https://doi.org/10.1162/003355302320935034.

McCoskey, S., Kao, C., 1998. A residual-based test of the null of cointegration in panel data. Econometric Reviews 17, 57–84. https://doi.org/10.1080/07474939808800403.

McKenzie, A.M., Goodwin, H.L., Carreira, R.I., 2009. Alternative model selection using forecast error variance decompositions in wholesale chicken markets. Journal of Agricultural and Applied Economics 41, 227–240.

Milauskas, S.J., Wang, J., 2006. West Virginia logger characteristics. Forest products journal 56, 19.

Moskalik, T., Borz, S.A., Dvořák, J., Ferencik, M., Glushkov, S., Muiste, P., Lazdiņš, A., Styranivsky, O., 2017. Timber harvesting methods in Eastern European countries: A review. Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering 38, 231–241.

Olagunju, K.O., Feng, S., Patton, M., 2021. Dynamic relationships among phosphate rock, fertilisers and agricultural commodity markets: Evidence from a vector error correction model and Directed Acyclic Graphs. Resources Policy 74, 102301. https://doi.org/10.1016/j.resourpol.2021.102301.

Omisakin, Olusegun, A., 2008. Oil price shocks and the Nigerian economy: a forecast error variance decomposition analysis. Journal of Economic Theory 2, 124–130.

Pan, X., Ai, B., Li, C., Pan, X., Yan, Y., 2019. Dynamic relationship among environmental regulation, technological innovation and energy efficiency based on large scale provincial panel data in China. Technological Forecasting and Social Change 144, 428–435. https://doi.org/10.1016/j.techfore.2017.12.012.

Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82, 669–688. https://doi.org/10.1093/biomet/82.4.669.

Pedroni, P., 2004. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. Econometric theory 20, 597–625. https://doi.org/10.1017/S0266466604203073.

Rephann, T.J., 2008. The economic impact of agriculture and forestry on the Commonwealth of Virginia. Center for Economic and Policy Studies.

Rickenbach, M., Steele, T.W., 2005. Comparing mechanized and non-mechanized logging firms in Wisconsin: Implications for a dynamic ownership and policy environment. Forest products journal 55, 21.

Rickenbach, M., Steele, T.W., 2006. Logging firms, nonindustrial private forests, and forest parcelization: evidence of firm specialization and its impact on sustainable timber supply. Can. J. For. Res. 36, 186–194. https://doi.org/10.1139/x05-238.

Scott, E., Hirabayashi, L., Graham, J., Franck, K., Krupa, N., Jenkins, P., 2020. Health and safety in the Maine woods: Assemblage and baseline characteristics of a longitudinal cohort of logging workers. American journal of industrial medicine 63, 907–916. https://doi.org/10.1002/ajim.23165.

Sheng, H.-C., Tu, A.H., 2000. A study of cointegration and variance decomposition among national equity indices before and during the period of the Asian financial crisis. Journal of Multinational Financial Management 10, 345–365. https://doi.org/10.1016/S1042-444X(00)00034-7.

Shivan, G.C., Potter-Witter, K., Blinn, C.R., Rickenbach, M., 2020. The Logging Sector in the Lake States of Michigan, Minnesota, and Wisconsin: Status, Issues, and Opportunities. Journal of Forestry, fvaa021. https://doi.org/10.1093/jofore/fvaa021.

Sims, C.A., 1980. Macroeconomics and reality. Econometrica: journal of the Econometric Society, 1–48. https://doi.org/10.2307/1912017.

Spirtes, P., Glymour, C.N., Scheines, R., Heckerman, D., 2000. Causation, prediction, and search. MIT press.

Stuart, W.B., Grace, L.A., Grala, R.K., 2010. Returns to scale in the Eastern United States logging industry. Forest Policy and Economics 12, 451–456. https://doi.org/10.1016/j.forpol.2010.04.004.

Swanson, N.R., Granger, C.W.J., 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. Journal of the American Statistical Association 92, 357–367. https://doi.org/10.1080/01621459.1997.10473634.

Taggart, D., Egan, A., 2011. Logging across borders and cultures: An example in northern Maine. Forest products journal 61, 561–569. https://doi.org/10.13073/0015-7473-61.7.561.

Wang, S., 1999. Silvicultural Contracting in British Columbia: A Transaction Cost Economics Analysis. Forest Science 45, 272–279. https://doi.org/10.5558/tfc74899-6.

Xu, Y., Smidt, M., Zhang, Y., 2014. Logging worker wage, performance, and experience. Forest products journal 64, 210–216. https://doi.org/10.13073/FPJ-D-14-00035.

Yin, R., 2001. Spotted owl protection, booming housing market, and log price changes in the Pacific Northwest. Natural Resource Modeling 14, 575–592. https://doi.org/10.1111/j.1939-7445.2001.tb00074.x.

Yin, R., Caulfield, J.P., 2002. A profile of timber markets in the US Southeast. Forest products journal 52, 25.

Yin, R., Caulfield, J.P., Aronow, M.E., Harris Jr, T.G., 1998. Industrial timberland: current situation, holding rationale, and future development. Forest products journal 48, 43. https://doi.org/10.1111/j.1939-7445.2001.tb00074.x.

Chapter 5. Nowcasting of lumber futures price with Google Trends Index: Using Machine Learning and Deep Learning Models

5.1 Introduction

Lumber futures have been traded at Chicago Mercantile Exchange since 1969 (Mehrotra and Carter 2017). Since the COVID-19 pandemic, the lumber futures price has experienced huge volatility. Figure 14 plots the daily opening price of lumber futures from May 3, 2011, to May 28, 2021. The opening price of lumber futures plummeted on April 1, 2020, and then returned to the normal level before the pandemic. After that, it continued to climb steeply and finally reached its highest point in 10 years on May 7, 2021, with $1,677 per thousand board feet (mbf). It was $425.9 per mbf on January 21, 2020, when the first COVID-19 case in the United States was confirmed (Sahu and Kumar 2020). The average opening price from May 2011 to January 2020 was $337 per mbf, while the average opening price from February 2020 to May 2021 was $698 per mbf. The unusual fluctuations exposed lumber futures products that were originally designed to hedge uncertainties to huge risks. Therefore, there is an urgent need to find a reliable method to predict the lumber futures price, which would help enterprises and investors hedge risks and make correct decisions in the market.

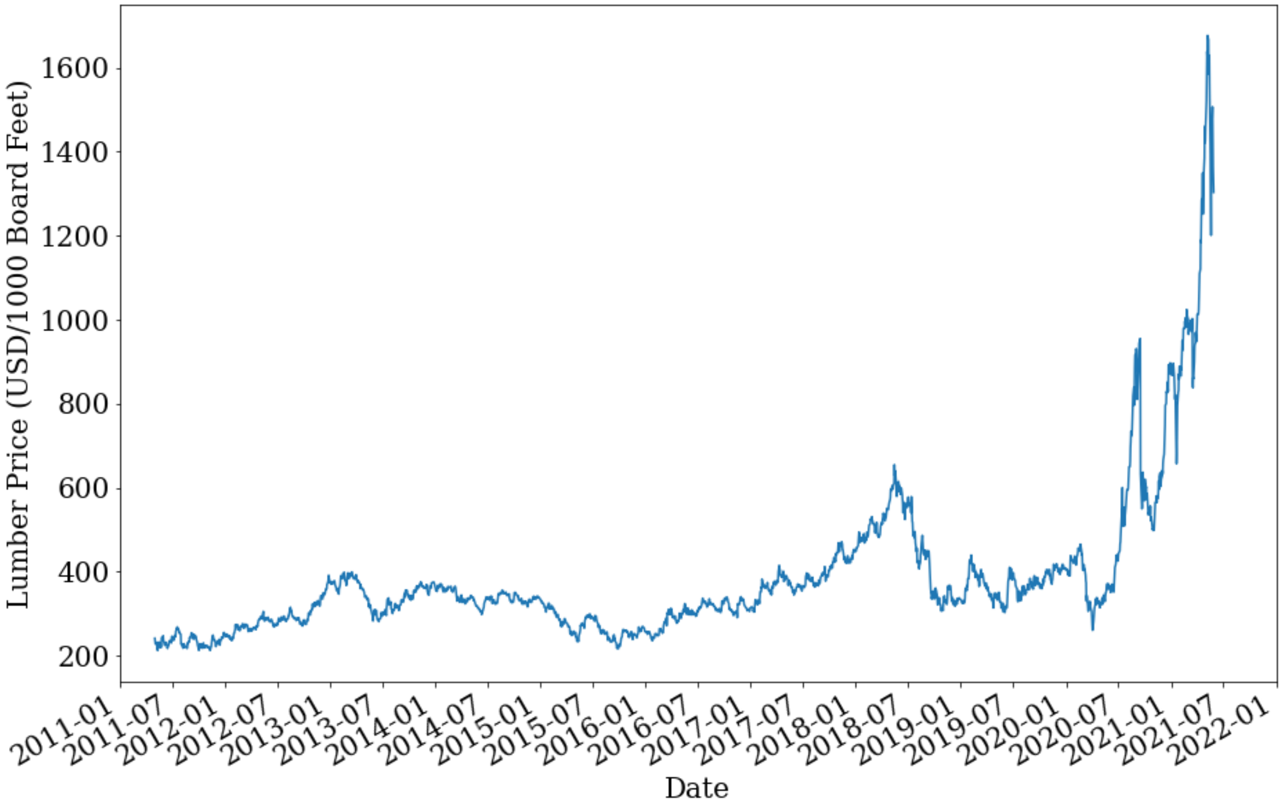

Figure 14. The opening prices of lumber futures, US, May 2011 -May 2021.

In recent decades, several lumber price prediction methods have been proposed, such as ordinary least-squares regression (Mehrotra and Carter 2017), vector autoregressive model (VAR) (Song 2006), autoregressive integrated moving average model (ARIMA) (Buongiorno and Balsiger 1977, Oliveira et al. 1977, Banaś and Utnik-Banaś 2021), seasonal autoregressive moving average model (SARIMA) (Banaś and Utnik-Banaś 2021), seasonal autoregressive moving average model with exogenous variables (SARIMAX) (Banaś and Utnik-Banaś 2021), forest simulation model (FORSIM) (Buongiorno et al. 1984), and sales & operations planning network model (Marier et al. 2014). Most of the literature on lumber price prediction is based on traditional statistical models (Marier et al. 2014), econometric models (Banaś and Utnik-Banaś 2021, Buongiorno and Balsiger 1977, Mehrotra and Carter 2017, Oliveira et al. 1977, Song 2006), or mathematical models (Buongiorno et al. 1984). So far, only one paper has used a recurrent neural networks model, which is a deep learning method to predict the closing price of lumber futures in the next few days using the price obtained from the previous few days (Verly Lopes et al. 2021).

In other domains, machine learning models and deep learning models were widely used for time series forecasting. A support vector machine (SVM) method was employed to forecast the daily electrical load (Singh and Mohapatra 2021) or wind speed (Gangwar et al. 2020). A random forest method was conducted in other studies to estimate poverty (Zhao et al. 2019) or the biomass weight of wheat (Zhou et al. 2016). XGBoost was run to forecast crude oil price (Gumus and Kiran 2017) or sales of the enterprise (Gurnani et al. 2017, Ji et al. 2019). Classification and regression tree (CART) was carried out to forecast precipitation (Choubin et al. 2018) or currency exchange rate (Haeri et al. 2015). And the deep learning models, including artificial neural network (ANN), recurrent neural network (RNN), and convolutional neural network (CNN), were applied to forecast construction material prices (Mir et al. 2021), photovoltaic power (Abdel-Nasser and Mahmoud 2019), gas demand (Su et al. 2019a), stock markets (Hoseinzade and Haratizadeh 2019), or river discharges (Awchi 2014). Overall, machine learning models and deep learning models have been widely employed to predict economic indicators, socioeconomic indicators, and science indicators. Machine learning models and deep learning models are statistical approaches. Compared to the traditional econometric models, they capture the hidden nonlinear characteristics among variables and provide more accurate predictions, while the econometric models are based on strict linear assumptions (Herrera et al. 2019) and might overfit the sample and yield forecasting error (Shobana and Umamaheswari 2021).

Some previous studies predicted the future lumber price based on the past values, which is an autoregressive technique (Song 2006). Other studies use some exogenous independent variables to predict the lumber prices, such as the construction confidence index (Banaś and Utnik-Banaś 2021) and specific characteristics of the lumber supply chain (Marier et al. 2014). Models that include exogenous independent variables can produce good prediction results because the exogenous variables normally contain more information. However, none of these studies included public attention as an exogenous variable. Google is the most popular search engine in the United States. Google Trends is a publicly available service provided by Google. It provides access to aggregated information about different search queries and how those queries change over time. The Google Trends index is an index measuring search volume of different queries over time. Users can use the Google Trends index to observe changes in the query volume of certain keywords over time and compare the query volume of different keywords over time. This provides an opportunity to capture the interest and concern of the public in real time without any cost. Therefore, Google Trends index is widely used to predict economic indicators and socioeconomic indicators, such as sales, unemployment, travel, consumer confidence (Choi and Varian 2012), consumer behavior (Carrière-Swallow and Labbé 2013), housing market (Dietzel 2016), the stock price (Hu et al. 2018), and so on.

This prospective study aims to use the Google Trends index of some keywords from the previous day to predict the next day's opening price of lumber futures. Nowcasting is the process of predicting the present, the very near future, or the very recent past value of an indicator based on real-time data (Banbura et al. 2010, Chumnumpan and Shi 2019). Nowcasting the opening price of lumber futures can help investors to take appropriate actions during the premarket trading hours between 8:00 a.m. to 9:30 a.m. Eastern each trading day. It would have a beneficial impact on hedging risks and expanding trade opportunities (Dungey et al. 2009). It would also be useful in helping enterprises navigate during normal and unusual times such as a pandemic. The statistical significance of the keywords of the Google Trends index will change over time. In other words, different factors have various effects on lumber futures price in different situations. The models can dynamically select the keyword variables in different time periods. As a result, the components of variables will change to capture dynamic trends of the real world. This study fills the gap in the literature by using machine learning and deep learning models to nowcast the lumber futures prices via Google Trends index.

This section consists of five sections. The “Data” section briefly introduces the data. The “Prediction Models” section describes the models adopted in this study. The “Results and Discussion” section presents and discusses the results, and the “Conclusion” section concludes this study.

5.2 Data

5.2.1 Data collection

The Chicago Mercantile Exchange lumber futures price daily data were extracted from Investing.com. The dataset includes opening price, closing price, highest price, and lowest price of lumber futures. The data are from May 2011 to May 2021, with a total of 2,523 entries of data. The opening price of lumber futures is plotted in Figure 14.

The actual Google search requests for some lumber price–related keywords were then extracted from Google Trends index to match the same time series as the lumber price datasets. Keyword variables include 2 by 4 (a length of sawn wood 2 inches thick and 4 inches wide), BDFT (board foot), CLT (cross-laminated timber), commodity, DIY (do it yourself), fire, forest products association, forestry, hardwood, harvest, home building, home improvement, home renovation, invest, logging, logs, lumber futures, lumber price, lumber yard, MDF (medium density fiberboard), OSB (oriented strand board), plywood, sawmill, softwood, stock market, timber, and wood. Research has seen an effect on the lumber prices for a reduction in the quality of softwood lumber or in that case any lumber. Hence, more general keywords were included instead of the specific kinds of lumber. For example, the Southern pine and Douglas-fir lumber, which are the two most commercially important types of softwood lumber, have not changed in strength and stiffness over the last five decades (Miyamoto et al. 2018, França et al. 2021, Shmulsky et al. 2021, Babula and Zhang 2019, Babula et al. 2012, Zhang and Sun 2001), and thus they were not included in the keywords.

Google Trends index will standardize the data to a scale of 0 to 100 to represent the "interest over time." But the scale of this data set will change if the same variable is colisted with other keywords or if the time range is changed. Therefore, it is important to always extract the same combination of words in the same time range during the modeling and prediction process to avoid restandardization of the same data set to different scales. However, the many years of daily keyword data cannot be downloaded directly from Google Trends. In order to avoid restandardization, application programming interface (API) was applied to extract Google Trends index data via R. The library "gtrendsR" on R was employed to extract the Google Trends index, and it retrieves the index via APIs. The descriptive statistics of opening price and closing price of lumber futures price and the whole Google Trends index of keywords is provided in Table 11.

Table 11. Descriptive statistics of lumber price and Google Trends Index, US, May 2011 – May 2021.

		Count	Mean	Std	Min	25%	50%	75%	Max
Google Trends Index	2 by 4	2523	0.014	0.017	0.000	0.004	0.007	0.018	0.148
	bdft	2523	0.003	0.024	0.000	0.000	0.000	0.000	0.500
	clt	2523	0.185	0.093	0.000	0.111	0.174	0.249	0.500
	commodity	2523	0.060	0.027	0.007	0.040	0.057	0.073	0.260
	DIY	2523	2.053	1.362	0.200	1.210	1.950	2.400	12.420
	fire	2523	31.476	9.235	7.360	25.440	30.680	36.580	100.000
	forest products association	2523	0.013	0.073	0.000	0.000	0.000	0.000	0.890
	forestry	2523	0.429	0.131	0.070	0.336	0.420	0.507	1.000
	hardwood	2523	0.082	0.034	0.020	0.060	0.080	0.100	0.210
	harvest	2523	0.357	0.181	0.050	0.240	0.300	0.450	2.240
	home building	2523	0.054	0.018	0.007	0.042	0.052	0.064	0.153
	home improvement	2523	0.041	0.060	0.000	0.010	0.020	0.030	0.400
	home renovation	2523	0.033	0.022	0.000	0.016	0.030	0.047	0.126
	invest	2523	1.523	0.694	0.304	1.035	1.382	1.849	6.000
	logging	2523	0.650	0.130	0.231	0.557	0.650	0.739	0.990
	logs	2523	0.042	0.021	0.010	0.020	0.040	0.060	0.140
	lumber futures	2523	0.0001	0.0007	0.0000	0.0000	0.0000	0.0000	0.0128
	lumber price	2523	0.001	0.004	0.000	0.000	0.000	0.001	0.044
	lumber yard	2523	0.008	0.007	0.000	0.003	0.006	0.011	0.046
	mdf	2523	0.197	0.078	0.018	0.143	0.191	0.248	0.470
	OSB	2523	0.226	0.123	0.000	0.140	0.211	0.291	1.000
	plywood	2523	0.777	0.273	0.207	0.577	0.740	0.918	2.000
	sawmill	2523	0.475	0.148	0.080	0.370	0.466	0.572	0.980
	softwood	2523	0.0002	0.0002	0.0000	0.0000	0.0002	0.0004	0.0028
	stock market	2523	0.707	2.003	0.020	0.150	0.280	0.540	35.000
	timber	2523	1.544	0.333	0.619	1.325	1.530	1.734	3.230
	wood	2523	12.518	4.947	3.630	9.180	11.760	15.360	37.600
Opening Price of Lumber Futures		2523	384.8	186.1	211.9	292.5	336.6	390.5	1677.0
Closing Price of Lumber Futures		2523	384.7	186.7	209.7	292.7	336.3	390.0	1686.0

5.2.2 Variable Selection

To increase the model interpretability, remove redundant or irrelevant variables, and reduce overfitting, least absolute shrinkage and selection operator (LASSO) was first applied to perform independent variable selection (Fonti and Belitser 2017). The LASSO estimate can be written as

$$\hat{\beta} = arg\ \min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{15}$$

where $\lambda \geq 0$ is a constant parameter that controls the strength of regularization. The value of λ is directly proportional to the amount of regularization (Muthukrishnan and Rohini 2016, Fonti and Belitser 2017). In the LASSO process, the variables that have nonzero coefficients after the regularization are selected as part of the model (Fonti and Belitser 2017). As a result, the lumber futures closing price, and the Google Trends index of the four terms “2 by 4,” “commodity,” “invest,” and “lumber futures” were selected as the feature inputs of the models (Table 12). Figure 15 plots the daily Google Trends index of the above keywords from May 3, 2011, to May 28, 2021.

Table 12. LASSO results.

Variable	LASSO	Variable	LASSO	Variable	LASSO
Close	0.002241	harvest	0	mdf	0
2 by 4	0.063729	home building	0	OSB	0
bdft	0	home improvement	0	plywood	0
clt	0	home renovation	0	sawmill	0
commodity	-0.35150	invest	0.007146	softwood	0
DIY	0	logging	0	stock market	0
fire	0	logs	0	timber	0
forest products association	0	lumber futures	-8.96809	wood	0
forestry	0	lumber price	0		
hardwood	0	lumber yard	0		

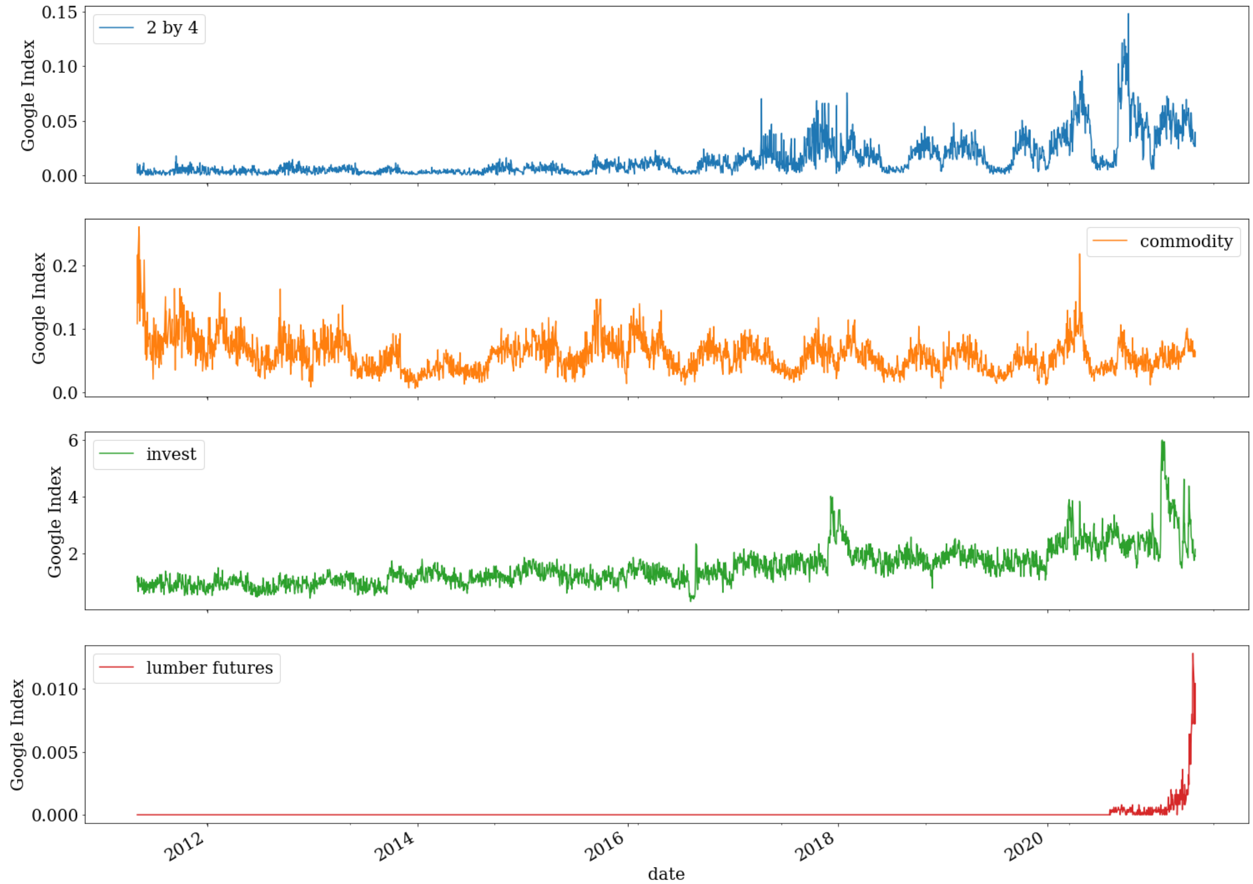

Figure 15. Google Trends Index after LASSO, US, May 2011 -May 2021.

5.2.3 Sample Splitting

Before building up the models, the dataset was divided into two subsets: a training set and a test set, which can avoid overfitting the models and improve the accuracy of the models (LeCun et al. 2015, Roelofs et al. 2019). The models will be trained on the training set, and the fitted models will be used to estimate the predicted value in the test set, which can provide an evaluation of the models. The different splitting rate of the data set is selected in respect to the object of characteristics of the studied subjects (Tao et al. 2020, Nguyen et al. 2021) and the sample size (Tai et al. 2019). In this study, considering that the lumber price does not fluctuate abnormally until the second half of 2020 and there are thousands of entries of samples, the splitting rate of the data set is determined to be 95 percent. The training set and the test set contain 95 and 5 percent of the total sample, respectively, which means the data of the first nine and a half years (May 3, 2011, to November 24, 2020) was used as the training set, and the data of the last six months (November 25, 2020, to May 28, 2021) will be used as the test set.

5.3 Prediction models

Machine learning (ML) models and deep learning (DL) models have emerged with the advent of big data technology and gained in popularity as frontier prediction methods (Liakos et al. 2018). Machine learning models are the algorithms of providing machines the ability to optimize the performance without being strictly programmed (Schmidt et al. 2019, Kadam et al. 2020). Machine learning models include support vector machine (SVM), random forest, XGBoost, classification and regression trees (CART), and many more (Friedman et al. 2001). Deep learning models are defined as representation-learning algorithms composed of processing units organized in input, hidden layers, and output layers (LeCun et al. 2015, Shrestha and Mahmood 2019). Deep learning models include artificial neural network (ANN), recurrent neural network (RNN), and convolutional neural network (CNN) (Miotto et al. 2018).

5.3.1 Machine Learning Models

5.3.1.1 Support Vector Machine

Support Vector Machine is an algorithm that maximizes a specific mathematical function based on a given data set (Noble, 2006). SVM can be applied to time series prediction by introducing kernel functions (Pyo et al., 2017). In the SVM, the input vector x is mapped to the high dimensional feature space using the nonlinear mapping function $\Phi(x)$ and run regression in the space (Wang et al., 2008). The SVM can be represented as the following equation:

$$\widehat{y_{SVM}} = b + \sum_i^n w_i \Phi_i(x) \quad (16)$$

where $\widehat{y_{SVM}}$ is the predicted value, parameters b and w_i can be estimated by minimizing the regularized risk function:

$$R(C) = C \frac{1}{n} \sum_{i=1}^n L_\varepsilon(y, \widehat{y_{SVM}}) + \frac{1}{2} \|w\|^2 \quad (17)$$

where C is a regularization constant, y is the actual value, L_ε is the loss function, $\frac{1}{2} \|w\|^2$ is a measurement of function flatness. By introducing the kernel function $K(x, y)$, the Eq. (17) can be transformed into the explicit form:

$$f_{SVM}(x, \partial_i, \partial_i^*) = \sum_{i=1}^n (\partial_i - \partial_i^*) K(x, x_i) + b \quad (18)$$

where ∂_i and ∂_i^* are the Lagrange multipliers which satisfy the condition: $\partial_i \times \partial_i^* = 0$, $\partial_i \geq 0$ and $\partial_i^* \geq 0$ (Choudhry and Garg, 2008; Wang et al., 2008). In this study, the $K(x, x_i)$ is the polynomial kernel function:

$$K(x, x_i) = (x x_i)^3 \quad (19)$$

where x_i is the sample in the training set (Choudhry and Garg, 2008).

5.3.1.2 Random Forest

Random Forest is an algorithm that obtains the output by combining many decision trees to form forests (Breiman, 2001). Specifically, it selects a bootstrap sample from the training set, which is selected randomly with replacement, and then obtains the optimal split point to split the node into two subtrees by minimizing MSE, which is called growing a random forest tree, T_m. After creation of M trees, the final output of Random Forest is defined as (Huang and Liu, 2019; Peng et al., 2021; Yoon, 2021):

$$\widehat{y_{RF}} = \frac{1}{M}\sum_{m=1}^{M} T_m(x) \tag{20}$$

5.3.1.3 XGBoost

XGBoost is a regression tree algorithm, which is also called Extreme Gradient Boosting. XGBoost is based on the Gradient Boosting Decision Tree algorithm and applies the addition of regularization terms to control the complexity of the model, which can prevent overfitting and improve the accuracy (Peng et al., 2019). As a result, the objective functions consist of two parts: training loss $L(\theta)$ and regularization $\Omega(\theta)$:

$$obj(\theta) = L(\theta) + \Omega(\theta) \tag{21}$$

where θ is the parameter (Gurnani et al., 2017; Peng et al., 2019). The training loss is defined as:

$$L(\theta) = \sum_{i=1}^{n}(y_i - \widehat{y_{XGB}}_i) \tag{22}$$

where y_i is the actual value. In the XGBoost, each inner node represents the value of the attribute test, and the leaf node with values represents a decision (Shilong and others, 2021). $\widehat{y_{XGB}}_i$ is the output, which is the sum of all predict values form M trees and can be written in the form:

$$\widehat{y_{XGB}}_i = \sum_{m=1}^{M} f_m(x_i)\,, f_m \in F \tag{23}$$

where m is the number of trees, x_i is the i^{th} training sample, f_m is the value for the m^{th} tree in the functional space F (Peng et al., 2019; Shilong and others, 2021).

The target function can be finally expressed as:

$$obj(\theta) = \sum_{i=1}^{n} L(y_i, \widehat{y_{XGB}}_i) + \sum_{m=1}^{M} \Omega(f_m) \tag{24}$$

5.3.1.4 Classification and Regression Trees

Classification and Regression Trees (CART) is a non-parametric statistical model, which is employed for classification problems or regression problems. If the output variable is continuous, the CART model will generate a regression tree. The CART tree is a hierarchical binary tree that is built up by splitting subsets of the data set by applying all output variables to generate two sub-nodes repeatedly. For determining the splitting, each predictor is evaluated to discover the best cut point, based on the least-squares deviation (LSD) impurity measure, *R(t)* (Mahjoobi and Etemad-Shahidi, 2008; Samadi et al., 2014):

$$R(t) = \frac{1}{N_{\omega(t)}} \sum_{i \in \omega} \omega_i \, f_i (y_i - \bar{y}_{CART}(t))^2 \qquad (25)$$

where $N_\omega(t)$ is the weighted number of records at node t, ω_i is the value of the weighting field for record i, f_i is the value of the repeat field, y_i is the value of the target field, and $\bar{y}_{CART}(t)$ is the mean of the output variable at node t.

5.3.2 Deep Learning Models

5.3.2.1 Artificial Neural Network

The Artificial Neural Network (ANN) Model connects the units called artificial neurons to generate complex networks (Kurbatsky et al., 2014; Su et al., 2019b). In each unit, there is an activation function, f, which applies the input variables, x_i, to generate the output value. The output of a unit conveyed to next unit as an input via a weighted connection. Given a unit, j, the output of this unit can be expressed as (Su et al., 2019b):

$$\widehat{y_{ANN_i}} = f_{ANN}\left(\sum_{i=1}^{n} \omega_{ij} x_i + t_j\right) \qquad (26)$$

where ω_{ij} is the connection weights, t_j is the bias term. The activation function, f_{ANN} is rectified linear unit activation function in this study. The ANN model in this study is composed of an input layer, 7 hidden layers, and an output layer. The output layer sums up the output of units from hidden layers. Different values of hyperparameter were tested and the model with the best performance has a batch size 8, epochs 100, an optimizer of Adam, loss function of mean squared error, and one hidden layer with 64 units in this study.

5.3.2.2 Recurrent Neural Network

Recurrent Neural Network (RNN) is a model of Neural Network. It applies the previous values of observations to calculate the future value by connecting the computational units from a directed circle (Moghar and Hamiche, 2020; Selvin et al., 2017). However, the RNN confronts two problems: vanishing gradient and exploding gradient (Bouktif et al., 2018). As a result, Long Short-Term Memory (LSTM) was introduced to solve these problems in this study. The U.S.ually hidden layers were replaced with LSTM cells. The LSTM cells consist of input gate, forget gate, output gate, and cell state, which makes it possible to control the gradient flow and then overcome the vanishing and exploding gradient problems (Bouktif et al., 2018; Selvin et al., 2017). The LSTM cell can be expressed as (Bouktif et al., 2020):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{27}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{28}$$

$$c_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{29}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{30}$$

$$h_t = o_t * \tanh(c_t) \tag{31}$$

where x_t is input vector at time *t*, h_{t-1} and h_t are output vector of hidden units at time *t-1* and time *t*, respectively. f_t, i_t and o_t are forget, input, and output gate vector, respectively. c_t is the cell state vector. W_* and b_* are the weight matrices and bias vector parameters of the LSTM unit, respectively. In this study, the RNN model is composed of an LSTM layer with 500 units and has epochs 50, batch size 9, an optimizer of Adam, and loss function of mean squared error. The activation function and recurrent activation function are hyperbolic tangent activation function and hard sigmoid activation function, respectively.

5.3.2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) is a class of feedforward neural networks, which can be effectively applied in image recognition, natural language processing, and time-series data prediction (Lu et al., 2020). CNN consists of convolution layer, pooling layer, and fully connected layers. It extracts data features via the convolution layer and connects the units locally using the pooling layer, which reduces the redundant features (Chen et al., 2021). Then it converts the features in the previous layers to the final output using fully connected layers, which can be expressed as (Balaji et al., 2018):

$$\widehat{y_{CNN}}_i^j = f_{CNN}(\textstyle\sum_k \widehat{y_{CNN}}_k^{j-1} w_{k,i}^{j-1}) \tag{32}$$

where $\widehat{y_{CNN}}_i^j$ is the output value of unit i at the layer *j*, $\widehat{y_{CNN}}_k^{j-1}$ is the output value of unit *k* at the layer *j-1*, f_{CNN} is the activation function. In this study, the activation function of CNN is rectified linear unit activation function. $w_{k,i}^{j-1}$ is the weight of the connection between unit k at layer *j-1* and unit *i* at layer *j*. In this study, the data is convoluted through a 1 dimensional convolution layer (Conv-1D layer) within 16 units, and then the Max Pooling layer. Next, the data is convoluted through another Conv-1D layer within 32 units, and then the Global Max Pooling layer. The activation function is Rectified Linear Unit. The CNN model has epochs 1500, an optimizer of Adam, and a loss function of mean squared error.

5.3.3 Evaluation of models

To evaluate the performance of these models, the Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Symmetric Mean Absolute Percentage Error (SMAPE) were used as the criteria. The measures are as follows:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{33}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{34}$$

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{nN}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{35}$$

$$\text{SMAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \tag{36}$$

where N is the number of training set samples or test set samples, y_i is a real value at time t, and $\hat{y}_i$ is the corresponding predicted value.

5.4 Results and discussion

In this study, a baseline model was established, based on the naïve forecasting method, to provide the required point of comparison when evaluating all other models.[1] Naïve forecasting is

[1] We have built up a multiple linear regression (MLR) model with the open price at time $t-1$ according to the recommendations. The MSE, MAE, MAPE, and SMAPE of the MLR model are 2067.48, 30.65, 2.89, and 2.90 percent, respectively. Overall, the performance is slightly better than the naïve forecasting, but does not differ substantially. Therefore, we decide to use the naïve forecasting model, the common baseline method in the machine learning research field. This also follows the zero rule algorithm for the baseline method (Choudhary and Gianey 2017).

the method in which actual values in the last period are simply taken as predicted values in this period. In the baseline model, the opening price at the previous time step $t - 1$ was used to be the predicted value at the time step t.[2]

The prediction results of different models of the test set are shown in Figure 16, which contains 127 observations from November 25, 2020, to May 28, 2021. All four machine learning models and three deep learning models showed strong predictive ability because the predicted lumber prices are close to the actual prices.

Figure 16 shows that the random forest, XGBoost, CART, ANN, RNN, and CNN models can capture the trends and dynamics in the test set, while the SVM model fails to identify the pattern in the highest price interval, which makes the nowcasting less accurate. It should be noted that the actual lumber price in the test set is much higher than that in the training set. Most of the machine learning and deep learning models can still capture the trends and identify the pattern. This shows that the machine learning and deep learning models have the ability to extract hidden features among variables in high-dimensional and multivariate data sets in a complex and dynamic environment (Köksal et al. 2011, Wuest et al. 2016).

From the overall performance, the ANN model performs better than other models. There is a large overlap between predicted prices and actual prices, especially for the prediction of an abnormal trend of rapid growth from mid-March 2021 to early May 2021. Moreover, the ANN model provides significantly better predictions than the baseline model. Although the random forest, XGBoost, CART, and RNN models are inferior to ANN, the predicted prices of these

[2] Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan-Quinn information criterion (HQIC) were employed to determine the lag order for the baseline model. Based on the selection criterion of the three models, the Lag 1 was selected because it has the smallest AIC, BIC, and HQIC values.

models were highly consistent with the actual observations. SVM and CNN models have the weakest prediction effects among the machine learning and deep learning models, respectively, although predicted prices of these two models are also roughly close to the actual prices. The SVM model overestimates the lumber price from mid-March to early May significantly, and the CNN model cannot capture the trend of rapid growth very well, compared with the other two deep learning models. This result might be explained by the fact that the CNN model does not depend on any information from previous observations to make a prediction (Selvin et al. 2017).

Figure 17 compares the average prediction performance between machine learning models, deep learning models, and the baseline model. Comparing the predictive performance of all seven models shows that the ANN model performs the best overall. The MSE, MAE, MAPE, and SMAPE of the test set are the lowest among these models. This may be explained by the good self-learning, self-adapting, and self-organizing ability of the ANN model, which can analyze the patterns and rules of observations through training (Su et al. 2019b). The RNN model is the second-best prediction performance model, which could be attributed to the good ability to use information from previous lags to predict the future values by RNN (Selvin et al. 2017). XGBoost gives more accurate predictions than other machine learning models, and it is also the third-best model among all seven models. ANN, RNN, XGBoost, random forest, CART, and CNN models provide more accurate results than the baseline model. In addition, the performance of the machine learning and deep learning models are generally better than traditional time series models. For example, Banaś and Utnik-Banaś (2021) forecasted round wood prices from 2019 Q1 to 2020 Q4 in Poland using ARIMA, SARIMA, and SARIMAX models, whose MAPE was 2.57, 2.20, and 1.75 percent on average, respectively. All the models except for SVM in this study have better performance than the ARIMA model. The ANN, RNN, XGBoost, random forest, and CART models in this study are better than the SARIMA model, and the ANN and RNN are better than the SARIMAX model.

Figure 16 and Figure 17 show that, compared with machine learning models, deep learning models are, on average, more capable of capturing the trends and providing more accurate predictions. This may result from the better overfitting reduce ability of deep learning models. This can also be seen in Figure 17. The fitting performance of the three deep learning models to the training set is worse than that of the machine learning models.

5.5 Conclusion

This study describes a new approach for nowcasting the lumber futures price using Google Trends index through machine learning models (SVM, random forest, XGBoost, and CART) and deep learning models (ANN, RNN, and CNN). We show that deep learning models generally give more accurate predictions than machine learning models. Among the seven models, the ANN model provides the best performance, followed by the RNN model. The comparison with the baseline model shows that the random forest, XGBoost, CART, ANN, RNN, and CNN models provide more accurate predictions than the baseline model. Our findings also imply that the Google Trends index, which reflects the dynamic changes of the interest and attention from the public, can provide enough information to be good predictors in nowcasting lumber futures prices.

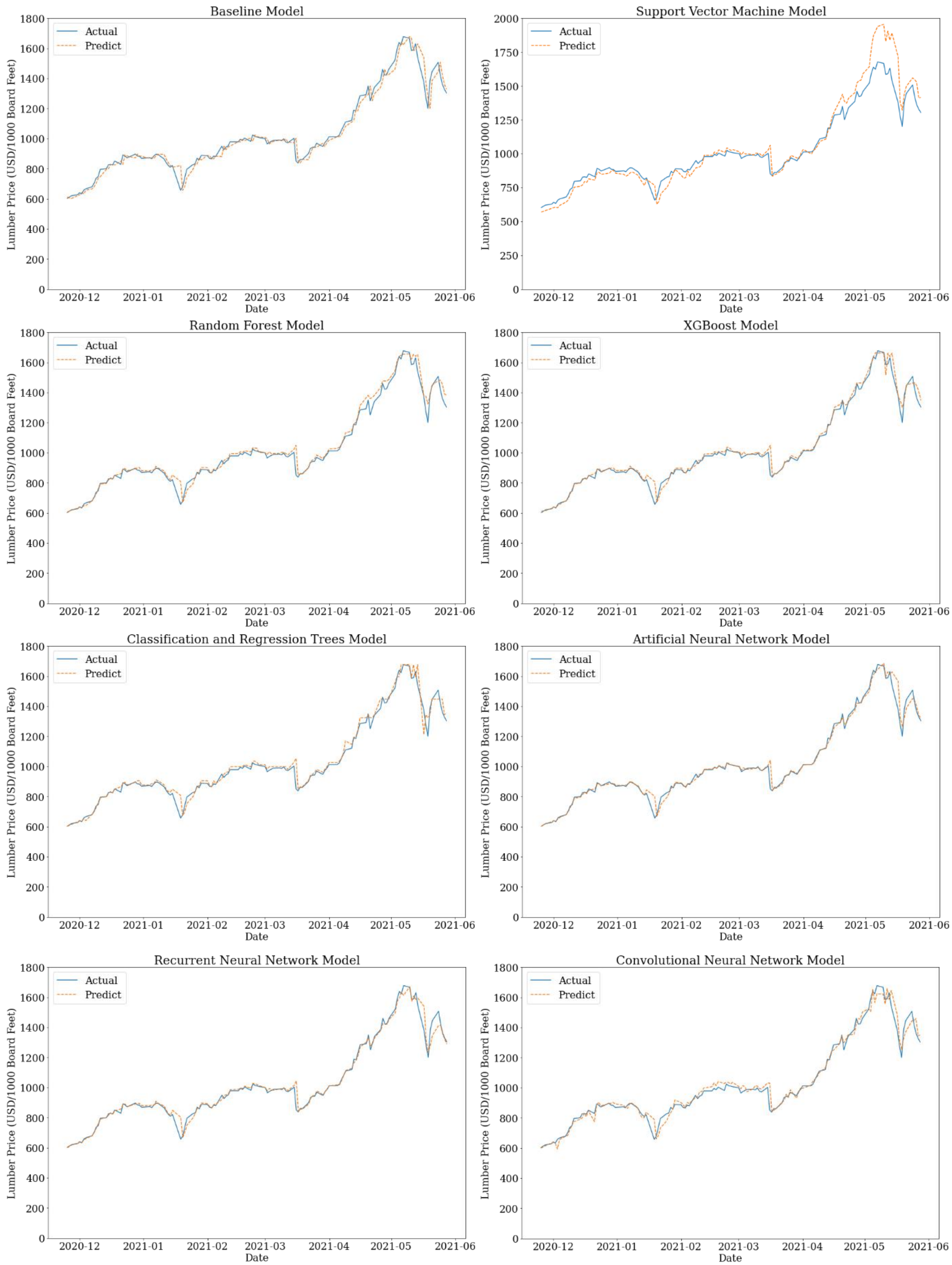

Figure 16. Models Fitting on Test Set.

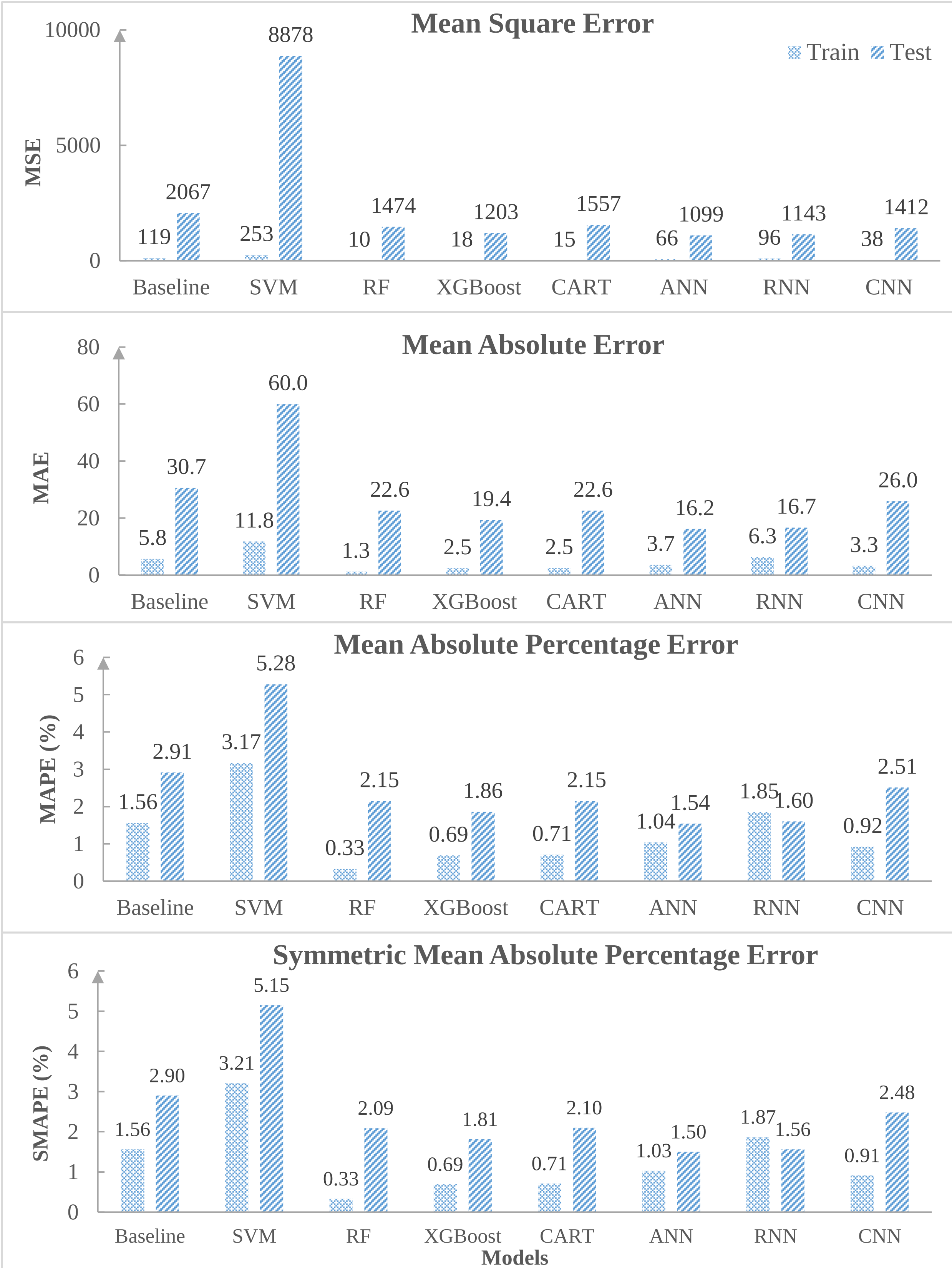

Figure 17. Models Evaluation.

By using the prediction methods and Google Trends index, investors can take appropriate measures to hedge risks and make profits during premarket trading hours. The high predictive power of this approach implies that the big data models should be added to the toolbox of investors and policymakers to predict other economic variables. One probable criticism to these methods being applied to predict the lumber futures price followed by appropriate actions is that it might enhance the lumber futures market volatility and further lead to the invalidation of the forecasting.

5.6 References

Abdel-Nasser, M. and K. Mahmoud. 2019. Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput. Appl. 31(7):2727–2740.

Awchi, T. A. 2014. River discharges forecasting in northern Iraq using different ANN techniques. Water Resour. Manag. 28(3):801–814.

Babula, R.A., Zhang, D., 2019. Assessing the role of futures position substitutability in a monthly model of factor demand for softwood lumber. Empirical Economics 56 (3), 1097–1116.

Babula, R.A., Zhang, D., Rothenberg, J.P., 2012. A Dynamic monthly demand model of US-produced softwood lumber with a futures market linkage. Journal of International Agricultural Trade and Development 8 (2), 149.

Balaji, A. J., D. H. Ram, and B. B. Nair. 2018. Applicability of deep learning models for stock price forecasting an empirical study on BANKEX data. Procedia Computer Sci. 143:947–953.

Banaś, J. and K. Utnik-Banaś. 2021. Evaluating a seasonal autoregressive moving average model with an exogenous variable for short-term timber price forecasting. Forest Policy Econ. 131:102564.

Banbura, M., D. Giannone, and L. Reichlin. 2010. Nowcasting. ECB Working Paper No. 1275.

Bouktif, S., A. Fiaz, A. Ouni, and M. A. Serhani. 2018. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies 11(7):1636.

Bouktif, S., A. Fiaz, A. Ouni, and M. A. Serhani. 2020. Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting. Energies 13(2):391.

Breiman, L. 2001. Random forests. Machine Learning 45(1):5–32.

Buongiorno, J. and J. W. Balsiger. 1977. Quantitative analysis and forecasting of monthly prices of lumber and flooring products. Agric. Syst. 2(3):165–181.

Buongiorno, J., F. Mey Huang, and H. Spelter. 1984. Forecasting the price of lumber and plywood: Econometric model versus futures markets. Forest Prod. J. 34(7).

Carrière-Swallow, Y. and F. Labbé. 2013. Nowcasting with Google trends in an emerging market. J. Forecast. 32(4):289–298.

Chen, Y., R. Fang, T. Liang, Z. Sha, S. Li, Y. Yi, W. Zhou, and H. Song. 2021. Stock price forecast based on CNN-BiLSTM-ECA Model. Scientific Programming 2021:2446543.

Choi, H. and H. Varian. 2012. Predicting the present with Google Trends. Economic Record 88:2–9.

Choubin, B., G. Zehtabian, A. Azareh, E. Rafiei-Sardooi, F. Sajedi-Hosseini, and Ö. Kişi. 2018. Precipitation forecasting using classification and regression trees (CART) model: A comparative study of different approaches. Environ. Earth Sci. 77(8):1–13.

Choudhary, R. and H. K. Gianey. 2017. Comprehensive review on supervised machine learning algorithms. In: 2017 International Conference on Machine Learning and Data Science (MLDS), December 4–15, 2017, Noida, India; IEEE Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 37–43.

Choudhry, R. and K. Garg. 2008. A hybrid machine learning system for stock market forecasting. World Acad. Sci. Eng. Technol. 39(3):315–318.

Chumnumpan, P. and X. Shi. 2019. Understanding new products' market performance using Google Trends. Australasian Marketing J. (AMJ) 27(2):91–103.

Xie, D. and S. Zhang. 2021. Machine learning model for sales forecasting by using XGBoost. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), January 15–17, 2021, Guangzhou, China; IEEE Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 480–483.

Dietzel, M. A. 2016. Sentiment-based predictions of housing market turning points with Google trends. Int. J. Housing Markets Anal. 9(1):108–136.

Dungey, M., L. Fakhrutdinova, and C. Goodhart. 2009. After-hours trading in equity futures markets. J. Futures Markets: Futures, Options, Other Derivative Prod. 29(2):114–136.

Fonti, V. and E. Belitser. 2017. Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics, 30 pp.

França, F. J. N., R. Shmulsky, J. T. Ratcliff, B. Farber, C. A. Senalik, R. J. Ross, and R. D. Seale. 2021. Yellow pine small clear flexural properties across five decades. Forest Prod. J. 71(3):233–239.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. The Elements of Statistical Learning. Springer Series in Statistics, New York.

Gangwar, S., V. Bali, and A. Kumar. 2020. Comparative analysis of wind speed forecasting using LSTM and SVM. EAI Endorsed Trans. Scalable Inf. Syst. 7(25):e1.

Gumus, M. and M. S. Kiran (Eds.). 2017. Crude oil price forecasting using XGBoost. In: 2017 International Conference on Computer Science and Engineering (UBMK), pp. 1100–1103.

Gurnani, M., Y. Korke, P. Shah, S. Udmale, V. Sambhe, and S. Bhirud. 2017. Forecasting of sales by using fusion of machine learning techniques. In: 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), February 24–26 2017, Pune, India; IEEE

Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 93–101.

Haeri, A., S. M. Hatefi, and K. Rezaie. 2015. Forecasting about EUR/JPY exchange rate using hidden Markova model and CART classification algorithm. J. Adv. Comput. Sci. Technol. 4(1):84–89.

Herrera, G. P., M. Constantino, B. M. Tabak, H. Pistori, J.-J. Su, and A. Naranpanawa. 2019. Long-term forecast of energy commodities price using machine learning. Energy 179:214–221.

Hoseinzade, E. and S. Haratizadeh. 2019. CNNpred: CNN-based stock market prediction using a diverse set of variables. Expert Syst. Appl. 129:273–285.

Hu, H., L. Tang, S. Zhang, and H. Wang. 2018. Predicting the direction of stock markets using optimized neural networks with Google Trends. Neurocomputing 285:188–195.

Huang, S. and S. Liu. 2019. Machine learning on stock price movement forecast: The sample of the Taiwan stock exchange. Int. J. Econ. Financial Issues 9(2):189.

Ji, S., X. Wang, W. Zhao, and D. Guo. 2019. An application of a three-stage XGBoost-based model to sales forecasting of a cross-border E-commerce enterprise. Math. Problems Eng. 2019.

Kadam, V. S., S. Kanhere, and S. Mahindrakar. 2020. Regression techniques in machine learning & applications: A review. Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) (10):826–830.

Köksal, G., I. Batmaz, and M. C. Testik. 2011. A review of data mining applications for quality improvement in manufacturing industry. Expert Syst. Appl. 38(10):13448–13467.

Kurbatsky, V. G., D. N. Sidorov, V. A. Spiryaev, and N. V. Tomin. 2014. Forecasting nonstationary time series based on Hilbert-Huang transform and machine learning. Automation Remote Control 75(5):922–934.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. Nature 521(7553):436–444.

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

Lu, W., J. Li, Y. Li, A. Sun, and J. Wang. 2020. A CNN-LSTM-based model to forecast stock prices. Complexity 2020.

Mahjoobi, J. and A. Etemad-Shahidi. 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. Appl. Ocean Res. 30(3):172–177.

Marier, P., S. Bolduc, M. B. Ali, and J. Gaudreault (Eds.). 2014. S&OP network model for commodity lumber products. In: Proceedings of the 10th International Conference on Modeling, Optimization, and Simulation (MOSIM), November 5–7, 2014, Nancy, France; CIRRELT, Québec, Canada.

Mehrotra, S. N. and D. R. Carter. 2017. Forecasting performance of lumber futures prices. Econ. Res. Int. 2017:1–8.

Miotto, R., F. Wang, S. Wang, X. Jiang, and J. T. Dudley. 2018. Deep learning for healthcare: Review, opportunities and challenges. Briefings Bioinformatics 19(6):1236–1246.

Mir, M., H. D. Kabir, F. Nasirzadeh, and A. Khosravi. 2021. Neural network-based interval forecasting of construction material prices. J. Building Eng. 39:102288.

Miyamoto, B. T., K. Cheung, M. Clauson, and A. Sinha. 2018. Revisiting the compression parallel to grain design values of Douglas-fir. Forest Prod. J. 68(2):132–137.

Moghar, A. and M. Hamiche. 2020. Stock market prediction using LSTM recurrent neural network. Procedia Computer Sci. 170:1168–1173.

Muthukrishnan, R. and R. Rohini. 2016. LASSO: A feature selection technique in predictive modeling for machine learning. In: 2016 IEEE International Conference on Advances in Computer Applications (ICACA), October 24-24, 2016, Coimbatore, India; IEEE Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 18–20.

Nguyen, Q. H., H.-B. Ly, L. S. Ho, N. Al-Ansari, H. van Le, Q. van Tran, I. Prakash, and B. T. Pham. 2021. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Math. Problems Eng. 2021.

Noble, W. S. 2006. What is a support vector machine? Nature Biotechnol. 24(12):1565–1567.

Oliveira, R. A., J. Buongiorno, and A. M. Kmiotek. 1977. Time series forecasting models of lumber cash, futures, and basis prices. Forest Sci. 23(2):268–280.

Peng, L., L. Wang, X.-Y. Ai, and Y.-R. Zeng. 2021. Forecasting tourist arrivals via random forest and long short-term memory. Cognitive Computation 13(1):125–138.

Peng, Z., Q. Huang, and Y. Han. 2019. Model research on forecast of second-hand house price in Chengdu based on XGboost algorithm. In: 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT), October 18-20, 2019, Jinan, China; IEEE Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 168–172.

Pyo, S., J. Lee, M. Cha, and H. Jang. 2017. Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets. PloS One 12(11):e0188107.

Roelofs, R., S. Fridovich-Keil, J. Miller, V. Shankar, M. Hardt, B. Recht, and L. Schmidt. 2019. A meta-analysis of overfitting in machine learning. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS), December 8 -14, 2019, Vancouver, Canada; Curran Associates, Inc., Red Hook, New York. pp. 9179–9189.

Sahu, K. K., and R. Kumar, R. 2020. Current perspective on pandemic of COVID-19 in the United States. J. Family Med. Primary Care 9(4):1784.

Samadi, M., E. Jabbari, and H. M. Azamathulla. 2014. Assessment of M5′ model tree and classification and regression trees for prediction of scour depth below free overfall spillways. Neural Comput. Appl. 24(2):357–366.

Schmidt, J., M. R. G. Marques, S. Botti, and M. A. L. Marques. 2019. Recent advances and applications of machine learning in solid-state materials science. NPJ Comput. Mater. 5(1):1–36.

Selvin, S., R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman. 2017. Stock price prediction using LSTM, RNN and CNN-sliding window model. In: 2017 International

Conference on Advances in Computing, Communications and Informatics (ICACCI), September 13-16, 2017, Udupi, India; IEEE Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 1643–1647.

Shmulsky, R., F. J. N. França, J. T. Ratcliff, B. Farber, C. A. Senalik, R. J. Ross, and R. D. Seale. 2021. Compression properties of small clear southern yellow pine specimens tested across five decades. Forest Prod. J. 71 (3):240–245.

Shobana, G. and K. Umamaheswari. 2021. Forecasting by machine learning techniques and econometrics: A review. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT), April 2-4, 2021, Pune, India; IEEE Computer Society's Conference Publishing Services (CPS), Piscataway, New Jersey. pp. 1010–1016.

Shrestha, A. and A. Mahmood. 2019. Review of deep learning algorithms and architectures. IEEE Access 7:53040–53065.

Singh, S. N. and A. Mohapatra. 2021. Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted SVM model. Appl. Soft Computing 107730.

Song, N. 2006. Structural and forecasting softwood lumber models with a time series approach. Louisiana State University and Agricultural & Mechanical College.

Su, H., E. Zio, J. Zhang, M. Xu, X. Li, and Z. Zhang. 2019a. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. Energy 178:585–597.

Su, M., Z. Zhang, Y. Zhu, D. Zha, and W. Wen. 2019b. Data driven natural gas spot price prediction models using machine learning methods. Energies 12(9):1680.

Tai, A. M. Y., A. Albuquerque, N. E. Carmona, M. Subramanieapillai, D. S. Cha, M. Sheko, Y. Lee, R. Mansur, and R. S. McIntyre. 2019. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. Artif. Intelligence Med. 99:101704.

Tao, H., A. O. Al-Sulttani, A. M. Salih Ameen, Z. H. Ali, N. Al-Ansari, S. Q. Salih, and R. R. Mostafa. 2020. Training and testing data division influence on hybrid machine learning model process: Application of river flow forecasting. Complexity 2020:8844367.

Verly Lopes, D. J., G. d. S. Bobadilha, and A. Peres Vieira Bedette. 2021. Analysis of lumber prices time series using long short-term memory artificial neural networks. Forests 12(4):428.

Wang, W., C. Men, and W. Lu. 2008. Online prediction model based on support vector machine. Neurocomputing 71(4–6):550–558.

Wuest, T., D. Weimer, C. Irgens, and K.-D. Thoben. 2016. Machine learning in manufacturing: advantages, challenges, and applications. Prod. Manufacturing Res. 4(1):23–45.

Yoon, J. 2021. Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. Comput. Econ. 57(1):247–265.

Zhang, D., Sun, C., 2001. US-Canada trade disputes and softwood lumber price volatility. Forest Products Journal 51 (4), 21–27

Zhao, X., B. Yu, Y. Liu, Z. Chen, Q. Li, C. Wang, and J. Wu. 2019. Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. Remote Sensing 11(4):375.

Zhou, X., X. Zhu, Z. Dong, and W. Guo. 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J. 4(3):212–219.

Chapter 6. Conclusion and future work

6.1 Conclusion

This dissertation aimed to contribute in the field of the forest management, the logging industry and the lumber futures through several perspectives. Firstly, it investigated the direction and magnitude of critical socioeconomic factors affecting the transformation of the households in China. Secondly, a quantitative analysis of the US logging industry data, presented employment and profitability trends, and explored contemporaneous causal relationships between employment and different economic variables were provided. Finally, the novel approach of utilizing the Google Trends Index to predict the lumber futures price using Machine Learning and Deep Learning Models was put forth.

Chapter 2 concluded that the age and education of the household heads, income, the holding areas of cropland, ecological forest, forestland, leasing forestland, and legally contracted forestland and their located provinces were found to be statistically significant in transforming the household's forest management. The factors that drive the transformation to various ownership types showed some variations as well.

Likewise chapter 3 concluded the logging industry has been experiencing reduced employment and aging workforce in the past two decades. This might be due to increased productivity from the technological advancement of mechanization, and reduced demand for logging. This chapter also identified that the reduced demand and increased operating costs led to poor profitability and a wave of closures of logging firms, which accelerated the management adjustment in the logging industry.

Chapter 4 concluded that logging production level affects employment directly and indirectly. An increase (decrease) in the level of logging production directly increases (decreases) wages, followed by an increase (decrease) in employment. Employment in the logging industry is most prominently explained by the production level (highest 52.0% at horizon 1-year), followed by the wage (highest 42.0% at horizon 20-year). In contrast, capital and product price have a limited influence on employment.

Lastly, Chapter 5 presented the idea for the first time that despite the high predictive power of Machine Learning and Deep Learning Models, on average, Deep Learning Models can better capture trends and provide more accurate predictions than Machine Learning Models. The Artificial Neural Network model is the most competitive, followed by the Recurrent Neural Network model. The Google Trends Index, which reflects the dynamic changes of the interest and attention from the public, can provide enough information to be good predictors in nowcasting lumber futures prices.

6.2 Future work

Rural households are gradually transforming around the world, so too is the forest management. China's households may have made substantial transformations from rural households in a relatively short period, however they are yet to reach the point similar to the post-industrial societies where forest management is aiming for more amenities and less timber production. We only identified FC, FFF and FSH as the transformed households, and may have omitted many other probable households who have carried out transformation. A more complete picture of the household transformation will help the policymakers plan for more comprehensive and future oriented policies.

Since the early 1970s, the employment in the logging industry in the US has been steadily declining, while the production level has increased significantly, mainly due to technological advancements. The logging industry is an important part of the timber supply chain and has an important impact on the sustainable forest management. Therefore, logging firms with high production efficiency will determine the future of forestry in the US. Future research is required to measure the capital productivity and total factor productivity of the logging firms at the industry-level and study the influencing factors as well.

The prosperity of the logging industry is highly dependent on the economic conditions with an impact on both demand and price. With the outbreak of COVID-19, the US economy fell into a recession again. Subsequently, the federal government launched multiple rounds of economic stimulus policies, which not only stimulated the economic recovery, but also promoted the prosperity of the real estate market. Future research can focus on the impact of COVID-19 as a natural experiment to study the consequences of the economic cycle on the logging industry, or the impact of the real estate market on the logging industry.

Machine Learning and Deep Learning Models provide accurate results for nowcasting lumber futures price. Further research might explore to apply more elaborate models, for example hybrid model, and more accurate training strategies to nowcast or forecast the lumber futures price many days ahead. The further research can also compare the performance between Machine Learning, Deep Learning Models and traditional econometric models.