

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THREE ESSAYS IN APPLIED ECONOMICS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Mohammad Javad Mehregan

2021

To: Dean John L. Volakis
School of International and Public Affairs

This dissertation, written by Mohammad Javad Mehregan, and entitled Three Essays in Applied Economics, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Cem Karayalcin

Jesse Bull

Mohammad Hadi Amini

Hakan Yilmazkuday, Major Professor

Date of Defense: July 1, 2021

The dissertation of Mohammad Javad Mehregan is approved.

Dean John L. Volakis
School of International and Public Affairs

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2021

© Copyright 2021 by Mohammad Javad Mehregan

All rights reserved.

DEDICATION

This thesis work is dedicated to my beloved wife, Bitu, who has been a constant source of support and encouragement during my Ph.D. program and life challenges. Thank you for being my inspiration, editor, proofreader, and sounding board. But most of all, thank you for being my best friend. I owe you everything. This work is also dedicated to my parents and sister, who have always loved me unconditionally and whose good examples have taught me to work hard for what I aspire to achieve. I am genuinely thankful for having you in my life. Without you, I would not be the person I am today.

ACKNOWLEDGMENTS

I would like to sincerely thank my supervisor, professor Yilmazkuday, for his guidance and support through this study. I would also like to thank professor Karayalcin, professor Bull, and professor Amini for their guidance and support.

I should also like to thank fellow graduate student Ehsan Asghari Ghara and all the graduate students at the Department of Economics. I would also like to thank my great friends Amin Shoja and Seyed Javad Nosratabadi for sharing their knowledge and experience with me.

ABSTRACT OF THE DISSERTATION
THREE ESSAYS IN APPLIED ECONOMICS

by

Mohammad Javad Mehregan

Florida International University, 2021

Miami, Florida

Professor Hakan Yilmazkuday, Major Professor

Governments worldwide support their national economies to obtain growth, adequate employment, and price durability. Regulation is a unique approach through which governments control the economy. Governments attempt to maintain and regulate the economy in various ways to guarantee that business fosters the common good. The range of government regulations is enormous and touches all areas of the economy and all features of daily life. Understanding the effect of regulations on the economy is essential since its outcomes can improve government interventions' efficiency. Applied economics can help measure the effectiveness of government intervention on economic outcomes.

This dissertation includes three essays in applied economics and tries to employ economical methods to understand the outcomes of selected government regulations on the economy and environment. The first essay estimates the effect of Iran's subsidy reduction on industrial workshops' total factor productivity (TFP). After the model was applied, results show that Iran's subsidy reduction on energy and food caused TFP to decrease overall for all industrial workshops. For some industries, however, TFP increased.

The second essay tries to determine whether there is a connection between weather conditions and the spread of the COVID-19 disease. Finding this connection could help governments make more informed policy decisions and better prepare for the next waves and subsequent pandemics. The outcomes indicate a negative relationship between the number of infected cases and daily minimum temperature in South Korea. Meanwhile, an increase in air pressure, humidity, and daily minimum wind speed was associated with a higher number of infections. This study focused on the COVID-19 pandemic from the perspective of weather conditions, but other essential factors, like mobility, can also affect the COVID-19 pandemic, and can be as critical as weather conditions.

In the last study, this dissertation evaluates how the 2017 Tax Cuts and Jobs Act (TCJA) affected small businesses' default rate on loans. It focuses on small business loans since activity in this segment is a primary goal of policymakers and shedding light on this area would help improve their policies regarding small businesses. The results show that the TCJA had a positive effect on avoiding defaults by small businesses. This kind of research could also prove helpful for policymakers since with it they can identify the results of their regulations.

TABLE OF CONTENTS

CHAPTER	PAGE
CHAPTER 1	1
1.1. Introduction.....	1
1.2. References.....	5
CHAPTER 2	6
The Effects of Energy Subsidy Reduction on Total Factor Productivity	6
2.1. Introduction.....	6
2.2. literature review	8
2.3. Iran's targeted subsidy scheme.....	13
2.4. The data and descriptive statistics.....	15
2.4.1. The data sources.....	15
2.4.2. The sample summary statistics	15
2.5. The empirical method and results	20
2.5.1. Estimating productivity response to the energy subsidy reduction:	20
2.5.1.1. Total factor productivity and production function estimation	22
2.5.1.2. Subsidy reform plan and Total factor productivity	27
2.5.2. Robustness checks:	29
2.5.2.1 Robustness checks based on provinces average temperatures.....	29
2.5.2.1 Robustness checks based on industrial energy consumption.....	33
2.6. Discussion of Results.....	37
2.7. Conclusions.....	38
2.8. Reference:	41
CHAPTER 3	45
Weather-based COVID-19 Spread Pattern and Public Policy Response.....	45
3.1. Introduction:.....	45
3.2. Literature review:.....	48
3.3. Data set description:.....	53
3.4. Methods:	58
3.4.1. Lag variables:.....	58
3.4.2. Variable selection:	60
3.5. Estimation Models:	60
3.5.1. Model Selection:	60
3.5.2. Regression analysis:.....	62
3.6. Results:.....	63
3.7. Robustness check:	66
3.8. Prediction:	68
3.9. Conclusion:	70
3.10. References:.....	73
CHAPTER 4	77
Small Businesses' Loan Default Rate Under 2017 Tax Regulation.....	77

4.1. Introduction.....	77
4.2.Literature Review.....	80
4.3. The data and descriptive statistics.....	85
4.3.1. The data sources.....	85
4.3.2. The sample summary statistics	87
4.3.3. Variable selection.....	91
4.4. Empirical Method and Results.....	94
4.4.1. Logistic Regression Analysis.....	95
4.4.2. Robustness checks:	99
4.4.2.1. Applied the model 5 with new variables:	100
4.4.2.2. Applied the Model 5 With New Datasets:	105
4.5. Discussion of Results:.....	107
4.6. Conclusions.....	108
4.7. Reference:	111
VITA	115

LIST OF TABLES

TABLE	PAGE
Table 2-1:energy price change after subsidy reduction in 2010	14
Table 2-2: Number of industries based on international standard industry system (ISIC)17	17
Table 2-3:Number of industrial workshops	19
Table 2-4:Energy consumption	21
Table 2-5:Production function components summary statistic	22
Table 2-6: Estimated total factor productivity for every industry from 2007 to 2012	25
Table 2-7: Estimated total factor productivity for every province from 2007 to 2012.....	26
Table 2-8:Regression results	28
Table 2-9: Division based on the annual air temperature of the provinces	31
Table 2-10:Regression results based on temperatures categories	32
Table 2-11:ISIC classifications based on energy using	34
Table 2-12:Regression results based on energy usage categories	36
Table 3- 1:Descriptive statistics of newly confirmed cases.....	57
Table 3- 2:Results of AIC and BIC test	59
Table 3- 3:Variable selection based on covariance between the variables	60
Table 3- 4: The main regression results	65
Table 3- 5:four different quantiles regression results based on population density	67
Table 3- 6:Regression results for the robustness check	68
Table 3- 7:Gyeongsangbuk-do predicted number of COVID-19 positive cases	69
Table 4- 1:Loan dataset detailing:.....	86
Table 4- 2:selected variables:.....	92

Table 4- 3:data description:	93
Table 4- 4:Regression results:	97
Table 4- 5:Results for robustness model 1.	101
Table 4- 6:Results for robustness model 2.	104
Table 4- 7:Results for robustness model 3.	106

LIST OF FIGURES

FIGURE	PAGE
FIGURE 2. 1:Industrial workshops in terms of major activities	16
FIGURE 2. 2: Number of observations	16
FIGURE 2. 3: Number of firms	18
FIGURE 2. 4: The whole industry TFP	23
FIGURE 2. 5: Iran weather map	30
FIGURE 2. 6: Percentage of energy consumption.....	35
FIGURE 3. 1: Place and the ways of infection in South Korea.....	49
FIGURE 3. 2: Test results during the timeline of this study.....	54
FIGURE 3. 3: Age-specific number of patients.....	55
FIGURE 3. 4: The 9 th highest number of infections case by regions in South Korea.....	56
FIGURE 3. 5: Number of infections for every province (time series plot).....	57
FIGURE 3. 6: Distribution of the dependent variable.....	61
FIGURE 3. 7: population density for all cities in South Korea.....	66
FIGURE 3. 8: Prediction based on the results.....	70
FIGURE 4. 1: Loan status from 2010 to 2019.....	85
FIGURE 4. 2: Loan status based on loan grades.....	87
FIGURE 4. 3: Loan status based on employment length.....	88
FIGURE 4. 4: Loan status based on account maturity.....	89
FIGURE 4. 5: Loan status based on home ownership status.....	89
FIGURE 4. 6: Loan status based on applicants' locations.....	90
FIGURE 4. 7: Ratio of loan default per status.....	91
FIGURE 4. 8: Selected variables correlation.....	94
FIGURE 4. 9: Number of borrowers based on loan grades.....	99
FIGURE 4. 10: Correlation matrix.....	100
FIGURE 4. 11: Relationship three correlated variables.....	102
FIGURE 4. 12: Relationships among three highly correlated variables.....	103
FIGURE 4. 13: Number of applicant's location around the country.....	105

ABBREVIATIONS AND ACRONYMS

TFP	Total Factor Productivity
GAM	Generalized Additive Model
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
TCJA	Tax Cuts and Jobs Act
PD	Probability of default Statistical
SCI	Center of Iran World Health
WHO	Organization
KCDC	Korea Centers for Disease Control Prevention

Chapter 1

1.1. Introduction

Government regulations can act as a double-edged sword. Governments try to put some restrictions on technology, capital, and other economic areas to enhance economic outcomes and living standards for their citizens. The results of these regulations are crucial since these outcomes can help governments make better decisions.

Applied economics is a technique that can be used in understanding the consequences of these regulations on the economy, environment, welfare, or other areas. Applied economics is also, a valuable tool for public policymakers. Many economists have used it to predict macroeconomic and microeconomic consequences of various policy proposals or evaluate effects of existing policies. Macroeconomic modeling is commonly used to predict changes in unemployment, economic growth, and inflation at the national, regional, and state levels (Sowell, 2008).

All three essays in this dissertation concern applied economics topics. The first estimates the effect of Iran's subsidy reduction on industrial workshops' total factor productivity (TFP). The second attempts to identify a connection between weather conditions and the spread of COVID-19 in South Korea. The final essay evaluates how the TCJA affected U.S. small businesses' loan default rate.

This dissertation endeavors to provide a new perspective on the applied economics literature on economic policymaking and its welfare results. Energy subsidies have a long history in Iran. Economists believe that energy subsidies are only the result of populist and populist policies that, although initially satisfying to the general public, ultimately do

nothing but weaken the economy and increase social inequality and public dissatisfaction with government performance (Durand-Lasserve et al., 2015). In this regard, in 2010 Iran's government decided to implement a subsidy reduction plan, and it had a profound effect on the economy. Indeed, affecting industrial productivity was the primary goal of this project.

Therefore, the first study of this dissertation aims to examine whether the subsidy reduction made industrial workshops more efficient. It also tries to understand how TFP shifts after subsidy reform is applied among different industry sectors. First, then, it is necessary to estimate TFP to understand the effect of subsidy cuts on this variable. TFP describes how inputs are applied in the process of production; they could change efficiency. TFP can also be considered an indicator of the efficiency of the whole economy since technological expansion and productivity are considered its main sub-sections (Zofio, 2007).

In a second step, the estimated TFP was used in the main regression to explore the results of the subsidy reduction plan on TFP. The conclusions show that the subsidy reform plan could not help industrial workshops become more productive as much as governments expected. Although some industries had different outcomes, on the whole, TFP decreased for the whole economy.

The statistical population of this project includes all industrial workshops located in the urban and rural areas of the country with an average of 10 to 49 employees between 2007 and 2012. The Statistical Center of Iran provided the industrial workshop data. In addition, some critical economic data, such as the consumer price index (CPI), was collected from Iran's central bank's website.

When the first positive case of COVID-19 was announced in South Korea on January 20, 2020, the government began to monitor the pandemic situation in an attempt to attempt to control it. A crucial question during the COVID-19 pandemic was whether there was a connection between weather conditions and the spread of the COVID-19 virus. Discovering this relationship could help governments make more informed policy decisions and become more prepared for the next waves and subsequent pandemics.

Therefore, in its second essay, this dissertation tries to study how different weather conditions might have affected the number of confirmed COVID-19 cases in South Korea. Another question is how effective government regulation was during this pandemic.

This study, then, intends to connect two different literature angles by applying climate indicators and government regulation in the same regression as right-hand-side variables.

The data set includes eleven thousand observations, including all the confirmed coronavirus positive cases. Sub-national level data and spatial variation in weather within the country were used. We collected daily COVID-19 confirmed case data from the official website of Korea Centers for Disease Control Prevention (KCDC) and daily weather data from NASA's website. City-level panel data analysis was used for this research. The data covered the information from nine provinces as well as 95 cities in South Korea.

According to the World Health Organization (WHO), COVID-19 symptoms appear on average 5–6 days after infection (WHO, 2019). To accommodate the disease onset and detection lag, we used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the appropriate number of lags to be included in the model. Based on the AIC and BIC, seven-day lags were selected as independent variables.

The Generalized Additive Model (GAM) was applied for this study, first because it offers an excellent fit for nonlinear and linear relationships since it is so adaptable. Second, generalized additive models can work with many distributions, such as gamma, normal, and Poisson. Lastly, it is prevalent in epidemiology, air quality, and medicine (Dominici et al., 2002).

The results indicate a negative connection between minimum daily temperature and the daily number of COVID-19 confirmed cases. According to the model, during winter, the number of infections would speed up again. Moreover, minimum wind speed and atmospheric pressure have a negative relationship with the dependent variable.

Governments intervene in the financial market in a variety of ways. To improve the effectiveness of government intervention, many researchers have conducted studies (Wang & Xu, 2011). Accordingly, this study tries to understand the effect of the 2017 Tax Cuts and Jobs Act (TCJA) on one aspect of the financial market by using the Lending Club data set. The last essay of this dissertation considers the effects of the TCJA on the probability of default (PD) from 2010 to 2019 using the logistic regression method.

The Lending Club data include information on 24,576 personal loan candidates with 36- or 72-month loans. This data set is public and includes various information about applicants. Only the loans with clear ending commitments were selected for this study. Therefore, the loans must have been either repaid in full or charged off to be included in our analysis. In a nutshell, this information contains borrowers' profiles and loan features.

Since the dependent variable is a dummy variable, it equals one if the borrowers fully pay off the loan; otherwise, it equals zero. The logistic regression method was applied

in this study. Zero means the applicants could not make the full payment by the due date, and therefore the loan was considered charged off.

The results of this study show that the TCJA helped small businesses avoid defaulting on their loans. All the selected variables show a statistically significant outcome. It can be concluded that all are essential features for lenders to consider when granting loans.

The following chapters introduce these essays.

1.2. References

- Durand-Lasserve, O., Campagnolo, L., Chateau, J., & Dellink, R. (2015). Modelling of distributional impacts of energy subsidy reforms: an illustration with Indonesia.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 395(10223), 497-506.
- Lu, H., Stratton, C. W., & Tang, Y. W. (2020). Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. *Journal of medical virology*, 92(4), 401-402.
- Sowell, T. (2008). *Applied economics: Thinking beyond stage one*. Hachette UK.
- Wang, C., & Xu, H. (2011). Government intervention in investment by Chinese listed companies that have diversified into tourism. *Tourism Management*, 32(6), 1371-1380.
- Zofio, J. L. (2007). Malmquist productivity index decompositions: a unifying framework. *Applied economics*, 39(18), 2371-2387.

Chapter 2

The Effects of Energy Subsidy Reduction on Total Factor Productivity

2.1. Introduction

In December 2010, Iran, implemented a subsidy reform plan for energy and food. Improving economic efficiency, preventing an uncontrolled increase in fuel consumption, and decreasing inequality in Iran were some of the subsidy reduction program's goals (Salehi-Isfahani, 2016). Subsidy reforms or cuts have been applied in multiple countries, with mixed results. The consequences of subsidy reforms are crucial for policymakers because they can use the results to boost the economy, enhance social regulations, and modify their environmental laws (Bernini & Pellegrini, 2011)

In this regard, this study examines the effects of Iran's 2010 subsidy reductions on Iranian industrial workshops' TFP. Usually, growth in technological innovation and efficiency are measured by TFP, which is regularly recognized as the primary proxy for the GDP growth rate (Harris, 2019). TFP also is a scale used to measure production efficiency; it measures how much output can be manufactured from a specific number of inputs (Mattsson, 2019).

Consequently, by using Iran's industrial workshops data set between 2007 to 2011 this paper aims to address the following questions:

1. Will industries become more productive and efficient when there is a reduction in the subsidy?
2. How will TFP change among different industries after applying subsidy reform?

First, we try to estimate TFP using production functions to answer the research question. Various techniques can be used to estimate TFP, but this study applies the parametric estimation method and production function to measure TFP. The endogeneity concern about this method was addressed by setting fixed effects as a variable in the regression (Dong & Li, 2014). The production function includes five components: output, value of capital, number of paid labors, material value, and total energy. For this model, total energy was calculated as the sum of the value of oil products, natural gas, coal, and electricity. Second, to answer the research questions, estimated TFP was used in the main regression to examine the effects of the subsidy reduction plan on TFP.

There is no agreement on the effect of subsidy reduction on firms' TFP since the results would vary by economic situation. Some previous studies have implied that subsidies may positively affect production, but at the same time, they could negatively affect productivity (Rizov, 2013; Ciaian & Swinnen, 2009). Iran's subsidy reduction also attracted critics; for example, some economists believe that targeting subsidies in the short term increased the price of energy carriers and reduced consumption, but financing was done incorrectly. In addition, a year after the subsidy reduction was implemented, international sanctions were imposed against the Central Bank of the Islamic Republic of Iran and sanctions had a highly negative effect on the economy.

This study's outcomes show that productivity decreased after the subsidy reduction law was implemented. Although some industries had different results, overall, TFP was reduced. There are many reasons for this outcome, but the sanctions were a primary one. The embargo made the transfer of money for buying essential items more complex; therefore, firms were unable to import energy-efficient equipment (Farzin et al., 2011).

Other outcomes of this study show that more investment in research and development can increase TFP. Previous studies indicated that investment in research and development caused more innovation, leading to higher productivity for firms (Griliches, 1998).

Several studies have been performed on subsidy and TFP, but this study is different since it used Iran as a case study. Iran is unique as a main oil-exporting country with slow long-run growth. Additionally, the data set provides a wide variety of information, such as the number of highly educated workers, the investment in research and development, and the number of skilled and unskilled workers. This rich data set helps us get more of the critical outcomes than the existing literature has.

Moreover, another unique contribution of this study is its use of both the subsidy reduction law and international sanctions that coincided during the timeline of this study. Thus, it connects two bodies of literature by using Iranian industrial workshops as a data set. Furthermore, few previous studies have used fixed effects to control endogeneity (e.g., Cameron et al., 2005; Alcalá & Ciccone, 2004; Aldieri & Cincera, 2009), but three fixed effects in this study were applied to get more critical results and control endogeneity.

The rest of the paper is organized as follows: Chapter 2 is a literature review, Chapter 3 explains the data-collection, data set features, descriptive statistics, and variable-selection methods. Chapter 4 reviews the empirical approach and outcomes. Chapter 5 discusses the results, and, finally, Chapter 6 draws conclusions from this study.

2.2. literature review

Governments usually provide subsidies as a particular amount of money to support industries or businesses aiming to make their goods or services affordable and competitive in the marketplace (Hall & Deardorff, 2006). Globally, subsidies are implemented in

multiple sectors, and energy is the prominent in this ranking. For example, in 2009, governments contributed around 409 billion dollars in subsidies to fossil fuel (International Energy Agency (IEA), 2012). Most of the granted subsidy (around 70%) was provided by developing countries, with North African and Middle Eastern governments providing the highest amount of energy subsidy globally. Among these countries, Iran ranks first in the amount of its subsidy. In 2009, Iran's government provided around 82 billion dollars' worth of subsidy to the fossil-fuel sector, an amount equal to 20% of Iran's GDP that year (IEA, 2009).

Although subsidy plays a notable role in the economy, such as keeping prices low enough to make tradable goods affordable to export, many studies suggest that it has adverse outcomes both globally and locally (see, e.g., Burniaux et al., 2009; Sobhani, 2001). In contrast to all suggested studies about subsidy reduction, some international organizations strongly encourage governments to adopt this policy. Some even provide plans to make this step more manageable and efficient for governments. For example, international energy agencies have offered a plan under subsidy reform to gradually reduce or eliminate subsidies. This plan offers many short and long-term economic and environmental benefits (IEA, 2009).

Subsidy reduction is not new, and many countries have experienced subsidy cuts at least once. For example, between 2015 and 2017, more than 40 countries, including India, Morocco, and China, implemented energy-subsidy reductions (Zinecker, 2018). Similarly, between 2010 and 2014, most countries in the MENA region implemented a subsidy-reduction plan for the first time. Subsidy cuts have also occurred during challenging moments, such as wars, social protests, or revolutions (Verme & Araar, 2016).

Like other countries in December 2010, Iran passed a subsidy-reduction law under the name “Targeted Subsidy Plan.” The goals of this plan were “to rejuvenate Iran’s economy, increase productivity, give it a new footing and bring it out of the slump it has been in for so long” (Tehran Times, 2011). More detail about this law can be found in the section on Iran’s targeted subsidy scheme.

Despite these suggestions, many governments do not support subsidy reduction plans since they believe it may bring political costs and unknown consequences (World Bank, 2007). A considerable number of studies have estimated the effect of subsidy-reduction policies on various aspects of the economy and environment. For example, in 2001 Myers and Kent examined the importance, purposes, and results of global subsidies on economics and the environment in five key areas: agriculture, fossil fuels, road transportation, water, and fisheries (Myers & Kent, 2001).

In 2001, De Moor examined the following issues in the context of both developing and developed countries:

- Differences in the amount of energy subsidy provided by governments in developed and developing countries, as well as a comparison of the distribution efficiency of subsidies among lower, middle, and higher household income levels.
- Effects of this policy on the economy, environment, and inequality.
- Effective solutions to implement subsidy reform more efficiently, especially in the energy sector (De Moor, 2001).

Some studies have tried to understand the effect of subsidies on research and development (R&D) in different countries. For example, in 2016, Bronzini and Piselli tried

to estimate the effects of subsidies on R&D and innovation in northern Italy in the case of small firms and startups using the regression discontinuity method. The results suggested that the policy could lead smart firms to apply for a grant. They also found a positive relationship between subsidies on R&D and the number of grant requests, particularly for smaller firms (Bronzini & Piselli, 2016).

In 2011 Boqiang and Zhujun attempted to examine the effect of energy subsidies on the economy in China using the price gap approach. Their findings show that the energy subsidy equaled nearly \$356.73 billion, or 1.43% of China's GDP in 2010. Their results also indicated that most of the government subsidies were allocated to the oil-products, electricity, and coal sectors. Additionally, applying the computable general equilibrium (CGE) model helped them estimate the consequences of energy subsidy reform in China. Their outcomes showed that subsidy reductions would decrease emissions and demand for energy but harm macroeconomics variables in the short run. In 2012, Lin et al. tried to predict the possible consequences of subsidy removal in various scenarios and suggested ways of decreasing the adverse effects of subsidy reform in China. They also attempted to estimate how subsidy reduction could affect global trade with China. They divided the world into five trade regions: China, OECD¹, Brazil, India, and the rest of the world and applied a multi-world-region general equilibrium model. The results showed that subsidy

¹ The Organization for Economic Cooperation and Development (OECD) included 37 countries and was organized in 1961 as a government conference.

elimination would affect trade competition and the economy in all regions. Additionally, if China implemented a subsidy-reduction policy, it would harm China's trade with the regions that did not have this policy in place (Lin et al., 2012).

In 2013, Plante tried to estimate how these subsidies would affect the steady-state levels of such macroeconomic aggregates as consumption, labor supply, and aggregate welfare by employing a small open-economy model with a non-traded sector. Their outcomes showed that aggregate welfare would be reduced by subsidies across steady states. He noted that aggregate welfare is significantly higher if a lump sum of equal value replaces the subsidies (Plante, 2013).

Isfahani and Mostafavi-Dehzooei used the panel data of Iranian household expenditures and income surveys for 2010 and 2011 to explain the cash-transfer program's effect on the labor supply. After the subsidy reduction, the government began to deposit cash to 70 million individual accounts in 2011. The findings showed the cash-transfer program did not negatively affect the labor supply but that it positively influenced self-employed women and men (Isfahani & Mostafavi-Dehzooei, 2018).

Another study by Farzin (head of the Subsidy Reform Headquarters of Iran in 2010) analyzed the economic and technical difficulties associated with the plan, including the cash transfers to households. It also examined the reform from a historical viewpoint, especially for the final stages of the preparation (Farzin et al., 2011).

As mentioned, many studies have examined various aspects of subsidy reduction on the economy and environment. Still, few have concentrated on how productivity can influence this policy (Bernini et al., 2017). Bergström tried to examine the effect of subsidy of public capital on the TFP of Swedish firms using the production function. Employing

the panel data of subsidized and unsubsidized Swedish firms from 1987 to 1993 helped explain how subsidies can facilitate firm growth. Still, it found no significant relationship between subsidy and productivity (Bergström, 2000).

Bernini et al. estimated the effect of subsidies on TFP for selected Italian firms in the long run. For the first stage, TFP was assessed employing a Stochastic Frontier Analysis (SFA). The application of regression discontinuity helped them determine the technological and technical productivity changes in subsidized selected firms. The outcomes indicated that subsidy cuts may decrease TFP for firms in the short run, but in the long run, TFP will increase (Bernini et al., 2017).

2.3. Iran's targeted subsidy scheme

The Law on Targeting Subsidies was introduced as a bill by the Ninth Government of the Islamic Republic of Iran in 2008. After a long delay, it was approved by the Islamic Consultative Assembly with some changes. The bill is part of an economic transformation plan put forward by President Mahmoud Ahmadinejad's government. On December 18, 2010, Ahmadinejad appeared on Iranian state television and officially announced the implementation of the law on targeted subsidies. Beginning the morning of December 19, costs of energy, water, and electricity were all affected, resulting in new prices throughout Iran (Gharibnavaz et al., 2015). Right after the subsidy reduction plan was passed by the government, the price of energy jumped dramatically; Table 2-1 details this increases.

Table 2-1:energy price change after subsidy reduction in 2010

Commodity	Unit of measurement	Before SR	After SR	Change in prices
Crude oil	cents/Liter	0.95	2	111%
Electricity	cents/KWH	1.7	4	135%
Water	cent/m ³	9	25	178%
LPG	cents/Liter	10	30	200%
Petrol (Regular)	cents/Liter	10	40	300%
Jet Fuel Prices	cents/Liter	10	40	300%
Petrol (Premium)	cents/Liter	15	80	433%
Kerosene	cents/Liter	1.5	10	567%
CNG	cent/m ³	4	30	650%
Diesel	cents/Liter	1.65	15	809%
Gas Oil	cents/Liter	1.5	15	900%
Natural Gas	cent/m ³	0.5	8	1500%
Fuel oil	cents/Liter	1	20	1900%

Source: Statistical Centre of Iran (2011)

Note: all the prices in columns 3 and 4 are reposted by dollar.

As a result of subsidy reform, consumption was drastically reduced in the sectors affected by this law in the first year. For example, the annual growth of gasoline consumption was lowered from +5% to -3.5%. Liquid gas consumption decreased by 12% and electricity consumption by 16%. Consequently, the country achieved about \$17.5 billion in foreign-exchange savings by reducing its energy consumption—an outstanding achievement. The price of gasoline sold inside Iran was 16% of the international price before starting the project, but it increased to 47% after implementing the project (Esmacili et al., 2013).

2.4. The data and descriptive statistics

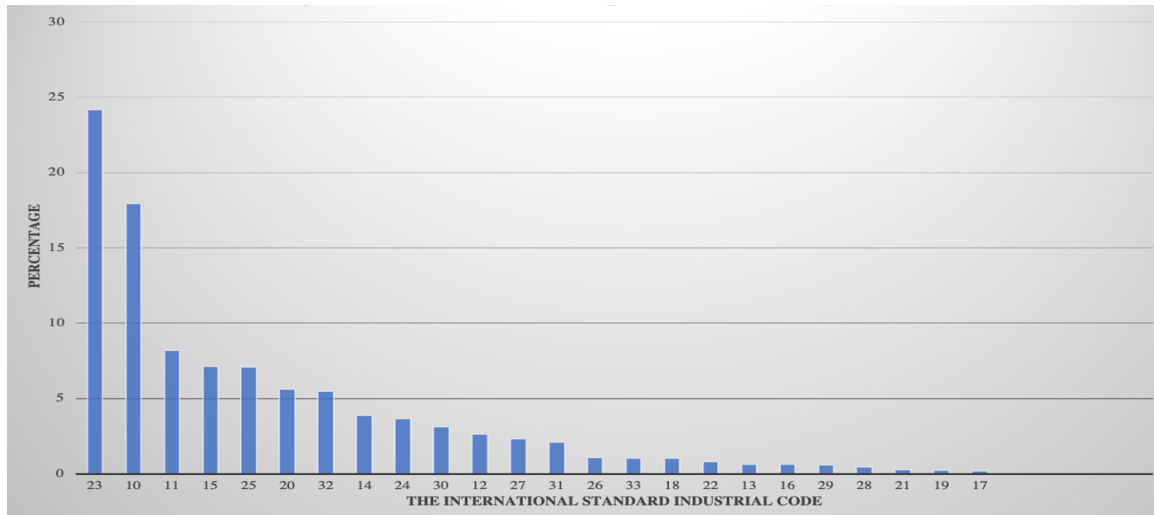
2.4.1. The data sources

The Statistical Centre of Iran collected the data set and uploaded it to its website. According to the data set provider, the main aim of collecting this data was to provide a comprehensive data set of the characteristics of industrial workshops, plan industrial development, adopt economic policies, and evaluate the results of technological development programs and commercial systems (Statistical Centre of Iran, 2011). The statistical population of this project includes all industrial workshops located in urban and rural areas of the country, with 10–49 employees on average between 2007 and 2012.

2.4.2. The sample summary statistics

The balanced panel data set includes 11,945 industrial workshops with 10–49 employees engaged in industrial activities between 2007 and 2012 across the country. These workshops belong to different industries based on their major activities: 24% are under the category “Manufacture of other non-metallic mineral products,” 18% under “Food and beverage industries,” 7% under “Manufacture of fabricated metal products, except machinery and equipment,” and the rest belong to other industries.

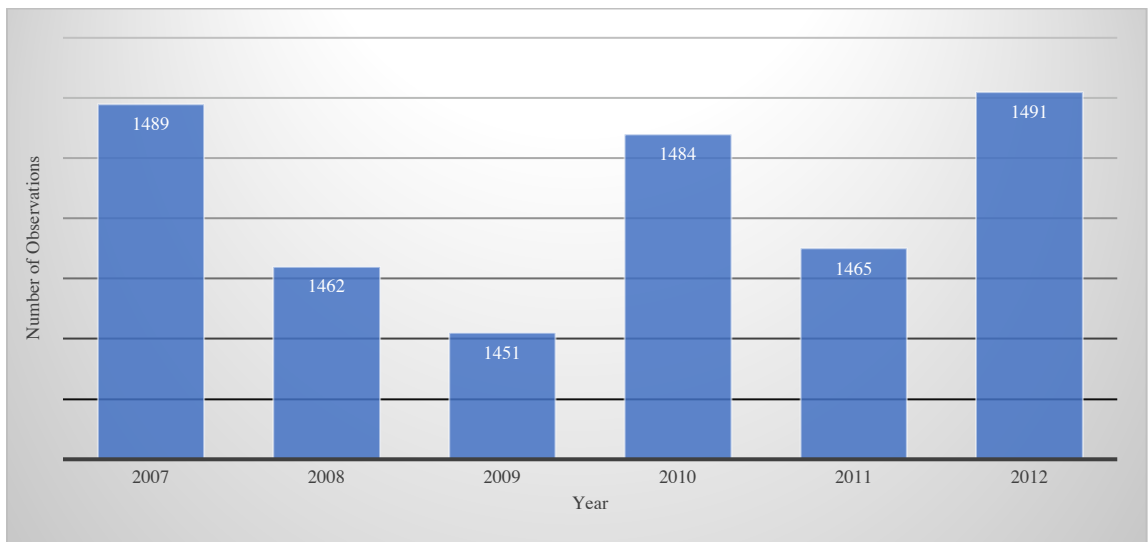
FIGURE 2. 1:Industrial workshops in terms of major activities



Note: This figure shows Percentage of the number of industrial workshops of 10-49 employees in terms of major activity
 Note: X axis indicates The International Standard Industrial Code (ISIC) more detail about ISIC indicate in table 1

Although the data set includes around 12,000 industrial workshops, only 8,842 industrial-level observations with a 4-digit level of ISIC have been made since the firm’s level data was not available.

FIGURE 2. 2: Number of observations



Note: This figure shows the number of observations each year.

Based on the international standard industry system (2-digit ISIC), 23 industries are included in this data set; they are detailed in Table 2-2

Table 2-2: Number of industries based on international standard industry system (ISIC)

ISIC	Industry name	Percent (%)
23	Manufacture of other non-metallic mineral products	24.15
10	Manufacture of food products	17.92
11	Manufacture of beverages	8.17
15	Manufacture of leather and related products	7.11
25	Manufacture of fabricated metal products, except machinery and equipment	7.06
20	Manufacture of chemicals and chemical products	5.6
32	Other manufacturing	5.49
14	Manufacture of wearing apparel	3.88
24	Manufacture of basic metals	3.66
30	Manufacture of other transport equipment	3.12
12	Manufacture of tobacco products	2.61
27	Manufacture of electrical equipment	2.33
31	Manufacture of furniture	2.09
26	Manufacture of computer, electronic and optical products	1.05
33	Repair and installation of machinery and equipment	1.03
18	Printing and reproduction of recorded media	1.02
22	Manufacture of rubber and plastics products	0.79
13	Manufacture of textiles	0.64
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	0.63
29	Manufacture of motor vehicles, trailers, and semi-trailers	0.56
28	Manufacture of machinery and equipment N.E.C.	0.43
21	Manufacture of pharmaceuticals, medicinal chemical, and botanical products	0.26
19	Manufacture of coke and refined petroleum products	0.24
17	Manufacture of paper and paper products	0.16

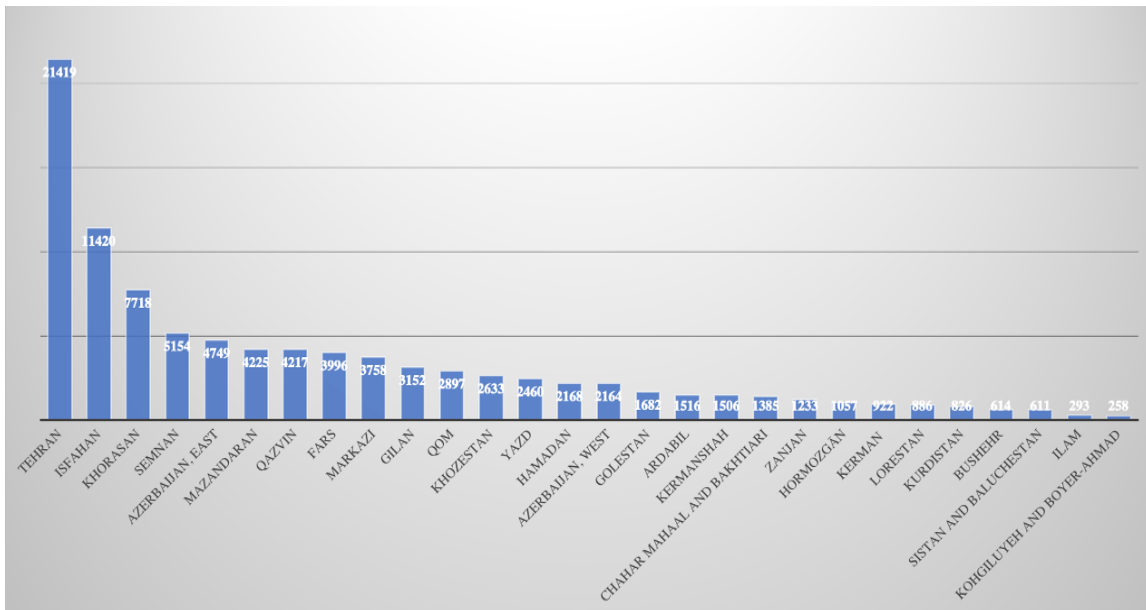
Note: 4_digits ISIC was used in this study. The 2-digit of ISIC code shows the division. 4-digits provides much more details compare to 2-digits ISIC.

The political divisions of the country, from highest to lowest, are province, city, district, and rural district. Each province comprises several cities, each city comprises several districts, and each district comprises several villages. Each village also includes several rural areas, places, and farms.

Province: A province is a part of the country with a definite geographical area, which joins several neighboring cities according to their political, social, cultural, economic, and natural situation.

City: A city is a unit of a province with a particular geographical area joined by several neighboring parts that, in terms of natural, social, economic, political, and cultural factors, have created a proportionate and homogeneous unit (Statistical Centre of Iran, 2010).

FIGURE 2. 3: Number of firms



Note: This figure shows the number of firms for every province

Around 240,000 workers contributed to this survey from all parts of Iran. Among all provinces, Tehran, the capital, has the highest percentage of workers, with 16%, followed by Isfahan, with 11%, and Khorasan, with 7%. Table 2-3 shows the number of

firms in each year for every province.

Table 2-3: Number of industrial workshops

Provinces	2007	2008	2009	2010	2011	2012
Tehran	3971	3964	3817	3376	3232	3059
Isfahan	2010	2002	1906	1876	1832	1794
Khorasan	1394	1391	1318	1273	1180	1162
Semnan	931	939	928	870	752	734
East ,Azerbaijan	861	872	796	733	732	755
Mazandaran	724	737	715	701	686	662
Qazvin	721	714	727	672	660	723
Fars	721	723	675	650	616	611
Markazi	641	672	631	611	540	663
Gilan	541	535	512	476	531	557
Qom	531	534	462	424	498	448
Khozestan	507	506	447	416	398	359
Yazd	431	433	412	394	379	411
Hamadan	369	372	360	357	367	343
West ,Azerbaijan	401	397	357	333	322	354
Golestan	251	256	253	250	351	321
Ardabil	341	343	282	215	175	160
Kermanshah	294	280	254	240	226	212
Chahar Mahaal and Bakhtiari	222	221	221	225	244	252
Zanjan	208	208	190	211	202	214
Hormozgān	181	189	175	200	171	141
KERMAN	149	157	148	145	126	197
Lorestan	119	118	103	218	177	151
Kurdistan	167	161	134	143	113	108
Bushehr	110	112	110	98	94	90
Sistan and Baluchestan	110	109	103	103	88	98
Ilam	61	57	58	51	39	27
Kohgiluyeh and Boyer-Ahmad	47	45	44	53	44	25

There are two ownership types for the selected industrial workshops in the data set: private, and public ownership. Private (non-cooperative) means workshops in which more than 50% of the capital belongs to either a few stockholders, workshop members, or a non-

governmental institution. Public refers to workshops where more than 50% of the capital belongs to ministries, government agencies, banks, municipalities, or other public-sector organizations (Statistical Center of Iran, 2011).

2.5. The empirical method and results

2.5.1. Estimating productivity response to the energy subsidy reduction:

Estimating productivity response to the energy subsidy reduction. TFP explains how inputs are applied in the process of production; they can change efficiency. TFP is sometimes mentioned as “multi-factor productivity,” meaning that it can be considered a pattern of knowledge or technology. The literature on estimating TFP can be categorized in four different groups (Diewert, 1981):

- Parametric estimation,
- Non-parametric indices,
- Exact index numbers, and
- Linear programming methods.

Parametric estimation employs parametric estimates of cost and production function to estimate TFP based on the Cobb–Douglas functions. Non-parametric indices apply non-parametric criteria, such as the Malmquist index (Malmquist, 1953), to estimate TFP. In an exact index numbers approach, exact index numbers are applied to estimate TFP. For example, in 2003, Ahn and Abt estimated price index amounts with the chain-type Fisher formula and applied price index accounts as the deflators without using published price index values. Finally, the last group uses linear programming methods to estimate TFP. In 1968, Aigner and Chu used this method for the first time to estimate how

technological change affected Yugoslavia’s firm’s TFP (Heshmati & Rashidghalam, 2020).

This study applies the parametric estimation method and production function to measure TFP. Two different approaches can be taken to estimating the effects of productivity using the parametric estimation method: calculating based on either production function or cost function. Both methods need to address the endogeneity issue. Therefore, the production function method was selected for this study, and the endogeneity concern was addressed by setting industry*year fixed effects as a variable in the regression (Dong & Li, 2014).

The production function includes five components (output, value of capital, number of paid laborers, material value, and total energy). For this model, total energy was calculated as the sum of the value of oil products, natural gas, coal, and electricity.

Table 2-4:Energy consumption

Year	2007	2008	2009	2010	2011	2012
oil products	65.1	73.2	64.4	57.1	38.8	35.9
Natural gas	140.4	147	158.7	187.3	214.3	221.6
coal	0.3	0.3	0.2	0.1	0.1	0.1
Electricity	30.7	32.2	34.4	37.3	41	42.4
Total Energy consumption	236.5	252.7	257.7	281.8	294.2	300

Source: Iran ministry of energy

Note: This table shows the final energy consumption of the industrial sector by energy carriers (million barrels of crude oil).

As shown in Table 2-4, the annual growth of energy consumption in the industrial sector during the timeline of this research was positive. In other words, energy consumption in this sector increased. It should be noted that, since 2010, when the targeted

subsidy plan was implemented, energy consumption has increased with a slower slope than in previous years.

Table 2-5: Production function components summary statistic

Variable	Mean	Median	Std	Min	Max	Count
Total Energy	46800.73	2734	338216.06	0	8315114	8842
Workers	776.33	248	2207.3	10	54871	8842
Inputs	902316.81	65197	7665078.57	0	203920050	8842
Out puts	1227534.92	111146	9172299.28	0	218036561	8842
Fixed capital	48531.64	2660	464373.19	0	19475739	8842

Since Table 2-5 shows considerable variation in some input components based on their standard deviations, the logarithm method was applied for all inputs to smooth out heterogeneity and reduce the effect of outliers (Wooldridge, 2006).

2.5.1.1. Total factor productivity and production function estimation

The following production function was applied to estimate TFP for industry k , year t , and province i based on Beveren's proposal (Beveren & Ilke, 2007).

All components were divided by each province's CPI to control the inflation rate fluctuation during the timeline of this study.

$$\text{Out}_{itk} = \beta_0 + \beta_1 \text{Capital}_{itk} + \beta_2 \text{Labor}_{itk} + \beta_3 \text{Material}_{itk} + \beta_4 \text{Total Energy}_{itk} + \text{i. Year} + \text{i. ISIC} + \text{i. Provinces} + \varepsilon_{itk} \quad (1)$$

While ε_{itk} estimates the average efficiency level over industry and over time, ε_{itk} can be decomposed into an unobservable and observable element. ε_{itk} is then decomposed into two elements, ω_{itk} and η_{itk}^q , for Regression 2.

$$\text{Out}_{itk} = \beta_0 + \beta_1 \text{Capital}_{itk} + \beta_2 \text{Labor}_{itk} + \beta_3 \text{Material}_{itk} + \beta_4 \text{Total Energy}_{itk} + \text{i. Year} + \text{i. ISIC} + \text{i. Provinces} + \omega_{itk} + \eta_{itk}^q \quad (2)$$

ω_{itk} represents industry-level productivity, and η_{itk}^q is an id component indicating accidental deviations from the average because of estimation error, unforeseen delays, or other external circumstances.

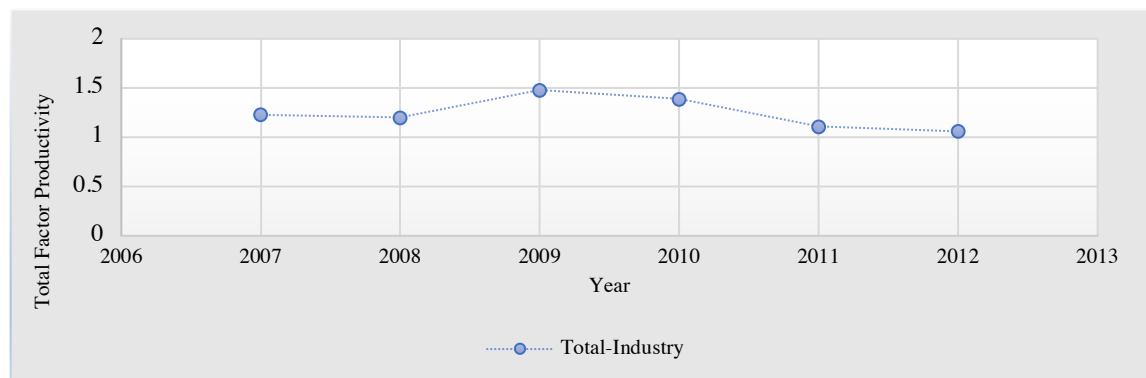
By solving equation 3 for $\hat{\omega}_{itk}$, the TFP is estimated, as shown in Regression 3.

$$\hat{\omega}_{itk} = \text{Out}_{itk} - \beta^{\wedge}_1 \text{Capital}_{itk} - \beta^{\wedge}_2 \text{Labor}_{itk} - \beta^{\wedge}_3 \text{Material}_{itk} - \beta^{\wedge}_4 \text{Total Energy}_{itk}$$

(3)

Figure 5 presents the yearly estimated average TFP for the whole industry during the timeline of this study. The results show that TFP decreased by 0.4 after 2010, the year the subsidy-reform plan was implemented.

FIGURE 2. 4: The whole industry TFP



Note: This figure shows the whole industry TFP from 2007 to 2012.

The result show that the average TFP declined for the whole industry after 2010, although not all industries experienced this outcome.

Based on the results, among all industries the three highest productivity rates are related to:

1. The manufacture of wood and products of wood and cork, except furniture; the manufacture of articles of straw and plaiting materials.

2. The manufacture of coke, refined petroleum products, and nuclear fuel; and
3. The manufacture of medical, precision, and optical instruments, watches, and clocks.

The three lowest rates of TFP are related to:

1. Recycling.
2. The manufacture of other transport equipment; and
3. The manufacture of paper and paper products.

Table 2-6 details this.

Table 2-6: Estimated total factor productivity for every industry from 2007 to 2012

ISIC		2007	2008	2009	2010	2011	2012
15	Manufacture of food products and beverages	1.23	1.2	1.48	1.39	1.11	1.06
16	Manufacture of tobacco products	1.2	1.35	1.53	1.63	1.15	1.47
17	Manufacture of textiles	1.19	1.04	1.06	1.16	1.24	1.39
18	Manufacture of wearing apparel; dressing and dyeing	1.25	1.2	1.35	1.58	2.07	1.18
19	Tanning and dressing of leather; manufacture of luggage	1.11	0.69	0.98	0.92	0.72	1.11
20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	2.14	1.95	1.71	2.36	2.71	3.17
21	Manufacture of paper and paper products	1.22	0.98	0.97	0.75	0.72	1.12
22	Publishing, printing and reproduction of recorded media	0.87	0.79	0.94	0.63	0.44	0.48
23	Manufacture of coke, refined petroleum products and nuclear fuel	1.27	1.27	4.47	2.96	1.92	2.56
24	Manufacture of chemicals and chemical products	1.59	1.21	1.1	1.18	1.17	1.12
25	Manufacture of rubber and plastics products	1.2	1.09	0.89	1.09	1.32	1.25
26	Manufacture of other non-metallic mineral products	1.2	1.16	1.12	1.19	1.16	1.28
27	Manufacture of basic metals	1.47	1.28	1.09	1.12	1.07	0.82
28	Manufacture of fabricated metal products, except machinery and equipment	1.22	1.16	1.1	0.94	1.17	1.01
29	Manufacture of machinery and equipment N.E.C.	1.23	1.22	1.12	1.05	1.06	0.87
30	Manufacture of office, accounting, and computing machinery	1.19	1.43	1.6	2.07	2.13	2.63
31	Manufacture of electrical machinery and apparatus N.E.C.	0.83	0.98	1.1	0.84	0.87	0.63
32	Manufacture of radio, television and communication equipment and apparatus	0.71	1.03	1.16	0.99	1.35	3.91
33	Manufacture of medical, precision, and optical instruments	1.34	1.56	2.43	1.62	2.32	2.21
34	Manufacture of motor vehicles, trailers, and semi-trailers	0.9	1	1.14	1.24	1.26	0.76
35	Manufacture of other transport equipment	1.01	0.68	0.76	0.8	0.8	0.6
36	Manufacture of furniture; manufacturing N.E.C.	0.9	0.95	0.84	0.57	0.68	0.48
37	Recycling	1.14	0.62	0.81	0.35	0.28	0.73
	Total Industry	1.23	1.2	1.48	1.39	1.11	1.06

TFP was also estimated for each of the provinces separately; the results are shown in Table 2-7. Based on these results, the highest TFP growth rates during the timeline of this study took place in Hormozgān, Kurdistan, Kohgiluyeh, and Boyer-Ahmad. By

comparison, the highest TFP decline rates belonged to Qom, Bushehr, and Ilam province.

See Table 2-7 for more details.

Table 2-7: Estimated total factor productivity for every province from 2007 to 2012

Provinces	2007	2008	2009	2010	2011	2012
Tehran	0.87	0.96	0.97	1.10	0.92	1.52
Isfahan	1.34	1.14	1.54	1.42	1.25	1.09
Khorasan	1.57	1.74	1.70	1.14	1.86	1.42
Semnan	0.88	0.84	0.93	0.95	0.72	0.82
West ,Azerbaijan	1.10	1.06	1.27	1.11	0.62	0.72
Mazandaran	1.25	1.01	1.05	0.88	0.56	0.72
Qazvin	1.83	1.57	1.52	1.50	0.99	1.01
Fars	1.91	1.79	2.51	3.41	1.40	1.74
Markazi	0.86	0.95	1.31	1.11	1.25	1.14
Gilan	2.41	1.66	2.72	1.83	1.58	1.58
Qom	1.00	0.94	1.01	1.27	0.31	0.24
Khozestan	1.54	1.53	1.84	1.93	1.37	1.39
Yazd	1.04	1.23	1.07	1.09	0.78	0.73
Hamadan	1.72	1.15	1.16	1.44	0.92	0.87
East ,Azerbaijan	1.42	1.39	2.67	2.57	1.65	1.61
Golestan	1.31	1.29	1.17	1.17	0.84	0.57
Ardabil	1.32	1.24	1.47	1.34	1.15	1.50
Kermanshah	1.87	0.99	1.53	1.16	1.02	0.76
Mahaal and Bakhtiari Chahar	1.60	0.82	0.84	1.04	0.65	0.68
Zanjan	1.16	0.89	0.92	0.65	0.51	0.62
Hormozgān	1.47	1.75	7.67	6.20	4.70	3.59
Kerman	2.31	1.31	2.49	1.50	1.22	1.86
Lorestan	1.45	1.40	1.34	1.45	0.91	1.59
Kurdistan	1.85	2.31	2.09	2.10	1.30	4.97
Bushehr	0.78	0.62	0.43	0.55	1.00	0.90
Baluchestan Sistan and	0.77	0.90	1.06	0.65	0.90	0.81
Ilam	0.59	0.58	0.61	0.53	0.37	0.32
and Boyer-Ahmad Kohgiluyeh	2.61	1.13	1.12	2.83	1.60	2.46
Iran	1.23	1.20	1.48	1.39	1.11	1.06

2.5.1.2. Subsidy reform plan and Total factor productivity

In the last section, TFP for each industry was estimated for each year. Thus, we tried to determine the effect of subsidy reduction on TFP to answer this study's questions. Regression 4 was estimated, and the results were reported in Table 2-8.

$$TFP_{itk} = \alpha_0 + \alpha_1 Control_{itk} + \alpha_2 R\&D_{itk} + \alpha_3 Import_{itk} + \alpha_4 Subsidy_t + \alpha_5 Sanction_t + year * ISIC_i + year * Provinces + provinces * ISIC + \varepsilon_{itk} \quad (4)$$

In this regression, R&D represents investment in the research and development sectors; the Import variable describes the importation of capital goods. Control is the control variable, and it includes Exchange, which indicates the real exchange rate. In the control variable, Util shows capital utilization, and Invento, as a part of the control variable, shows the inventory changes in this model. Moreover, to get more accurate results, three interactions were applied in the regression: year * ISIC_i, year * Provinces, and provinces * ISIC.

The variable sanction is the dummy variable; it is zero before 2011, and otherwise 1. In June 2011, the United Nations passed Security Council Resolution 1984, and based on this regulation, supported Iran's sanction for one year. Subsidy also is a dummy variable; it is equal to zero before the year 2010, and otherwise 1.

Table 2-8:Regression results

Model	1	2	3	4
Dependent	TFP	TFP	TFP	TFP
Subsidy	-0.8*** (0.21)	-1.14*** (0.22)	-1.24*** (0.21)	-1.22*** (0.22)
Sanction	-1.52*** (0.38)	-1.93** (0.24)	-2.03*** (0.17)	-2.03*** (0.19)
R&D	1.07** (0.09)	1.63** (0.11)	1.61** (0.13)	1.61** (0.13)
Import	-0.91*** (0.01)	-0.84*** (0.01)	-0.81*** (0.01)	-0.18*** (0.01)
Util		7.67* (1.01)	8.71* (1.21)	8.75* (1.22)
Exchange		1.51* (0.005)	1.50* (0.006)	1.50* (0.006)
Invento			-0.11* (0.001)	-0.11* (0.001)
Dummy-private ownership				0.26* (0.08)
_cons	6.67 (0.18)	6.29 (0.18)	6.19 (0.19)	6.04 (0.19)
N	8842	8842	8842	8842
R-sq	0.73	0.79	0.81	0.81
Std errors in parentheses	* p<0.05	** p<0.01	*** p<0.001	

Note: Result were reported based on model 4 since it has the highest R-squared value.

Note: This table shows the results of regression 4 in the first model control is equal to zero.

Note: There are two distinct ownerships in dataset public and provide ownership

The results show that the subsidy-reduction plan decreased TFP during the timeline of this study by -1.22 points.

Additionally, there is a positive relationship between investment in research and development (R&D) and TFP, and it indicates that more investment in this section will increase the TFP by 1.61 points.

Other results show that import of capital goods caused TFP to be reduced among industries by -0.18 points and that higher capital use in the industrial areas leads to higher TFP.

Moreover, the outcomes confirm that increased inventory causes lower TFP by -0.11 points. Furthermore, industries with a higher number of private-ownership firms had a higher TFP compared to firms with public ownership for the timeline of this study.

2.5.2. Robustness checks:

A common activity in empirical research is a “robustness check.” Researchers analyze how particular “core” regression coefficients perform when the regression term is adjusted by adding or eliminating regressors. If the coefficients are credible and robust, this is usually interpreted as proof of structural validity (Lu & White, 2014).

Consequently, this study performed a robustness check to confirm the results. Two models were applied for robustness testing. First, the data set was divided into smaller portions based on the province’s average temperature, and regression four was run on these four data sets. Second, the data set was categorized based on industry energy consumption, and the primary regression was run to confirm the main results.

2.5.2.1 Robustness checks based on provinces average temperatures

Some studies indicated that there is an inevitable relationship between TFP and temperature, since severe weather conditions cause more energy consumption (e.g.,

Table 2-9: Division based on the annual air temperature of the provinces

Cold	Hot	Very Hot	Moderate
Ardabil	Tehran	Kerman	Gilan
Azerbaijan, East	Fars	Khorasan, South	Golestan
Azerbaijan, West	Qom	Khuzestan	Mazandaran
Hamdan	Khorasan, Razavi	Bushehr	Isfahan
Kermanshah		Hormozgan	
Ilam	Kohgiluyeh and Boyer-Ahmad		Khorasan, North
Lorestan	Alborz	Sistan and Baluchestan	Kurdistan
Zanjan	Markazi	Yazd	
Chahar Mahaal and Bakhtiari		Semnan	

Based on these categories, most of Iran's provinces were categorized as hot or very hot temperature areas, followed by provinces with cold temperatures. Iran has a variable climate; for example, in the northwest, winter is cold, with heavy snowfall and sub-glacial temperatures, spring and autumn are relatively mild, and summers are dry and hot. The winters are mild in the south, and the summers are very hot, with the average daily temperature in July exceeding 38 °C (100.4 °F). In Khuzestan, the summer heat is associated with high humidity (Iran Meteorological Organization, 2014).

Table 2-10:Regression results based on temperatures categories

Model	Cold	Hot	Very Hot	Moderate
Dependent	TFP	TFP	TFP	TFP
Subsidy	-1.6** (0.31)	-2.04* (0.52)	2.31*** (0.47)	-0.72** (0.22)
R&D	1.07** (0.09)	1.23** (0.11)	1.81** (0.23)	1.11* (0.19)
Import	-0.91*** (0.11)	-0.84*** (0.21)	-0.81*** (0.01)	-0.18*** (0.08)
Util	6.43* (2.24)	7.67* (1.01)	8.71* (1.21)	6.75* (1.22)
Exchange	1.02 (0.01)	1.51* (0.005)	1.50* (0.006)	1.50* (0.006)
Invento	-0.05* (0.007)	-0.07** (0.004)	-0.9* (0.009)	-0.04 (0.007)
ownership	0.12** (0.14)	0.12** (0.07)	0.12** (0.29)	0.26* (0.08)
Sanction	-1.02* (0.61)	-1.32* (0.39)	-1.27* (0.46)	-0.92** (0.63)
_cons	4.67 (0.19)	2.29 (0.11)	3.19 (0.21)	4.04 (0.29)
N	4256	2205	1841	540
R-sq	0.52	0.43	0.38	0.21
Std errors in parentheses	* p<0.05	** p<0.01		*** p<0.001

The outcomes shown in Table 2-10 reveal that subsidy reduction caused TFP for industries located in very hot temperature provinces increase more compared to those in other regions. The second highest TFP decrease belongs to industries located in the hot areas,

with a value of -2.04. Substitute reform change TFP for six provinces with moderate temperature by a value of -0.72, however, meant that they were affected less than other groups.

2.5.2.1 Robustness checks based on industrial energy consumption

According to the U.S. Department of Energy, industries are categorized into three different groups based on energy consumption. These groups are high energy-intensive, moderate energy-intensive, and low energy-intensive; more details about this matter are shown in Table 2-11. Model 4 regression was applied to every group to examine the robustness for this study.

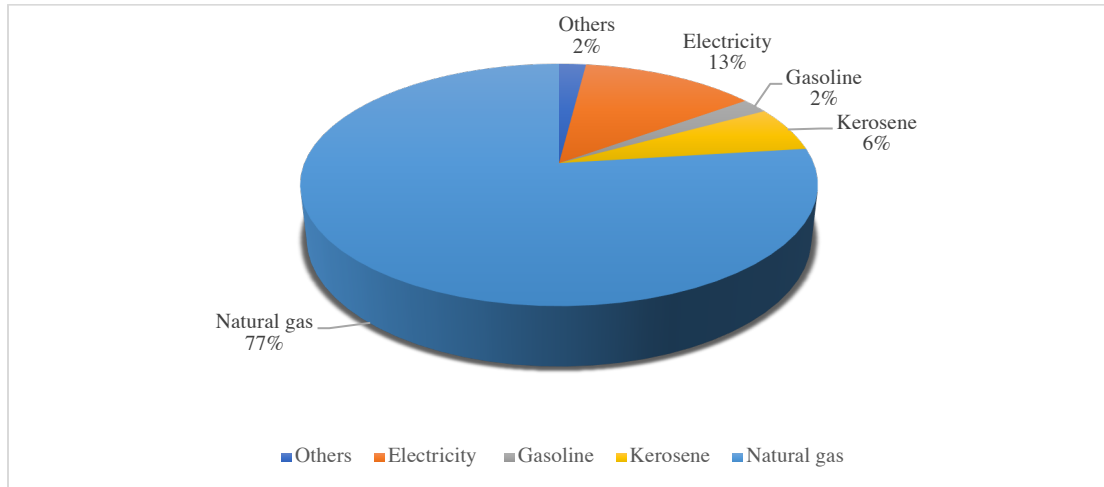
Table 2-11: ISIC classifications based on energy using

The intensity of energy consumption	ISIC	Description of activities
	17	Manufacture of textiles
High energy -intensive	21	Paper and paper products
	23	Coke and refined petroleum products
	24	Chemical products
	26	Non-metallic mineral products
	27	Manufacture of basic metals
	15	Food products and beverages
	18	Wearing apparel, dressing and dyeing
Moderate energy-intensive	19	Manufacture of leather products
	20	Wood and wood products
	22	Printing and publishing
	24	Rubber and plastic products
	28	Fabricated metal products
	16	Tobacco products
	29	Machinery and equipment N.E.C.
	30	Office, accounting and computing machinery
Low energy-intensive	31	Electrical machinery and apparatus N.E.C.
	32	Radio, TV and communication equipment
	33	Medical, precision and optical instruments
	34	Motor vehicles, trailers and semi-trailers
	35	Other transport equipment
	36	Furniture and other manufacturing N.E.C.
	37	Recycling

Moreover, the consumption of energy by type of carrier in the timeline of this study shows that natural gas and electricity were both consumed more than other energy carriers in industrial workshops with ten employees. Their share of natural gas was 77.1%

(equivalent to 178 million barrels of crude oil), and electricity shares were 12.7% (equivalent to 29 million barrels of crude oil) of all energy carriers.

FIGURE 2. 6:Percentage of energy consumption



Note: This figure shows percentage of energy consumption by carrier type in industrial workshops of 10 employees and more

The study of the amount of energy consumed by industrial workshops based on type of activity shows the groups with the highest energy consumption to be “production of other non-metallic mineral products” and “production of basic metals.” They are followed by the groups “chemical and chemical products industries” and “coke coal industries, oil refineries, and nuclear fuels.” The energy consumption of “production of other non-metallic mineral products” was equivalent to 69 million barrels of crude oil, and the “production of basic metals” used the equivalent of 61 million barrels of crude oil (30% and 26.1% of the total, respectively). The lowest energy consumption was in “recycling of metallic and non-metallic waste.”

The Model 5 regression was run on these three groups, and results are reported in Table 2-12.

Table 2-12:Regression results based on energy usage categories

Model	High	Low	Moderate
Dependent	TFP	TFP	TFP
Subsidy	-1.52* (0.46)	-1.15* (1.92)	-1.02** (0.51)
R&D	2.46** (0.59)	3.57** (0.91)	1.09* (0.71)
Import	-1.83** (1.59)	-1.03*** (2.09)	-0.95*** (1.04)
Util	3.17* (0.47)	5.29 (0.4)	4.21 (0.82)
Exchange	1.39 (0.93)	2.16* (0.28)	2.04* (0.76)
Invento	-0.19* (0.011)	-0.26** (0.031)	-0.34 (0.017)
Dummy-private ownership	0.32** (0.08)	0.47 (0.15)	0.51* (0.01)
Sanction	-0.62 (0.91)	-1.02* (0.49)	-0.83** (0.45)
_cons	6.32 (0.25)	9.15 (0.31)	7.93 (0.17)
N	1989	4327	2560
R-sq	0.47	0.58	0.53
Std errors in parentheses			
*p<0.1	**p<0.05	***p<0.01	

Note: This tables shows the outcomes of model 5 regression on three selected groups

The results indicate that industries with low energy use were the least affected by the subsidy reform in 2010 compared to two other groups. The highest effect of subsidy reduction on TFP rate belongs to high energy consumption sectors, with a value of -1.52.

Hence, both methods confirm the main results and indicate that the variable of interest (subsidy reduction) has a negative effect on TFP.

2.6. Discussion of Results

This study shows that TFP decreased after the subsidy reduction law was implemented. Although some industries, such as “manufacture of wood and products of wood and cork, except furniture” and “manufacture of articles of straw and plaiting materials (ISIC-20)” had different outcomes, overall, TFP was reduced.

Some previous studies imply that subsidies may positively affect production but at the same time could negatively affect productivity (Rizov, 2013; Ciaian & Swinnen, 2009). Subsidies may negatively influence firms’ productivity since production composition would be altered, causing inefficient allocations. As a result, beneficiary firms may adjust their performance and begin spending on subsidy-seeking projects that are comparatively less productive (Alston & James, 2002).

In 2011, the sanctions imposed by the Central Bank of Iran caused the rial (Iran’s currency) to be dramatically devalued against the U.S. dollar, by around 300%. It led to price increases for imported raw materials and different types of equipment. Sanctions also made the transfer of money for buying essential items more difficult (Farzin et al., 2011). All these events caused sanctions to have a highly negative effect on the economy. Therefore, it caused a reduction in TFP since sanctions made it more difficult to import more efficient industrial equipment and materials (Nosratabadi, 2019). The results of this study are consistent with this conclusion. Sanctions reduced TFP by -2.03 points.

Other outcomes of this study show more investment in research and development caused TFP to increase by 1.61. Previous studies indicated that investment in the research

and development section caused more innovation, leading to higher productivity for firms (Griliches,1998).

There are two types of ownership among these industrial workshops: private and public. The results of this study show that TFP is higher by 0.26 points for firms with private ownership. Empirical studies have shown that the work incentive is higher for private companies' employees, but the effects of ownership on productivity is ambiguous (Gao, 2013).

Other outcomes indicated that the import of capital goods caused TFP to decrease among industries by -0.18 points. In addition, higher capital use in industrial areas led to a higher TFP (by 8.75), and the results are statistically significant. Moreover, the outcomes confirm that increase in inventory lowers TFP by 0.11 points.

2.7. Conclusions

The economic history of Iran, especially after the emergence of oil led to a significant government role in the economy and various direct and indirect subsidies to industry. Productivity is a critical factor in promoting economic growth that every Iranian government has tried to address since the 1979 revolution.

In 2010, the most significant economic intervention in the history of the country was performed. In general, the government took steps to liberalize prices and allocate the proceeds to the fields of, for example, development, household, production, or health. In the meantime, production barriers were given much attention due to price liberalization, especially in the field of energy and raw materials. Based on this law, the subsidy of energy and food was decreased, causing the price of these items to increase substantially. It also had a significant effect on all economic sectors, such as industries, households, and all

other sectors of the economy. In Iran, the highest amount of energy consumption belongs to the industrial sectors (Central Bank of the Islamic Republic of Iran, 2015). This is mainly due to old technologies, worn-out equipment, and in particular a lack of proper management of energy consumption in this sector (Statistical Center of Iran, 2016).

Subsidy reform can lead the net cost value of products to increase, affecting companies' financial condition when they are forced to keep the prices unchanged. The rising costs of production inputs, especially energy, reduce corporate profit margins and put companies at risk of bankruptcy. Meanwhile, some companies have improved their situation after implementing the targeted subsidies law. According to Iran's government, some important goals of this law were improving living standards, stopping wasteful consumption and enhancing efficiency, decreasing fuel smuggling, and reducing income disparities (Hassanzadeh, 2012).

This study examined the effects of subsidy reduction plan on the TFP of industrial workshops with ten or more employees. In the first step, we estimated TFP using production functions. In fact, the technological upgrade of an economy is evaluated by the annual growth in TFP (SEROT, 1993). Then, to answer the research questions, estimated TFP was used in the main regression to examine the effects of the subsidy-reduction plan on TFP.

The data set was collected from the Iran Statistics Center and Central Bank of Iran's websites. The data set used was provided in Farsi and translated into English.

Several studies have been performed on subsidy and TFP, but this study is different since it used Iran as a case study and because both the subsidy-reduction law and international sanctions coincided with the timeline of this study.

Based on this study's findings, the observed firms did not achieve an increase in productivity after the subsidy reduction was implemented. Hence, government may need to revise the law to support industrial workshops to become more productive. The finding also shows that subsidy reduction can affect TFP differently for different geographical locations based on their average temperature. Although the government attempted to control the effect of subsidy reduction by increasing energy prices based on geographical location for households, the law did not follow the same path for the industry. Moreover, the outcomes suggest that governments should modify the regulations for different industries based on their level of energy use (high, moderate, or low).

This study's results can help modify the current subsidy-reduction law or inspire future regulations in this area to become more efficient. The results even can be helpful for other countries that want subsidy reform and for which increasing productivity is a concern. Lastly, the result can help firms become more productive by allocating their inputs more efficiently.

Overall, the findings show that subsidy reduction could not help industrial workshops become as productive as governments expected. A year after the subsidy-reform plan was implemented, the international sanctions imposed on Iran had had a significant effect on the economy. Hence, the government could not solely consider the results of this regulation

2.8. Reference:

- Aldieri, L., & Cincera, M. (2009). Geographic and technological R&D spillovers within the triad: Micro evidence from US patents. *The Journal of Technology Transfer*, 34(2), 196-211.
- Alesina, A., Spolaore, E., & Wacziarg, R. (2005). Trade, growth and the size of countries. In *Handbook of economic growth* (Vol. 1, pp. 1499-1542). Elsevier.
- Araar, A., & Verme, P. (2012). Reforming subsidies: A toolkit for policy simulations. The World Bank.
- Ayakwah, A., & Mohammed, J. (2014). Fuel price adjustments and growth of SMEs in the New Juaben Municipality, Ghana. *International Journal of Small Business and Entrepreneurship Research*, 2(3), 13-23.
- Baier, S. L., Dwyer Jr, G. P., & Tamura, R. (2006). How important are capital and total factor productivity for economic growth? *Economic Inquiry*, 44(1), 23-49.
- Bernini, C., & Pellegrini, G. (2011). How are growth and productivity in private firms affected by public subsidy? Evidence from a regional policy. *Regional Science and Urban Economics*, 41(3), 253-265.
- Bernini, C., & Pellegrini, G. (2011). How are growth and productivity in private firms affected by public subsidy? Evidence from a regional policy. *Regional Science and Urban Economics*, 41(3), 253-265.
- Breisinger, C., Engelke, W., & Ecker, O. (2011). Petroleum subsidies in Yemen: leveraging reform for development. The World Bank.
- Bronzini, R., & De Blasio, G. (2006). Evaluating the impact of investment incentives: The case of Italy's Law 488/1992. *Journal of urban Economics*, 60(2), 327-349.
- Cameron, G., Proudman, J., & Redding, S. (2005). Technological convergence, R&D, trade and productivity growth. *European Economic Review*, 49(3), 775-807.
- Chen, J. Y., Dimitrov, S., & Pun, H. (2019). The impact of government subsidy on supply Chains' sustainability innovation. *Omega*, 86, 42-58.
- Davis, L. W. (2014). The economic cost of global fuel subsidies. *American Economic Review*, 104(5), 581-85.
- De Moor, A. (2001, May). Towards a grand deal on subsidies and climate change. In *Natural Resources Forum* (Vol. 25, No. 2, pp. 167-176). Oxford, UK: Blackwell Publishing Ltd.

- Dong, B., & Li, X. (2014). Estimating outage costs in China: a modified tobit model. *Emerging Markets Finance and Trade*, 50(5), 182-192.
- Esmaeili, A., Karami, A., & Najafi, B. (2013). Welfare effects of alternative targeted food subsidy programs in Iran. *Food security*, 5(3), 451-456.
- Fisher-Vanden, K., & Jefferson, G. H. (2008). Technology diversity and development: Evidence from China's industrial enterprises. *Journal of Comparative Economics*, 36(4), 658-672.
- Fisher-Vanden, K., Mansur, E. T., & Wang, Q. J. (2015). Electricity shortages and firm productivity: evidence from China's industrial firms. *Journal of Development Economics*, 114, 172-188.
- Flachenecker, F., Bleischwitz, R., & Rentschler, J. E. (2017). Investments in material efficiency: the introduction and application of a comprehensive cost-benefit framework. *Journal of Environmental Economics and Policy*, 6(2), 107-120.
- Gao, H., Harford, J., & Li, K. (2013). Determinants of corporate cash policy: Insights from private firms. *Journal of Financial Economics*, 109(3), 623-639.
- Gharibnavaz, M. R., & Waschik, R. (2015). Food and energy subsidy reforms in Iran: A general equilibrium analysis. *Journal of Policy Modeling*, 37(5), 726-741.
- Griliches, Z. (1998). Issues in assessing the contribution of research and development to productivity growth. *NBER Chapters*, 17-45.
- Harris, R., & Li, S. (2019). Government assistance and total factor productivity: firm-level evidence from China. *Journal of Productivity Analysis*, 52(1), 1-27.
- Harris, R., & Trainor, M. (2005). Capital subsidies and their impact on Total Factor Productivity: Firm-level evidence from Northern Ireland. *Journal of Regional Science*, 45(1), 49-74.
- Heshmati, A., & Rashidghalam, M. (2020). Estimation of technical change and TFP growth based on observable technology shifters. *Journal of Productivity Analysis*, 53(1), 21-36.
- Jiang, H., Hu, Y., Zhang, H., & Zhou, D. (2018). Benefits of downward earnings management and political connection: evidence from government subsidy and market pricing. *The International Journal of Accounting*, 53(4), 255-273.
- Kim, W., Chattopadhyay, D., & Park, J. B. (2010). Impact of carbon cost on wholesale electricity price: a note on price pass-through issues. *Energy*, 35(8), 3441-3448.

- Lee, C. C., & Chiu, Y. B. (2011). Electricity demand elasticities and temperature: Evidence from panel smooth transition regression with instrumental variable approach. *Energy Economics*, 33(5), 896-902.
- Lin, B., & Jiang, Z. (2011). Estimates of energy subsidies in China and impact of energy subsidy reform. *Energy Economics*, 33(2), 273-283.
- Lin, B., & Li, A. (2012). Impacts of removing fossil fuel subsidies on China: How large and how to mitigate? *Energy*, 44(1), 741-749.
- Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of econometrics*, 178, 194-206.
- Männasoo, K., Hein, H., & Ruubel, R. (2018). The contributions of human capital, R&D spending and convergence to total factor productivity growth. *Regional Studies*, 52(12), 1598-1611.
- Mattsson, P. (2019). The impact of labour subsidies on total factor productivity and profit per employee. *Economic Analysis and Policy*, 62, 325-341.
- Moffat, J. (2014). Regional Selective Assistance in Scotland: Does it make a difference to plant productivity? *Urban studies*, 51(12), 2555-2571.
- Myers, N. (2001). *Perverse subsidies: how tax dollars can undercut the environment and the economy*. Island Press.
- Nosratabadi, J. (2019). Employment, Skill Upgrading, and International Trade: The Case of Sanctions Against Iran. *Skill Upgrading, and International Trade: The Case of Sanctions Against Iran* (December 3, 2019).
- Pellegrini, G., Bernini, C., & Cerqua, A. (2015). Public subsidies, TFP and Efficiency: a tale of complex relationships.
- Pelucha, M., Kveton, V., & Potluka, O. (2019). Using mixed method approach in measuring effects of training in firms: Case study of the European Social Fund support. *Evaluation and program planning*, 73, 146-155.
- Plante, M. (2014). The long-run macroeconomic impacts of fuel subsidies. *Journal of Development Economics*, 107, 129-143.
- Rahmani, O., Rezania, S., Beiranvand Pour, A., Aminpour, S. M., Soltani, M., Ghaderpour, Y., & Oryani, B. (2020). An Overview of Household Energy Consumption and Carbon Dioxide Emissions in Iran. *Processes*, 8(8), 994.
- Rahmati, M. H., & Karimirad, A. (2017). Subsidy and natural resource curse: Evidence from plant level observations in Iran. *Resources Policy*, 52, 90-99.

- Rentschler, J., Kornejew, M., & Bazilian, M. (2017). Fossil fuel subsidy reforms and their impacts on firms. *Energy Policy*, 108, 617-623.
- Reza Farzin, M., Guillaume, D. M., & Zyteck, R. (2011). Iran-the chronicles of the subsidy reform. *IMF Working Papers*, 1-28.
- Roach, B., & Harris, J. M. *Energy Economics and Policy*.
- Salehi-Isfahani, D. (2016). Energy subsidy reform in iran. In *The Middle East Economies in Times of Transition* (pp. 186-195). Palgrave Macmillan, London.
- Salehi-Isfahani, D., & Mostafavi-Dehzoeei, M. H. (2018). Cash transfers and labor supply: Evidence from a large-scale program in Iran. *Journal of Development Economics*, 135, 349-367.
- Sekar, C., Rai, R. V., & Ramasamy, C. (1996). Role of minor forest products in tribal economy of India: a case study. *Journal of Tropical Forest Science*, 280-288.
- Serot, D. E. (1993). Estimating total factor productivity using parametric, nonstochastic cost frontiers. *Journal of Productivity Analysis*, 4(4), 407-418.
- Sijm, J., Neuhoff, K., & Chen, Y. (2006). CO2 cost pass-through and windfall profits in the power sector. *Climate policy*, 6(1), 49-72.
- Stern, D. I. (2012). Interfuel substitution: a meta-analysis. *Journal of Economic Surveys*, 26(2), 307-331.
- Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, 26(1), 98-128.
- Vedenov, D., & Wetzstein, M. (2008). Toward an optimal US ethanol fuel subsidy. *Energy Economics*, 30(5), 2073-2090.
- Verme, P. (2016). Subsidy reforms in the Middle East and North Africa region: a review.
- Worrell, E., Bernstein, L., Roy, J., Price, L., & Harnisch, J. (2009). Industrial energy efficiency and climate change mitigation. *Energy efficiency*, 2(2), 109.

Chapter 3

Weather-based COVID-19 Spread Pattern and Public Policy Response

3.1. Introduction:

The COVID-19 pandemic has weakened the economy and changed billions of lives around the world. The pandemic and its associated lockdown regulations worldwide substantially shocked the global economy's supply and demand. Most countries were not ready for or immune to such a strong shock. Almost all parts of the economy were affected by the COVID-19 pandemic; for example, in the tourism sector, demand decreased significantly, as, on the one hand, people were afraid of traveling, and on the other hand, government regulations made travel almost impossible (Wu et al., 2020).

Researchers worldwide continue to struggle to identify possible responses, determine the results of regulations, and understand efficient ways to control the pandemic. In some cases, these studies have helped governments control the pace of spread by determining the outcomes of some regulations, such as quarantines and lockdowns. Some studies have tried to understand how government lockdown regulations have changed mobility and controlled the virus spread (see, e.g., Pepe et al., 2020; Kishore et al., 2020).

The effect of weather conditions on respiratory diseases is one of the most critical factors indicated by researchers. Respiratory diseases are widespread globally and are the third cause of non-communicable mortality (Gakidou et al., 2017). Many studies have confirmed positive connections between public health and weather indicators, such as temperature, humidity, or wind speed (Beggs & Bambrick, 2005; McMichael et al., 2006).

Previous studies have confirmed a strong relationship between SARS, another member of the coronavirus family, with weather conditions (see, e.g., Yuan et al., 2005). This relationship could also be true for COVID-19 since it is a member of the corona families.

One crucial question during the COVID-19 pandemic is whether there is a connection between weather conditions and the COVID-19 virus spread. This study tries to answer this question. Determining this connection could help governments make more informed policy decisions and become more prepared for the next waves and next pandemics. Some previous studies (Barreca & Shimshack, 2012; Casanova et al., 2010; Chan et al., 2011; Shaman & Kohn, 2009; van Doremalen et al., 2013) showed that seasonal and geographic climatic change could affect the spread of respiratory diseases, and winters are the traditional prevalence peaks of these diseases. These studies have concentrated mainly on influenza, SARS, and MERS-COV, but few of them provided valuable information about COVID-19 since it is distinctive from other recognized viruses in its spread and pathogenicity.

Furthermore, the relationship between governments policies and the spread of COVID-19 is another direction of the respiratory-disease literature, which can help governments control the pandemic. There are several studies about government policy and its effects on controlling respiratory diseases. The close relationship between government and society is not inevitable during abnormal conditions like the COVID-19 pandemic. This study also tries to identify the effect of government regulations on controlling the COVID-19 transmissions. On February 23, 2020, the South Korean government announced a Contagious Disease Alert, level 4 (red) for the whole country to control the pandemic. Based on this regulation, the government began to restrict foreigners from

entering the country and stop all schools, universities, restaurants, bars, and public facilities from operating. Finding how much government regulation effectively controlled the pandemic can also help the South Korean government manage the pandemic better.

To answer the research questions, first, we tried to find the correct number of lag variables by applying the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Weather condition indicator variables are required lags since, according to the World Health Organization (WHO, 2019), COVID-19 symptoms appear on average 5–6 days after infection. Most previous studies about the relationship between respiratory diseases and weather conditions have used lag variables (see, e.g., Chan et al., 2011; Barro, 1991; Barreca & Shimshack, 2012).

Second, the Generalized Additive Model (GAM) was applied to answer the research question. The GAM was selected for this study because it can work with many distributions, such as Gamma, Normal, and Poisson, and the distribution of dependent variables in this study is Poisson. This model is common in studies of epidemiology, air quality, and medicine. More details about the GAM are described in the Model Selection section.

The outcomes prove that higher minimum temperature and higher relative humidity lead to lower transmission. So, based on the results during winter, the number of infections would be expected to speed up again. Therefore, governments need to be prepared and pass regulations to reduce the COVID-19 infection rate. Moreover, minimum wind speed and atmospheric pressure have a negative relationship with the dependent variable. One should also note that the Infectious Disease Alert, level 4 (red) policy lowered infection numbers. Other studies have confirmed that government regulations cause fewer infections

worldwide, especially when some implemented rules to restrict mobility, such as lockdowns.

Compared to previous studies, this research uses more features that can affect the spread of the COVID-19 virus. This study is different since it tries to connect two different literature angles by applying climate indicators and government regulation in the same regression as right-hand side variables. The Infectious Disease Alert, level 4 (red) policy caused a decrease of more than 10% in mobility for the whole country; this policy can help control the mobility effect in this study since the city-level mobility data for South Korea are not available. Other contributions of this study include finding reliable predictive models for daily COVID-19 positive cases in countries with the most advanced health care systems and forecasting numbers of infections by applying the GAM.

COVID-19 is a new virus, and few studies have tried to investigate its effect on, for example, the economy or environment. This research reveals new approaches to recognizing and stopping the spread of this virus with its connection to weather conditions and relevant parameters.

This paper is organized as follows. Section 2 presents the related works. Section 3 introduces our new methodology and the proposed approaches. Section 4 presents the experimental observations. Finally, the conclusions and possible future outcomes are introduced in Section 5.

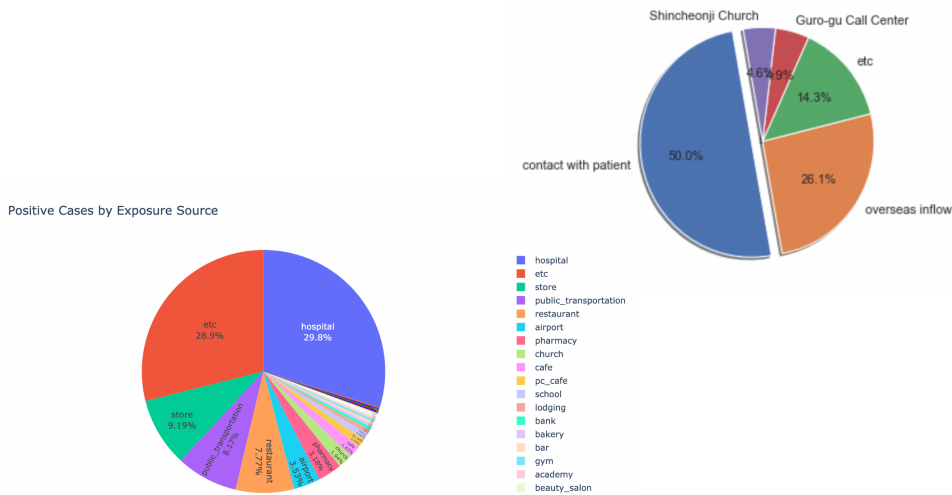
3.2. Literature review:

Coronaviruses (CoVs) belong to the Coronaviridae family, a group of enveloped, positive-sensed, single-stranded RNA viruses (Ye et al., 2020). These types of viruses are

named CoVs due to their crown shape (Su et al., 2016). Coronaviruses are a widespread group of viruses that are prevalent among animals. In some exceptional cases, they can be transferred from animals to humans, called zoonotic transmission (The US Centers for Disease Control and Prevention, 2019).

Since December 31, 2019, when the first positive case of COVID-19 was detected in Wuhan, China, people worldwide have str with this disease (Lu et al., 2020). Based on recent studies, COVID-19 transmission has been mostly happening through physical contact; for example, a study in China reported that 41% of patient infections was due to person-to-person hospital visits (Wang et al., 2020). Furthermore, our results are consistent with this fact in the case of South Korea during the timeline of this study (as shown below).

FIGURE 3. 1:Place and the ways of infection in South Korea



The literature has two strands: disease spread patterns and weather connection, and public policy. First, let's focus on the literature on the relationship between weather conditions and respiratory disease. Epidemiological and laboratory research have recognized a connection between weather conditions and respiratory diseases (Chan et al.,

2011; Tan et al., 2005). Weather conditions have an essential role in human health. A study was performed by Cheng et al. in 2018 in the five most populated Australian cities (Sydney, Melbourne, Brisbane, Adelaide, and Perth) from 2000 to 2009, which analyzed the relationship between mortality rate and temperature variability. The results show that any significant temperature change can be associated with a higher mortality rate.

The risk of transmission can differ based on weather conditions related to respiratory diseases' durability on various surfaces (Casanova et al., 2010; van Doremalen et al., 2013). There is extensive literature about patterns between weather conditions and virus mortality. Additionally, another study demonstrated a significant relationship between weather conditions and the pneumonia mortality rate (Bull, 1980).

In 2003, Bi et al. tried to estimate how weather conditions could change the speed of spread by applying data for a daily number of positive SARS cases in Beijing and Hong Kong during the 2003 pandemic. This study indicated that higher temperatures could decrease infection; air pressure, however, was positively correlated with the number of positive SARS cases. Likewise, a study in South Korea showed that the number of positive influenza cases increased significantly on days with lower temperature and relative humidity (Park et al., 2019).

In 2018, Lu et al. conducted a study to investigate the relationship between the number of varicella infections and climatic factors in Guangzhou, China, between 2006 and 2018. They applied distributed nonlinear lag models to estimate the correlation between weather factors and the number of varicella infections. Their conclusions indicate that air pressure, sunshine hours, wind speed, and rainfall were positively correlated with

the dependent variable. Furthermore, the results show that the relationship between average temperature and varicella is nonlinear and a U-shape curve.

In 2020, Bashir conducted a study to analyze the relationship between weather indicators and the COVID-19 mortality rate in New York City. The weather condition variables in this study include average temperature, minimum temperature, maximum temperature, rainfall, average humidity, wind speed, and air quality. The outcome indicated that average temperature, minimum temperature, and air quality were significantly associated with the COVID-19 pandemic.

In 2021, Hossain et al. aimed to investigate the effect of weather-condition factors on the number of COVID-19 infected cases in South Asian countries—India, Afghanistan, Pakistan, Bangladesh, and Sri Lanka—from the day the first case was infected to August 31, 2020 in each country. The climate indicators in this study include aggregate rainfall, relative humidity, surface pressure, maximum and minimum temperature (in °C), maximum air pollutants, and maximum wind velocity (m/s). The Autoregressive Integrated Moving Average with Explanatory Variables (ARIMAX) model was applied as an estimator model. The outcomes show that aggregated rainfall, air pollution, and maximum wind velocity play an essential role in spreading coronavirus in these countries.

In 2021, Zhang et al. tried to estimate the effect of temperature and humidity on COVID-19 infection numbers for 1,236 areas worldwide. The weather data were collected from a large-scale satellite, and the results indicate that temperature and relative humidity have a negative relationship with the global spread of COVID-19.

Most previous research in this section has indicated that temperature and humidity negatively affect respiratory disease infections. So, in dry cold weather conditions, more

people are expected to get respiratory diseases. This phenomenon occurs since airways are saturated with a narrow layer of liquid, and when dry air is inhaled, that liquid evaporates faster than normal. This makes the throat dry, and a parched throat leads to irritation and swelling, which worsens the symptoms of respiratory diseases. Cold weather boosts the production of mucus, the protective layer of the throat. The mucus produced in cold weather, however, is especially concentrated and dense, which can lead to obstructions in the respiratory system and raise the likelihood of catching a cold or other respiratory disease (Wu et al., 2020).

There is another relevant strand of the respiratory disease literature. There are several studies on government policy and its effects on controlling respiratory diseases. For example, Nidhi and Jayaraman tried to estimate the effect of Indian government regulations on the amount of carbon dioxide released by factories. They used the information of people who had been hospitalized for respiratory illnesses in seven Delhi hospitals between 1998 and 2004.

For the patient population, they applied the Poisson regression model. Their results indicate that the regulations caused a notable reduction in the monthly average density of sulfur dioxide; the monthly average density of nitrogen dioxide, however, increased by 2%. These changes in air quality reduced the number of respiratory patients significantly, by about 12%.

A study in Vietnam examined the effectiveness of government policy during the coronavirus pandemic. The study's timeline was divided into four periods based on four principal government regulations: emergency responses, border and entry control measures, social isolation, and financial support. The findings show that the Vietnamese

policy system responded quickly, proactively, and virtually at multiple authority levels (Le et al., 2021).

Another research by Wielechowski et al. in 2020 tried to determine how government lockdown regulations in Poland affected mobility and the spread of COVID-19 between March 2 and July 19, 2020. The results indicated that lockdown decreased the number of positive COVID-19 cases in Poland but that social distance regulations were more effective than lockdown.

The close relationship between government and society is not inevitable during an abnormal condition like the COVID-19 pandemic. For example, one study on Vietnam showed how close communication among the government, social media, and science journalism caused public health sustainability during the COVID-19 pandemic (Le et al., 2020).

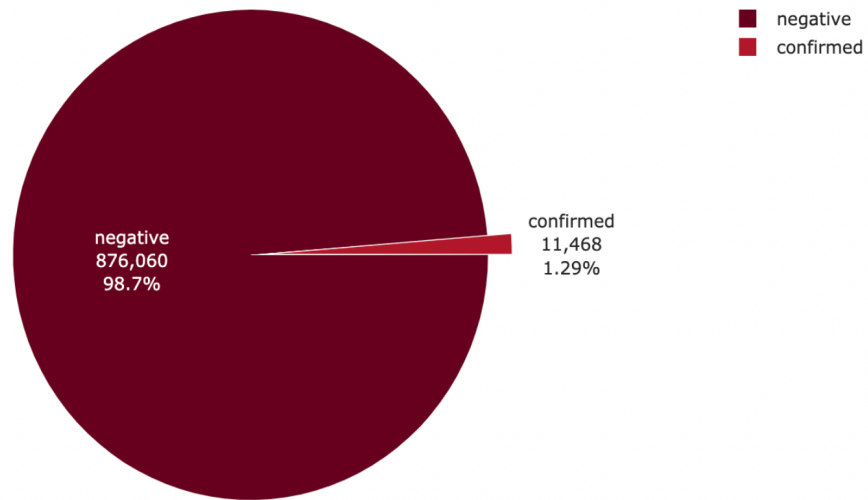
New technologies can also help governments manage the COVID-19 pandemic. A study in South Korea showed that using new technology helps the government decrease the pace of COVID-19 spread between January 2019 to June 2020 (Jeong et al., 2020). For example, using a mobile case-tracing application, mobile self-quarantine safety protection application, mobile self-diagnosis application, and drive-thru screening centers were beneficial for South Kores.

3.3. Data set description:

The data set includes 11,000 observations including all the confirmed coronavirus positive cases. The South Korean data set is a reliable source because of the country's aggressive testing protocol and sizable testing capacity. During this study's timeline, around 500,000 individuals were tested—1% of South Korea's population. Sub-national

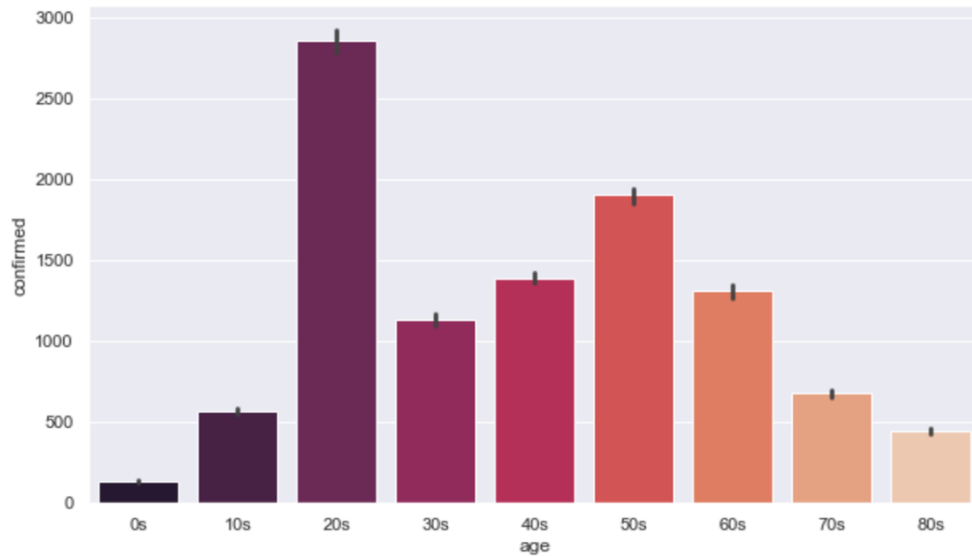
level data and spatial variation in weather within the country were used. We collected daily COVID-19 confirmed case data from the official website of KCDC and daily weather data from NASA's website. City-level panel data analysis was used for this research. The data covered the information of nine provinces as well as 95 cities in South Korea.

FIGURE 3. 2: Test results during the timeline of this study



The COVID-19 data set consists of the following three categories: patient data, policy data, and local data. The patient data category includes per-patient symptom_onset_date, confirmed_date, and state_deceased. Symptom onset date is when each patient first displayed symptoms of COVID-19, and confirmed date is the date that the patient was diagnosed. An age-specific number of patients was also indicated in this category. As shown in Figure 3, people in their twenties have the highest infection rate, but the lowest rate belongs to infants and kids.

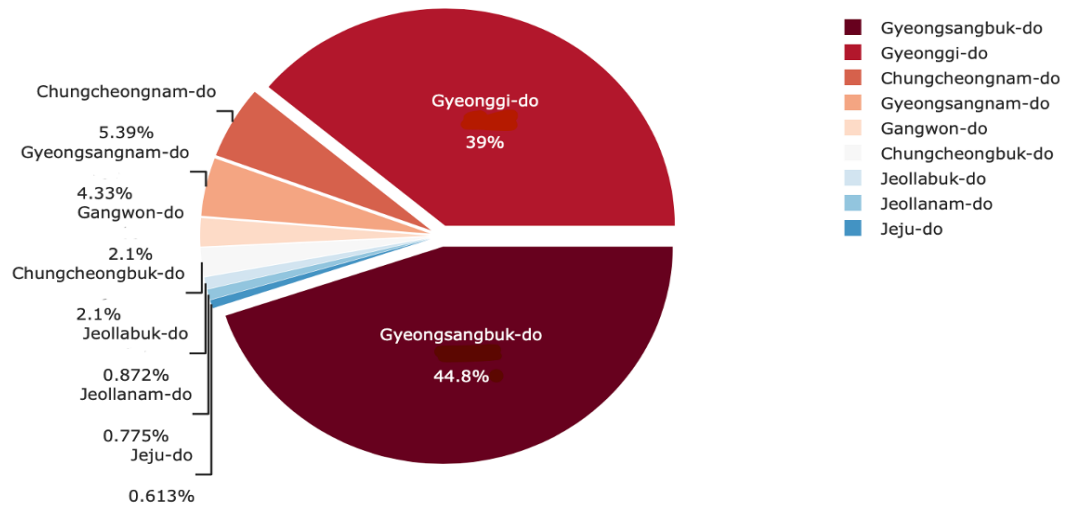
FIGURE 3. 3:Age-specific number of patients



The second category is regional data, which includes number of screening centers, land size, and population of each province. It also contains 178 cities in 17 South Korean provinces (126.30–129.47° east longitude and 33.45–38.19° north latitude).

The second category is regional data, which include the number of screening centers, land size, and population of each province. It also contains 178 cities in 17 South Korean provinces (126.3–129.47° east longitude and 33.4–38.19° north latitude). Among these 15 provinces, Gyeongsangbuk-do had the highest COVID-19 spread rate (35.2% of all cases detected during that study period), and the first COVID-19 case was seen in this province. Figure 4 shows more details about this.

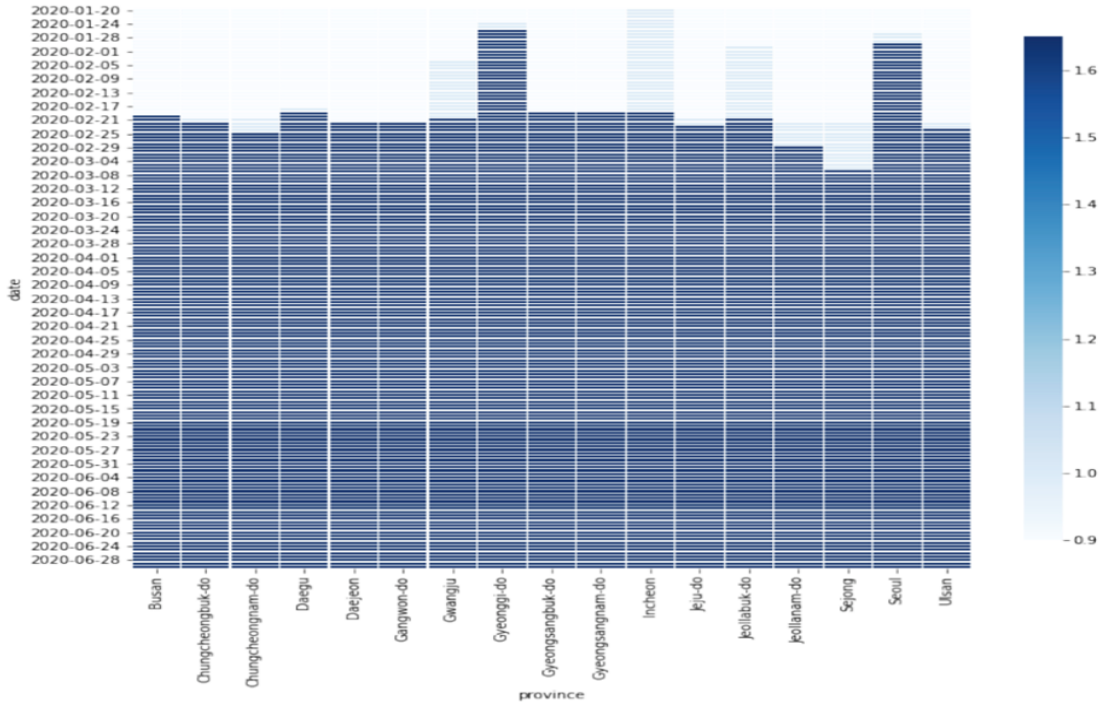
FIGURE 3. 4: The 9th highest number of infections case by regions in South Korea



The first case of COVID-19 in South Korea was detected on January 20, 2020: a Chinese woman aged 35 traveling from Wuhan to South Korea. On January 24, however, a 55-year-old South Korean male worker traveling from Wuhan with flu symptoms was detected as the first Korean COVID-19 positive case. From that day, all parts of the country have struggled with this disease. Lastly, policy data contains descriptions and the start and end dates of eight preventative policies enacted by the Korean government. These eight policies include a broad type of policies, such as social, educational, and immigration. To help understand the policies' effectiveness, the cumulative number of confirmed patients is appended as part of the data set.

The timeline of this study is January 20 to June 28, 2020; during the timeline of this study, from May 19 to June 12, 2020, South Korea faced the highest number of infections almost in all provinces, as shown in Figure 5.

FIGURE 3. 5: Number of infections for every province (time series plot)



The model also controlled for population density, the number of hospitals per 1,000 people, elderly population ratio, social distance order, day of the week, and young population ratio. Table 3 shows summary statistics of selected variables in this study.

Table 3- 1: Descriptive statistics of newly confirmed cases

Variable	Mean	Min	Max
Confirm cases	0.36	0	92
Humidity	73.6	49.1	100
Min-temperature	0.85	-16.89	14.71
Min-Wind	2.3	0.01	12.15
pressure	100.11	94.05	103.28

Although other weather condition variables applied in the regression are useful for explaining the relationship between weather conditions and the number of positive coronavirus cases, minimum temperature is the variable of interest for this study for the following reasons. Firstly, among all explanatory variables, it has the highest correlation with the independent variable.

Secondly, there is a strong correlation between temperature and health, as critical temperatures can jeopardize health condition by affecting the body's ability to control its internal temperature. There is also a strong relationship between extreme temperature and cardiovascular disease, respiratory disease, cerebrovascular disease, and diabetes-related conditions (Anderson et al., 2009).

Lastly, since temperature varies by season, it is more predictable than other climate variables. Hence, governments can be more prepared and face an issue more effectively based on temperature data. Furthermore, this study's dependent variable is count data, which is reflected in the number of occurrences of behavior in a fixed period, such as the number of positive COVID-19 cases (Cox et al., 2009). A proper model, then, is required to work with this type of data, which will be explained in more detail in the Model Selection section.

3.4. Methods:

This section discusses selecting an optimal number of lags for some selected variables and addressing the multicollinearity issue.

3.4.1. Lag variables:

Lag variables. According to WHO (2019), COVID-19 symptoms appear on average 5–6 days after infection. To accommodate the disease onset and detection lag, we

used The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the appropriate number of lags to be included in the model. Also, most previous studies on the relationship between respiratory diseases and weather conditions have used lag variables. A study tried to estimate the relationship between weather condition and varicella (chicken pox) cases using lag variables between seven and 21 days in Guangzhou (South China). The results showed positive relationships between aggregate rainfall, air pressure, sunshine hours, and the number of positive varicella cases (Zhang et al., 2017). In this research, the cumulative lag effects (lag0_2, lag0_3, lag0_4, lag0_5, lag0_6, lag0_7, lag0_8, lag0_9, and lag0_10) were examined, and seven-day lags was the best fit based on AIC and BIC.

AIC (Akaike, 1974) and BIC (Bayesian,1978) are valuable methods for estimating a predicted model’s likelihood or estimating anticipated values. For model selection criteria, minimum BIC and AIC were applied. The most suitable model is a model with the minimum AIC and BIC (Arabnia & Quoc-Nam, 2015).

Table 3- 2:Results of AIC and BIC test

Number of Lags	AIC	BIC
MA (0-2)	20571.99	21056.61
MA (0-3)	20442.46	20919.61
MA (0-4)	20410.02	20886.82
MA (0-5)	20332.09	20809.71
MA (0-6)	20297.28	20774.08
MA (0-7)	20089.09	20565.89
MA (0-8)	20186.12	20721.02
MA (0-9)	20198.24	20831.02
MA (0-10)	21187.04	20941.15

3.4.2. Variable selection:

To avoid multicollinearity, only one variable among minimum, maximum, and average temperature was selected for the regression. Choosing the variable with the highest correlation with the dependent variable (the number of COVID-19 confirmed cases) can solve this problem and increase the predictive power of the model (Ross, 2001). The same logic was applied for the wind speed variables. The daily minimum temperature and minimum wind speed were selected since they have a higher correlation with COVID-19 infection rate than the mean or maximum temperature and wind speed in the data set. Table 3 shows the summary statistics of these variables.

Table 3- 3: Variable selection based on covariance between the variables

	Confirmed Cases	Avg_Temp	Max_Temp	Min_Temp	Avg_Wind	Max_Wind	Min_Wind
Confirmed Cases	1	-0.05	-0.04	-0.05	-0.03	-0.03	-0.04
Avg_Temp	-0.04	1	0.92	0.91	0.51	0.57	0.3
Max_Temp	-0.04	0.92	1	0.72	-0.09	-0.5	-0.1
Min_Temp	-0.05	0.91	0.72	1	0.15	0.12	0.14
Avg_Wind	-0.03	0.03	-0.09	0.15	1	0.94	0.83
Max_Wind	-0.03	0.57	-0.05	0.12	0.94	1	0.68
Min_Wind	-0.04	0.51	-0.1	0.14	0.83	0.68	1

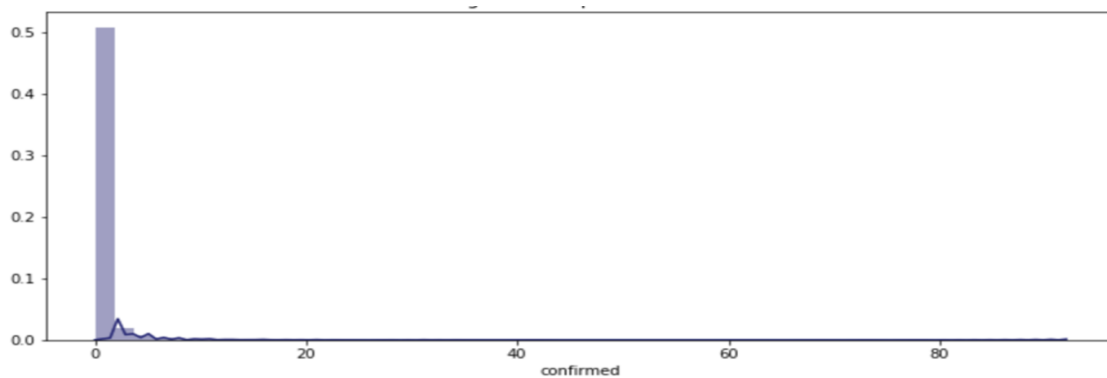
3.5. Estimation Models:

3.5.1. Model Selection:

The counted number of occurrences is the dependent variable of this study (number of COVID-19 infection). This class of data set is known as count data, which is non-negative and discrete. The ordinary linear regression models cannot fully explain this type of model for the following reasons. First, there is a high chance the model provides negative predicted values, and these results are impossible in theory.

Second, this type of data is skewed positively since many non-negative or zero values are in the data set. So, the data set distribution will be far from the normal distribution because of the zero values (Gardner et al., 1997). Count data define how often an event occurs; frequently, Poisson distribution is applied for this kind of data (Cameron et al., 2013). It is done the opposite way of the normal distribution data (Rainer, 2008). Figure 4 shows the distribution of number of positive COVID-19 cases in South Korea.

FIGURE 3. 6: Distribution of the dependent variable



For the following reasons the GAM was applied for this study. First, GAM implements a great fit for nonlinear and linear relations since it is easily adaptable. Second, GAMs can work with many distributions, such as Gamma, Normal, and Poisson. Lastly, it is very prevalent in epidemiology, air quality, and medicine (Dominici et al., 2002). The main models applied in environmental epidemiology are GAM and GLM², especially when

² GLM stands for The General Linear Model

the dependent variable is count data. Moreover, in these models, Poisson distribution is usually preferred with the logarithmic responding variable (Yang et al., 2017).

3.5.2. Regression analysis:

The GAM with Poisson distribution was applied in this regression based on earlier studies (Basu et al., 2008; Almeida et al., 2010). The following regression was applied in this study:

$$\begin{aligned} \text{LogE}(y_{it}) = & \alpha_0 + \text{Humidity}_{it-7} + \text{Pressure}_{it-7} + \text{Min_Tempreture}_{it-7} \\ & + \text{Min_Wind}_{it-7} + \text{IDA_RED}_i + \text{day}_t + \text{city}_i + \text{City} * \text{time} \\ & + \varepsilon_{it} \end{aligned} \quad (1)$$

In this model, the dependent variable ($E(y_{it})$) is the expected number of new daily infections on day t for city i . To control population density for each city, we divided the dependent variable by the population. The GAM model uses the logarithmic dependent variable to fix the residuals' non-normality (Ives, 2015; Warton et al., 2016). Since there are several days in the data set with zero number of infections, and logarithm zero is undefined, the dependent variable was increased by one to solve the algorithm zero issue (Liu et al., 2020).

The intercept of the model is α_0 , and IDA_RED is a dummy variable; it is zero from January 20, 2020 to February 23, 2020, otherwise one. In this study, we focused on the COVID-19 pandemic from the perspective of weather conditions, but other essential factors, like mobility, can affect the COVID-19 epidemic. We could not use this data in our model since the mobility data is not provided at the city level for South Korea by Google or Apple. Both companies solely present country-level data for South Korea. So, instead of using Google or Apple mobility data, the IDA_RED is a dummy variable used

to control mobility. IDA stands for Contagious Disease Alert, level 4 (red), which the government announced on February 23, 2020 for the whole country. It is zero from January 20, 2020 to February 23, 2020 and otherwise one. Based on KCDC, it caused a more than 10% decrease in mobility for the whole country on average after the government implemented this regulation, even taking into account that South Korea never expected strict limitations or lockdowns like Germany, Italy, and other countries.

At Infectious Disease Alert, level 4 (red), the government can restrict foreigners from entering the country and force all schools, universities, restaurants, bars, and all public facilities to cease operating. It also gives the government the ability to control public activities. The last time alert level 4 was announced was in 2009, to control influenza A (H1N1; Bloombrage, 2020). Finally, to control for day and city fixed effects, day_t and $City_i$ are added to the model (Amuakwa-Mensah et al., 2017).

3.6. Results:

The results indicate a negative relationship between minimum daily temperature and the daily number of COVID-19 confirmed cases, and the same pattern exists between the dependent variable and humidity. Most previous studies support the results about the negative relationship between temperature and respiratory diseases (Oliveiros et al., 2020; C. Wang et al., 2020; M. Wang et al., 2020), although there is no agreement about the effect of humidity. Some studies show low humidity aggravates asthma and allergic rhinitis (Mäkinen et al., 2009; Deng et al., 2020); others indicate that high humidity can increase the chance of infection by influenza and respiratory syncytial virus (RSV; Paynter, 2014; Freitas, 2009).

The reason behind these results relates to the survival of respiratory viruses (e.g., influenza, flu) in cold weather with very low humidity being close to 24 hours, whereas in high humidity and hot temperature it survives for less than one hour (Elert, 2013). Moreover, as shown in Table 5, a 1 °C rise in minimum daily temperature (lag0–7) led to a 0.88% decrease in the daily number of COVID-19 confirmed cases. For example, when the temperature decreased by 1 °C in a city with a population of 1 million, 880 people could get infected.

Furthermore, one should note that putting into effect the Infectious Disease Alert, level 4 (red) policy decreased the number of infections by 0.78%. Other studies confirm that government regulations lead to fewer infections worldwide, especially some implemented rules to decrease mobility, such as lockdown. For example, in 2020, Méndez-Arriaga showed that a Mexican lockdown phase 1 reduced the number of COVID-19 infections by more than 36%. Moreover, in 2020, Yilmazkuday implies that less mobility led to fewer COVID-19 disease cases and deaths in 130 countries.

Table 3- 4: The main regression results

	Model (1)	Model(2)	Model(3)
Variables	Coef		
Min_Temp	-0.68*	-0.81**	-0.88***
Humidity		-1.02***	-1.12***
Min_Wind		1.31**	1.49***
IDA_RED	- 0.51	- 0.69***	- 0.78***
Pressure			2.2***
R-sq.(adj)	0.57	0.64	0.68
Observations	18334	18334	18334
Notes: *** p<0.01, ** p<0.05, * p<0.1.			

Other findings show a negative relationship between relative humidity and the number of COVID-19 infections; a 1 °C increase in humidity caused a 1.12% decrease in risk of getting the disease. This phenomenon happened because the virus, which stays in the air, fades more quickly when the humidity is higher. In environments with faster virus reduction, fewer virus elements survive in the air, a condition associated with a decreased risk of infection (Wang et al., 2020).

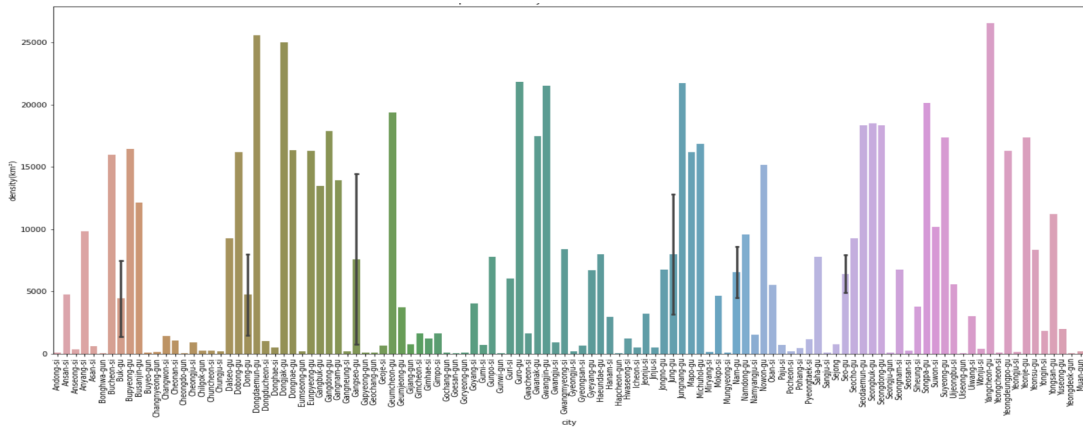
The results confirm that one-inch mercury (inHg) increases in pressure led to a 2.2% higher chance of COVID-19 infection. The outcomes further show a positive relationship between minimum wind speed (km/h) and positive coronavirus cases. With a 1 km/h upsurge in wind speed, COVID-19 infection risk rises by 1.49 %. Other studies confirmed these results in other parts of the world (see, e.g., Şahin, 2020; Casanova et al., 2010).

In summary, the finding is to some extent consistent with the previous outcomes of studies on influenza (see, e.g., Lowen et al., 2007; Shaman & Kohn, 2009) and SARS (see, e.g., Chan et al., 2011).

3.7. Robustness check:

For a robustness check, we ran the same regression on different country locations based on population density. Among all cities, the highest population density was in Yangcheon-gu in Seoul, with a population density of 26,553.00/km², and the lowest was in Bonghwa-gun in Gyeongsangbuk-do, with 27.02/km², as shown in Figure 7.

FIGURE 3. 7: population density for all cities in South Korea



The whole data set is divided into four even groups based on population density. The first quartile (0–25%) has the lowest population density, and the last quartile (75–100%) has the highest one. The results of these regressions are indicated in Table 6. The results show a negative relationship between the dependent variable and humidity in all four groups, with the lowest in the first quartile (0–25%). The strongest relationship, however, belongs to the last quartile (75–100%), with the highest population density.

Weather pressure and the number of positive COVID-19 cases have a positive relationship in all groups, but it is not statistically significant for the last quartile. The relationship between Min_temperature with a seven-day lag and dependent variable results support our findings, which indicate that a lower temperature causes a higher number of

infections. Higher wind speed also increases the number of positive cases for all four groups, like the results of the main regression.

Table 3- 5:four different quantiles regression results based on population density

Variables	The main regression	0%_25%	25%_50%	50%_75%	75%_100%
Humidity	-1.12***	-0.99***	-1.02**	-1.15***	- 1.20***
Pressure	2.2***	4.19***	3.12***	4.88*	2.64
Min_Temp	-0.88***	-0.53***	-0.48***	-0.83*	-0.90**
Min_Wind	1.49***	1.52***	0.11**	1.07*	0.98*
IDA_RED	-0.78***	-0.41*	-0.48***	-0.7*	-0.96***
R-sq.(adj)	0.67	0.58	0.58	0.67	0.49
Observations	18334	4532	4632	4635	4532

Notes: *** p<0.01, ** p<0.05, * p<0.1.

The same regression was applied for maximum and average temperature instead of minimum temperature to check the accuracy of the main regression. Similar processes for maximum and average wind speed are employed instead of minimum wind speed. The results are shown in Table 5. Thus, four regressions were executed for the first model, and the maximum temperature was adopted instead of minimum temperature. The results are the same for all variables when compared to the main regression. The maximum temperature shows a negative relationship with the dependent variable, as with the minimum temperature.

Model 2 employed the average temperature in the regression rather than minimum temperature, and the results for all other variables were nearly the same compared to the main regression. The relationship between the number of infections and the average temperature was negative.

In Model 3, maximum wind speed was replaced with minimum wind speed and applied in the regression. The results for all other variables are almost the same compared to the main regression. The relationship between the number of positive COVID-19 cases and the maximum wind speed is positive, as is the relationship between minimum wind speed and the number of infections.

In the last regression, the average wind speed was applied instead of minimum wind speed. We got almost the same results for this regression compared to the main regression, and based on these results, it can be concluded that higher average wind speed can increase the number of infections. Hence, all four models support the main results and show high-grade accuracy for this study.

Table 3- 6:Regression results for the robustness check

Variables	Model_1 (Max_Temp)	Model_2 (Avg_Temp)	Model_3 (Max_Wind)	Model_4 (Avg_Wind)
Humidity	- 0.11***	- 0.02**	- 0.11***	- 1.10***
Pressure	2.18***	1.98***	2.11*	2.30***
Min_Temp	N/A	N/A	- 0.45**	- 0.45**
Max_Temp	-0.52***	N/A	N/A	N/A
Avg_Temp	N/A	- 0.19**	N/A	N/A
Min_Wind	1.48***	1.48**	N/A	N/A
Max_Wind	N/A	N/A	0.16**	N/A
Ave_Wind	N/A	N/A	N/A	-0.96*
IDA_RED	- 0.84**	- 0.59**	- 0.65*	-0.74*
R-sq.(adj)	0.63	0.61	0.59	0.54
Observations	4532	4632	4635	4532

3.8. Prediction:

To check the accuracy of the model, the province of Gyeongsangbuk-do in South Korea was selected to predict the number of infections between May 2 and May 9, 2020,

and to compare the results with the actual number of positive COVID-19 cases during the same timeline for this province. According to Table 6, model accuracy is around 70%.

Table 3- 7:Gyeongsangbuk-do predicted number of COVID-19 positive cases

Date	Min-temp	Humidity	Min-Wind	Pressure	Predicted Number	Actual Number	Error	Accuracy
5/2/20	10.4	75.43	0.67	97.78	976	1266	290	77%
5/3/20	9.04	81.3	0.39	98.2	982	1360	378	72%
5/4/20	88	66.31	1.37	98.35	923	1350	427	68%
5/5/20	6	60.72	2.89	98.85	918	1261	343	73%
5/6/20	7.32	60.08	2.27	98.56	932	1192	260	78%
5/7/20	13.82	90.24	3.02	97.7	901	1336	435	67%
5/8/20	10.06	77.72	2.8	97.55	912	1366	454	67%
5/9/20	8.78	61.49	2.78	97.75	981	1266	285	77%

Additionally, to help governments make better decisions and become more efficient in responding to the pandemic, two cities with nearly the same population were selected: Detroit, Michigan, with 670,031 people in the midwestern United States, and El Paso, Texas with 681,728 people in the southern United States. The number of COVID-19 cases in the fall of 2020 were forecasted based on weather information from November 1, 2, and 3, 2019. Based on the results, the pandemic should be more challenging for Detroit than El Paso.

FIGURE 3. 8: Prediction based on the results



The results indicated that a 1 °C decrease in minimum daily temperature led to a higher chance of infections for Detroit; therefore, this kind of climate and geographical information can be used by governments to prioritize locations and make more effective decisions for contributing resources.

3.9. Conclusion:

To help governments fight the COVID-19 pandemic, this study tried to see how different weather conditions affected the number of COVID-19 confirmed cases in South Korea. The GAM was applied as the estimation model.

The analysis started by finding the correct number of lags for weather condition variables, since there is a delay between infection and symptoms. The right variables had to be selected to avoid multicollinearity. The outcomes prove that higher minimum temperature and higher relative humidity lead to a lower transmission rate.

Thus, based on the results during winter, the number of infections should be expected to speed up again. Therefore, governments need to be prepared and pass

regulations to reduce the COVID-19 infection rate. Moreover, minimum wind speed and atmospheric pressure have a negative relationship with the dependent variable. There are some limitations to this study, which are mentioned in the following section.

First, controlling city-level mobility can make results more accurate and help predict previous outcomes (Chang et al., 2021), but data on this are not available. Secondly, the results are conducted based on the specific period in South Korea, which means the outcomes were estimated under the hypothesis of no COVID-19 vaccine availability.

Finally, our results are described on statistical patterns, but random control experiments or epidemiological analysis are required to explore more accurate outcomes.

This study's results are aligned with previous studies on the transmission of SARS or other respiratory illnesses based on climate conditions (Tan et al., 2005; Lin et al., 2006; Liu et al., 2018; Park et al., 2020). Some researchers argue that global climate change is a main reason for the increasing pace of respiratory disease spread (Bezirtzoglou et al., 2011; Anwar et al., 2019; Casadevall, 2020), but this study focuses solely on weather condition indicators to determine how these indicators can affect the speed of COVID-19 transmission.

This research provides valuable results for decision-makers to efficiently combat the COVID-19 pandemic. The results may be notable not only to evaluate what occurred in the first stage of the COVID-19 pandemic in South Korea but also to recognize locations that require more attention to control the next phases of COVID-19 or future respiratory diseases. This study's outcomes show which climate indicators mostly affect the transition of the virus. The outcomes can be applied to increase general awareness and recognize

critical factors that affect the transmission of COVID-19. Furthermore, it is desirable to extend social distancing and lockdown until the vaccine becomes available or the temperature increases, which can reduce infection numbers.

Weather conditions are not the main determining factor in the spread of the COVID-19 virus. Government regulations and public awareness could all contribute to controlling the transmission speed. As temperature decreases in the fall and winter, the spread will probably speed up. Therefore, governments can control the COVID-19 pandemic with a dynamic policy.

It might be too soon to examine South Korea's regulative plan outcomes as regulations continue to develop and the battle against COVID-19 proceeds. Meanwhile, the South Korean COVID-19 regulative results can serve as an example to help other countries support their nations by passing various regulations. Although this study focuses only on the first phase of South Korea COVID-19 regulations, outcomes show that these regulations were notably effective against the COVID-19 pandemic.

Although this study concentrates on spread of COVID-19 in South Korea, the conceptual framework can be used for other cities or countries. The model can be applied to daily data since we fitted the model on the daily number of infections in South Korea.

3.10. References:

- Anderson, B. G., & Bell, M. L. (2009). Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States. *Epidemiology* (Cambridge, Mass.), 20(2), 205.
- Arabnia, H. R., & Tran, Q. N. (2015). *Emerging trends in computational biology, bioinformatics, and systems biology: algorithms and software tools*. Morgan Kaufmann.
- Barreca, A. I., & Shimshack, J. P. (2012). Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *American journal of epidemiology*, 176(suppl_7), S114-S122.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics*, 106(2), 407-443.
- Bashir, M. F., Ma, B., Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of the Total Environment*, 728, 138835.
- Bi, P., Wang, J., & Hiller, J. E. (2007). Weather: driving force behind the transmission of severe acute respiratory syndrome in China?. *Internal medicine journal*, 37(8), 550-554.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.
- Casanova, L. M., Jeon, S., Rutala, W. A., Weber, D. J., & Sobsey, M. D. (2010). Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Applied and environmental microbiology*, 76(9), 2712-2717.
- Chan, K. H., Peiris, J. M., Lam, S. Y., Poon, L. L. M., Yuen, K. Y., & Seto, W. H. (2011). The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Advances in virology*, 2011.
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840), 82-87.
- Cheng, J., Xu, Z., Bambrick, H., Su, H., Tong, S., & Hu, W. (2019). Impacts of heat, cold, and temperature variability on mortality in Australia, 2000–2009. *Science of The Total Environment*, 651, 2558-2565

- Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of personality assessment*, 91(2), 121-136.
- Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of statistics* (Vol. 30, pp. 351-413). Elsevier.
- Dignum, F., Dignum, V., Davidsson, P., Ghorbani, A., van der Hurk, M., Jensen, M., ... & Verhagen, H. (2020). Analysing the combined health, social and economic impacts of the coronavirus pandemic using agent-based social simulation. *Minds and Machines*, 30(2), 177-194.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin*, 118(3), 392.
- Hossain, M. S., Ahmed, S., & Uddin, M. J. (2021). Impact of weather on COVID-19 transmission in south Asian countries: An application of the ARIMAX model. *Science of The Total Environment*, 761, 143315.
- Ives, A. R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution*, 6(7), 828-835.
- Jayaraman, G. (2007). Air quality and respiratory health in Delhi. *Environmental monitoring and assessment*, 135(1), 313-325.
- John, S. D. (2009). „Supreme Emergencies, Epistemic Murkiness and Epistemic Transparency”. *Philosophy of Management*. 1 aprilie 2009.
- Kishore, N., Kahn, R., Martinez, P. P., De Salazar, P. M., Mahmud, A. S., & Buckee, C. O. (2021). Lockdowns result in changes in human mobility which may impact the epidemiologic dynamics of SARS-CoV-2. *Scientific reports*, 11(1), 1-12.
- Lu, H., Stratton, C. W., & Tang, Y. W. (2020). Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. *Journal of medical virology*, 92(4), 401-402.
- Lu, J. Y., Zhang, Z. B., He, Q., Ma, X. W., & Yang, Z. C. (2020). Association between climatic factors and varicella incidence in Guangzhou, Southern China, 2006–2018. *Science of The Total Environment*, 728, 138777.
- Méndez-Arriaga, F. (2020). The temperature and regional climate effects on communitarian COVID-19 contagion in Mexico throughout phase 1. *Science of the Total Environment*, 735, 139560.

- Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., & Tizzoni, M. (2020). COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific data*, 7(1), 1-7.
- Ross, M. L. (2001). *Timber booms and institutional breakdown in Southeast Asia*. Cambridge University Press.
- Şahin, M. (2020). Impact of weather on COVID-19 pandemic in Turkey. *Science of The Total Environment*, 728, 138810.
- Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C., Zhou, J., ... & Gao, G. F. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in microbiology*, 24(6), 490-502.
- van Delft, A., & Eichler, M. (2018). Locally stationary functional time series. *Electronic Journal of Statistics*, 12(1), 107-170.
- Wang, J., Tang, K., Feng, K., Li, X., Lv, W., Chen, K., & Wang, F. (2020). High temperature and high humidity reduce the transmission of COVID-19. arXiv preprint arXiv:2003.05003.
- Warton, D. I., Lyons, M., Stoklosa, J., & Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods in Ecology and Evolution*, 7(8), 882-890.
- Wielechowski, M., Czech, K., & Grzęda, Ł. (2020). Decline in Mobility: Public Transport in Poland in the time of the COVID-19 Pandemic. *Economies*, 8(4), 78.
- Wong, L. Y. R., Lui, P. Y., & Jin, D. Y. (2016). A molecular arms race between host innate antiviral response and emerging human coronaviruses. *Virologica Sinica*, 31(1), 12-23.
- Yang, Z., Wang, X., Chi, X., Zhao, F., Guo, J., Ma, P., ... & Long, G. (2017). Correction for Yang et al., "Neglected but Important Role of Apolipoprotein E Exchange in Hepatitis C Virus Infection". *Journal of virology*, 91(24).
- Ye, Z. W., Yuan, S., Yuen, K. S., Fung, S. Y., Chan, C. P., & Jin, D. Y. (2020). Zoonotic origins of human coronaviruses. *International journal of biological sciences*, 16(10), 1686.
- Yilmazkuday, H. (2021). Stay-at-home works to fight against COVID-19: international evidence from Google mobility data. *Journal of Human Behavior in the Social Environment*, 31(1-4), 210-220.

- Yuan, J., Yun, H., Lan, W., Wang, W., Sullivan, S. G., Jia, S., & Bittles, A. H. (2006). A climatologic investigation of the SARS-CoV outbreak in Beijing, China. *American journal of infection control*, 34(4), 234-236.
- Zhang, C., Liao, H., Strobl, E., Li, H., Li, R., Jensen, S. S., & Zhang, Y. (2021). The role of weather conditions in COVID-19 transmission: A study of a global panel of 1236 regions. *Journal of cleaner production*, 292, 125987.
- Zhang, Z., Werner, F., Cho, H. M., Wind, G., Platnick, S., Ackerman, A. S., ... & Meyer, K. (2017, February). A framework for quantifying the impacts of sub-pixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bi-spectral method. In *AIP Conference Proceedings* (Vol. 1810, No. 1, p. 030002). AIP Publishing LLC.
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, 579(7798), 270-273.

Chapter 4

Small Businesses' Loan Default Rate Under 2017 Tax Regulation

4.1. Introduction

Small businesses play a crucial role in the U.S. economy. In 2010 alone, small businesses in the U.S numbered more than 27 million and generated around 50% of the U.S. gross domestic product (GDP; Office of Advocacy, 2010). Small businesses contribute strongly to the economy and socio-economic development sectors. Usually, they create jobs, innovate, and provide solutions to attain financial prosperity and independence for all people. Small businesses also facilitate economic growth by creating employment possibilities for people whom large corporations may not hire (Graham, 1999).

Access to capital is a crucial concern for small businesses and start-ups. Without adequate capital, small businesses cannot produce new goods or services or expand to meet demand. Liquidity deficiency is the main reason for small businesses' bankruptcy (Coleman, 2000). Unlike large corporations, small businesses and start-ups typically are not admitted to the traditional capital market. On the other hand, small businesses depend profoundly on financial institutions' loans, home equity loans, and loans from family and friends. Banks are vital external sources for accessing capital (Watson et al., 1998). Also, from the lenders' viewpoint, the primary purpose of lending money is capital return, ideally within a desirable time frame and with high interest on capital. Generally, financial institutions try to secure their money by choosing safe borrowers with a degree of longevity

or with assets as collateral. Many studies in this area have indicated the essential features of safe borrowers (Chu et al., 2017).

Governments play the most critical role in this area, and their decisions have a profound influence on financial markets. Governments try to balance the demand and supply sides to control the market correctly (Grody et al., 2018). In particular, in all financial markets, governments play a significant role in managing risk and support the economy by implementing various regulations, such as decreasing tax or offering low-interest loans (Sanfelici & Halbert, 2018).

In this regard, in 2017, the government implemented the new Tax Cuts and Jobs Act (TCJA) to support the economy. Based on this regulation, the business tax was reduced from 35% to 15%. This study evaluates how the 2017 TCJA affected the small business loan default rate by applying the logistic regression method on a data set from Lending Club. This research focuses on small businesses' loans since activity in this segment is one of the foremost goals of policymakers and shedding light on this area would help policymakers improve their policies on small businesses (Nowak et al., 2015). We tried to focus primarily on two features of the borrowers' characteristics and loans. Examples of borrower features are annual income, average debt-to-income ratio, or homeownership status. Loan features include the number of applicants, maturity date, and interest rate.

More than 100 variables are provided by Lending Club and selecting the correct number of variables was the first challenge in this study. Different methods, such as the Least Absolute Shrinkage and Selection Operator (LASSO), were applied to avoid multicollinearity and to select the proper variables. LASSO is a method for determining and shrinking by minimizing squared errors in linear regression models (Tibshirani, 1996).

This method also appears to be the most suitable technique for variable selection and estimation of the probability of small businesses' default rate (see Tian et al., 2015). Finally, this research examines how the 2017 TCJA changed the small businesses' loan default date. To achieve this goal, the dummy variable was created to assess the effects of the 2017 tax cut regulation. Lastly, logistic regression was applied to estimate the model since the dependent variable was binary in this study (zero if an applicant charged-off loan, otherwise one).

The results show that the 2017 tax cut regulation helped small businesses decrease defaulting on their loans. In addition, the outcomes of this research specified some essential features of borrowers and loans for determining the chance of default. For example, applicants with longer-term loans (60 months versus 36 months) present a higher probability of default, as do applicants with higher debt-to-income ratios.

This study contributes to the existing literature in three significant directions. Firstly, it tries to connect two different literature angles by applying borrowers' features and government regulation in the same regression as the right-hand side variables. Previous studies have only attempted to explore the borrowers' features that can be essential in forecasting default rates (see, e.g., Davis et al., 2016; Kumar et al., 2016; Jagtiani & Lemieux, 2018). Governments usually play the most crucial role in the financial market. As the primary regulator, the government can affect saving, lending, and risk hedging in the financial market. So, ignoring government regulation in this research area might make the results inaccurate (Kevin, 2010).

Including more risk factors than did previous studies is the second contribution of this study. Lending Club usually updates its data set yearly and provides more information

about loan applicants. The availability of more statistical information can enable us to analyze more independent variables compared to previous studies.

Lastly, this study applied three variable-selection methods to select the correct number of dependent variables. Most previous studies either did not apply any models or only used one model to select variables. These methods include cross-validation, the adaptive LASSO model, and the BIC. Only variables were selected for this study, which were indicated by all three methods at the same time.

The rest of the paper is designed as follows. Section 2 reviews the previous literature, and Section 3 explains the data collection, data set features, descriptive statistics, and variable-selection methods. Section 4 reviews the empirical approach and outcomes. Section 5 discusses the results, and, finally, Section 6 present the conclusion of this study.

4.2.Literature Review

On December 22, 2017, the TCJA was announced. This regulation was the most comprehensive update of U.S. tax law since the Tax Reform Act of 1986 (Auerbach, 2018). This law made several adjustments to business income tax. Some important features are:

- The business tax was reduced from 35% to 15%;
- The corporate alternative minimum tax was eliminated;
- This allowed firms engaged in manufacturing in the United States to choose between the full expensing of capital investment and the deductibility of interest paid (Cole, 2016).

Many researchers criticized this law, calling it a reward to firms, and believed it would increase inequality; advocates of the new bill, however, maintained that the regulation would expand the economy to balance the drop in tax declarations (Wagner et

al., 2018). Based on the Congressional Budget Office (CBO) report of April 2018, the tax cut regulation led personal income tax to shrink by \$65 billion. The corporate tax also declined by \$94 billion, and around \$3 billion declined in other tax categories. Other outcomes of the tax cut regulation include:

- Production and investment moved upward.
- An extension of the debt to GDP ratio.
- A lowered possibility of conversions (relocating companies headquartered in the United States abroad (CBO, 2018)).

The TCJA regulations had a remarkable effect on small businesses. It caused a reduction in personal income tax; hence, small businesses could benefit from this bill. For a small business, such as sole proprietorship, LLC, or partnership with entity pass-through operation, a significant change would reduce the individual income tax rate (Rogers, 2019).

In addition, small businesses are now allowed to decrease 20% of their qualified business income, which means that the revenue generated by small business can be deducted from any applicable costs. All these effects led to a higher cash-to-debt ratio for corporations and small businesses. Therefore, under the TCJA, corporate and small business debt decreased by 7–9% (Divino et al., 2021). Besides some regulations and bills that help borrowers decrease their default rates, other factors also matter for lenders to reduce the probability of borrower default.

Typically, providing a loan to a small business is riskier and more costly for lenders since they have a higher default rate and have fewer defenses against downward trends in the economy compared to corporations (Mach et al., 2014). There is a vast body of literature on risk management and recommendations and on avoiding bankruptcy or

decreasing the risk of default. The main differences among these studies are due to using different analytical methods and critical variables. Statistical models are prevalent in this matter. Most studies have tried to determine a causal relationship between the loan status and other essential variables on the right-hand side of the regression.

For example, Chen (2017) estimated the probability of default for a different group of borrowers, such as small business, educational, medical, and debt consolidation using the data set of Paipai (a Chinese online peer-to-peer [P2P] lending borrower) and applying logistic regression. The results showed that the debt consolidation group had a higher risk of default than other groups. Djurovič (2017) examined the relationship between default rate and loan features, including both the purpose of the loan and the loan duration (36 and 72 months). The findings imply a strong relationship between default rate and loan characteristics. In addition, loans of 72 months' duration were found to be much riskier than those of 36 months.

Serrano-Cinca et al. (2015) studied factors defining loan default and P2P lending. They employed 24,449 applicants' data from Lending Club between 2008 and 2015. Their findings indicated that factors like candidate income, homeownership status, and credit score play a critical role in estimating the probability of default.

Croux et al. (2020) estimated the risk of default among different groups and examined the essential features by employing logistic regression. Their results showed that occupation and homeownership can play a critical role in borrower risk. Their findings also showed that small business funding and medical financing loans have the highest default rate.

Artificial Neural Network (ANN) classification methods have attracted the attention of researchers in this area to estimate the probability of default (PD). In 2017, Bequé and Stefan (2017) applied different algorithms of machine learning, such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and logistic regression models to estimate the probability of default. The results suggest that logistic regression is a reliable and accurate prediction model. Jin and Zhu (2015) tried to forecast the performance of P2P loans by adopting five machine learning techniques—two decision trees (DTs), two neural networks (NNs), and one support vector machine (SVM)—and used two metrics. The study was done on the Lending Club data set, and their findings showed that the loan maturity, debt-to-income ratio, annual self-reported income, loan amount, and credit score play an essential role in estimating the probability of default. All five models showed almost equal prediction accuracy.

Others (e.g., Wagner et al., 2018) aimed to study stock market performance since the TCJA. Corporate tax cuts caused an upward trend for the stock market, resulting in an additional return of around 15% on the market.

All of the mentioned studies studied either how borrowers' profile and loan features affect default rate or how the TCJA affected different aspects of the economy; this study, however, connects these two approaches by determining how the TCJA has affected small businesses' default rate. Moreover, it predicts what factors matter most in predicting the loan repayment performance of lenders.

This study aims to evaluate how the TCJA affected small businesses' loan default rate by applying logistic regression method on the Lending Club data set. This research focuses on small businesses' loans since activity in this segment is one of the foremost

goals of policymakers and shedding light on this area would help policymakers improve their policies about small businesses (Nowak et al., 2015).

Small and medium-sized firms play a crucial role in both developing and developed countries' economies. Since 1989, 93.5% of the net U.S. new jobs have been created by small businesses (Spence & Rutherford., 2003). Thus, helping sustain and grow small businesses is a win-win condition for the government. Small and mid-size companies require funding to start up, stay open, or expand. Based on a recent survey by the National Small Business Association, about 71% of participants believe that lack of liquidity is their main obstacle to growth. They also believe that a lack of capital for prevents them from expanding or causes them to close (Baker et al., 2021). Usually, governments try to pass laws and provide some easing situations to help businesses. These regulations range from granting low-interest loans to businesses or reducing tax rates so that businesses can update their equipment, enhance their energy efficiency, or make any other improvements (Vismara,2016). The role of governments in the financial market, however, is a long-standing debate among economists worldwide. One risks that threatens companies is low liquidity. Failure to repay loans can make borrowing money more difficult for borrowers because lenders either will not lend at all or will lend only at a very high interest rate to these types of applicants (Brock & Evans, 1989).

Additionally, lenders need to decrease the risk of default as much as they can to survive. Since corporations and small businesses are the main customers of financial institutions, this kind of supportive law can be crucial for both the supply and demand sides of these situations (Sher, 2016).

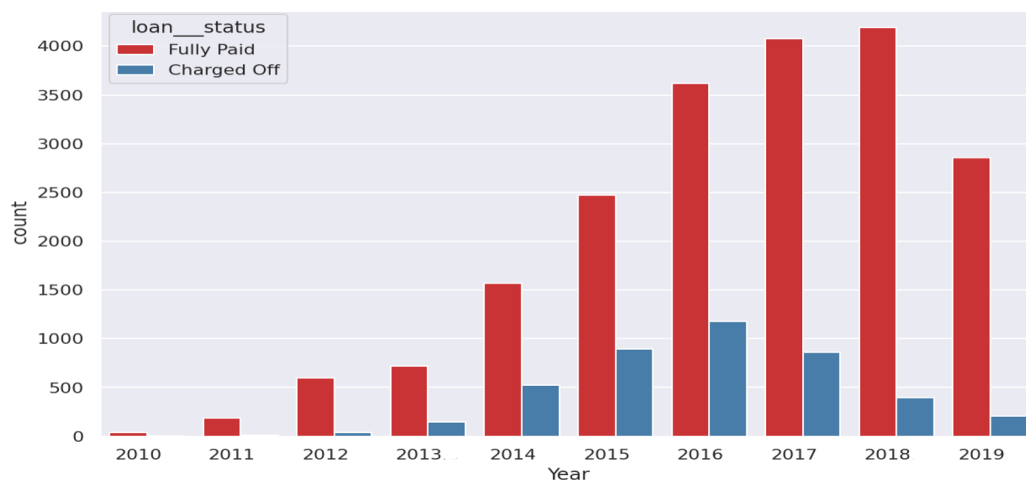
4.3. The data and descriptive statistics

4.3.1. The data sources

The Lending Club data include 24,576 personal loan candidates' information with a 36- or 72-month loan duration. This data set is public and includes various aspects of applicants. Only the loans with clear ending commitments were selected for this study. Therefore, the loans must have been either repaid in full or charged off to be included in our analysis. In essence, this information contains borrowers' profiles and loan features.

The purpose of these loans is to provide financing to small businesses with different ownership types, individual applicants, or joint applications with two co-borrowers. Almost 92% of loans were requested individually and the rest with co-borrowers. The data were posted on the public website by Lending Club, which has information from 2010 to 2019 with monthly updates for every loan payment. Although the data begin in 2010, only 1% of the data belongs to this year. Around 50% of data relates to 2016, 2017, and 2018. Figure 1 shows the number of observations every year, including applicants' loan status.

FIGURE 4. 1: Loan status from 2010 to 2019



Note: The figure shows the number of applicants based on their loan status per year.

The data provide information about borrowers' profiles, such as annual income and employment length; other information about the loan features, like interest rate, issue date, and last payment date are also included in this data set.

Table 1 represents the sample, which includes 24,576 personal loans from Lending Club granted in 2010–2019. The table describes two main parts: first, loan information, including the number of loans, percentage of charge-off, and amount paid off during this study. The second part shows some critical features of applicants. Based on the below Table, 16.69% of personal loans between 2010 and 2019 were defaulted on by applicants. Around 70% of the granted loans had three years of maturity, and the rest had five years of maturity.

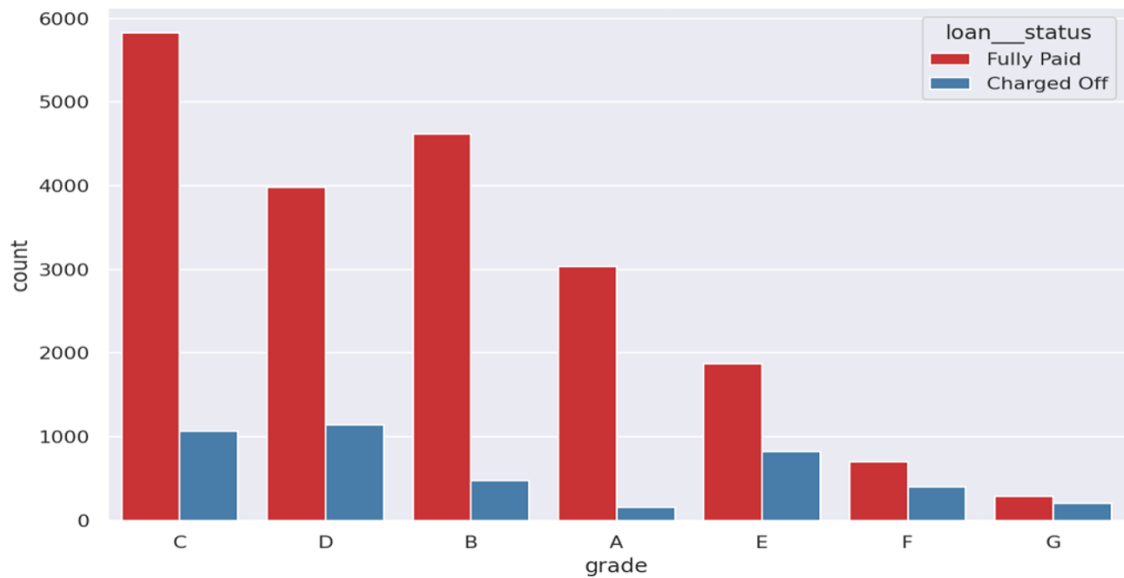
Table 4- 1: Loan dataset detailing:

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
<i>Loan information</i>											
Number of applicants	44	199	634	875	2088	3065	3362	4583	4795	4931	24576
Percentage %	0.2	0.8	2.6	3.5	8.5	12.5	13.7	18.6	19.5	20.1	100
Number of Charged off	6	14	39	149	518	889	1155	817	349	162	4098
Percentage %	0.1	0.3	1	3.6	12.6	21.7	28.2	19.9	8.5	4	16.7
Number of paid off	38	185	595	722	1570	2475	3635	4121	4234	2900	20475
Percentage %	0.2	0.9	2.9	3.5	7.7	12.1	17.8	20.1	20.7	14.2	83.3
The number of payments on the loan: 36 months	1	6	512	657	1501	2428	3011	3692	3306	2269	17383
Percentage %	2.3	3	80.8	75.1	71.9	79.2	89.6	80.6	68.9	46	70.7
The number of payments on the loan: 72 months	43	193	122	218	587	637	351	891	1489	2662	7193
Percentage %	97.7	97	19.2	24.9	28.1	20.8	10.4	19.4	31.1	54	29.3
<i>Applicants profile information</i>											
Average of annual income provided by the borrower in \$ thousand	82.4	75.8	81.1	93	95.7	96.8	95.1	96.9	98.9	111.4	97.9
Average of debt-to-income ratio provided by the borrower	10	11.7	13.1	14.3	14.8	15.6	15.6	15.5	15.9	17.4	15.7
Homeownership status: RENT	17	84	276	382	949	1212	766	1886	2328	3149	11417
Percentage %	0.1	0.7	2.4	3.3	8.3	10.6	6.7	16.5	20.4	27.6	46.5
Homeownership status: MORTGAGE	27	102	317	411	926	1460	1961	2092	1953	1414	10663
Percentage %	0.3	1	3	3.9	8.7	13.7	18.4	19.6	18.3	13.3	43.4
Homeownership status: OWN	0	13	41	82	213	393	635	605	514	368	2496
Percentage %	0	0.5	1.6	3.3	8.5	15.7	25.4	24.2	20.6	14.7	10.2

4.3.2. The sample summary statistics

Lending Club categorizes borrowers from A to G to estimate the risk of default by borrowers. Applicants with the lowest risk of default are categorized as Group A, and applicants with the highest risk of charge-off are categorized as Group G. Figure 2 shows the number of charge-offs and fully paid loans based on the specific categorical variable. Applicants with an LC grade of C have the highest number of observations; around 28% of applicants are assigned to Group C. The second and third rank belongs to Groups D and B, respectively. Therefore, applicants in Group C rank first in fully paid loans. The lowest number of charge-offs belongs to applicants in Group A, is 3% of all reported numbers of charge-offs.

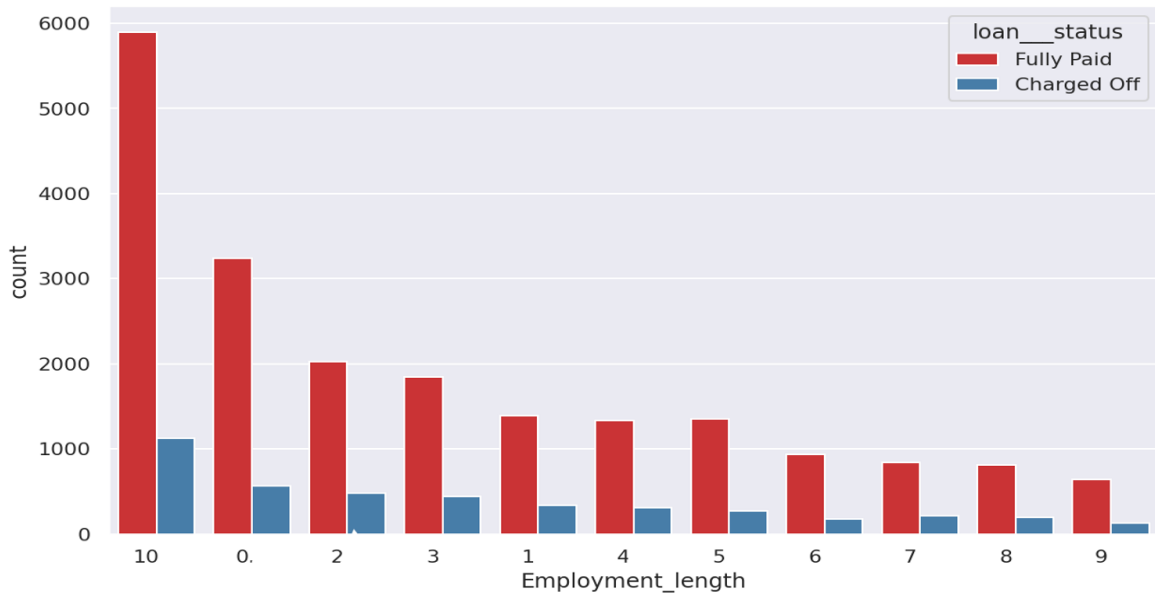
FIGURE 4. 2: Loan status based on loan grades



Note: This figure shows the number of applicants with different LC grade based on their loan status.

Around 28% of the observations belonged to applicants with an employment duration of ten years or more. Candidates with less than one year of employment ranked second, with an observation rate of 15%. The lowest number of observations belonged to borrowers with nine years' employment—the charged-off rate in this category was 11%. The highest rate of charged-off among all categories belonged to applicants with seven years' employment, and the lowest rank went to candidates with less than a year's employment length, equal to 4% charged-off.

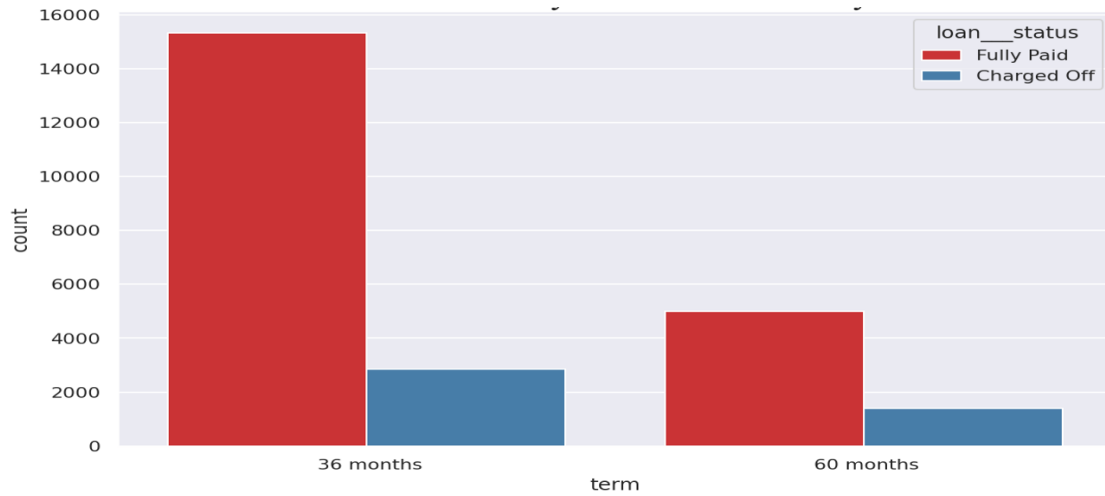
FIGURE 4. 3: Loan status based on employment length



Note: This figure shows the number of applicants with different employment length based on their loan status

Around 74% of personal loans were granted with a maturity date of 36 months, from which about 15% were charged-off by borrowers. On the other hand, loans with a 72-month maturity date had a 22% default rate. Figure 4 details this.

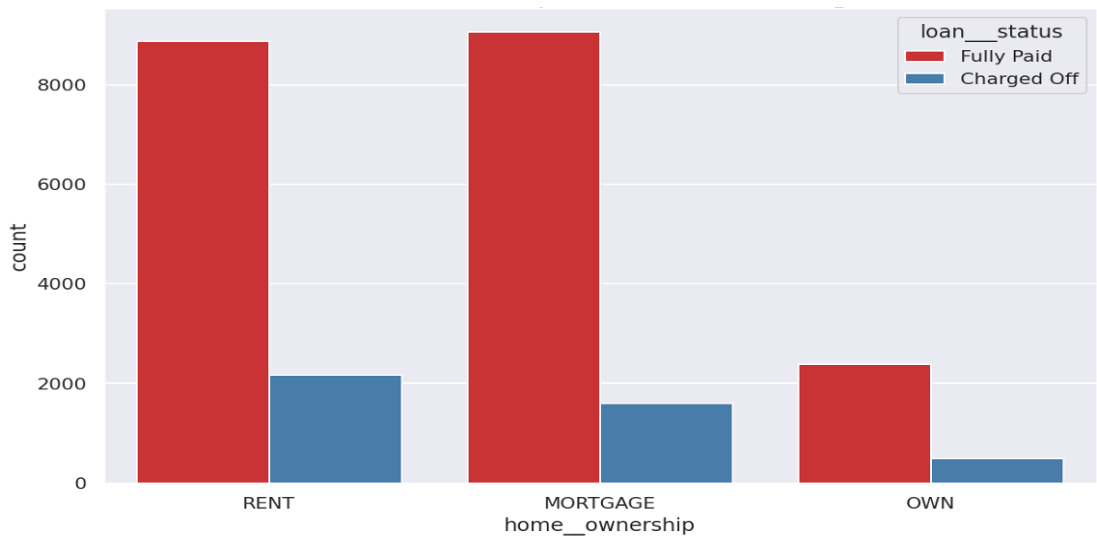
FIGURE 4. 4: Loan status based on account maturity



Note: This figure shows the number of applicants with different loan maturity dates based on their loan status.

Most applicant's homeownership status was rental: 45% of all borrowers, compared to 43% mortgage based, and 12% outright home ownership. Among these categories, the candidates with mortgage homeownership status had the highest percentage of charged-off (15%).

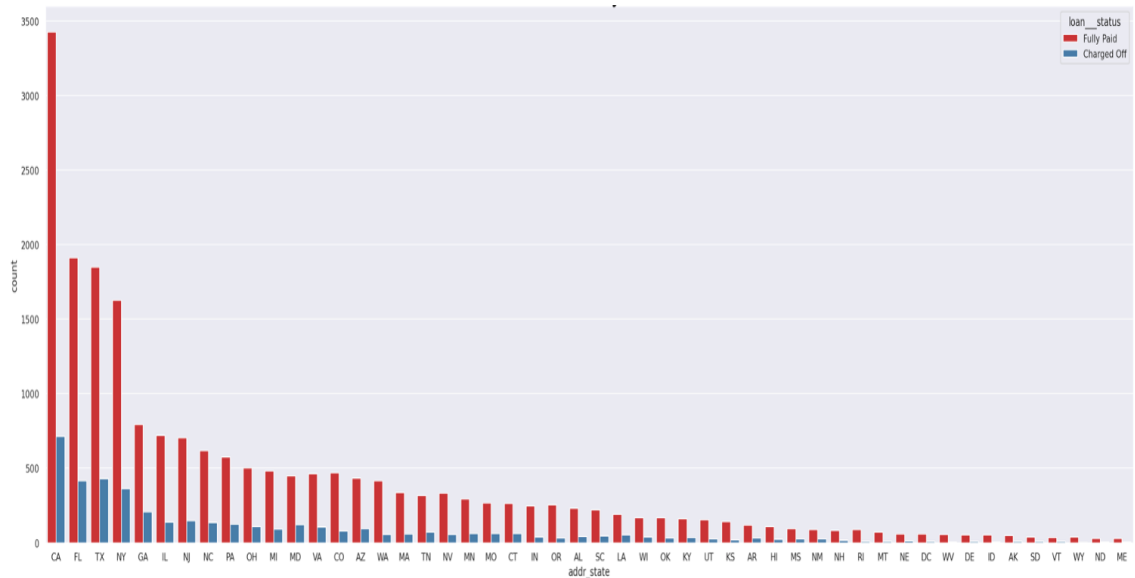
FIGURE 4. 5: Loan status based on home ownership status



Note: This figure shows the number of applicants with different home ownership status based on their loan status

The largest number (42%) of borrowers were in California, Florida, Texas, and New York. Among all states, California had the highest rank, with 16% of applicants. Florida and Texas were second, each with 9% of applicants, and New York was third, with 8% of borrowers.

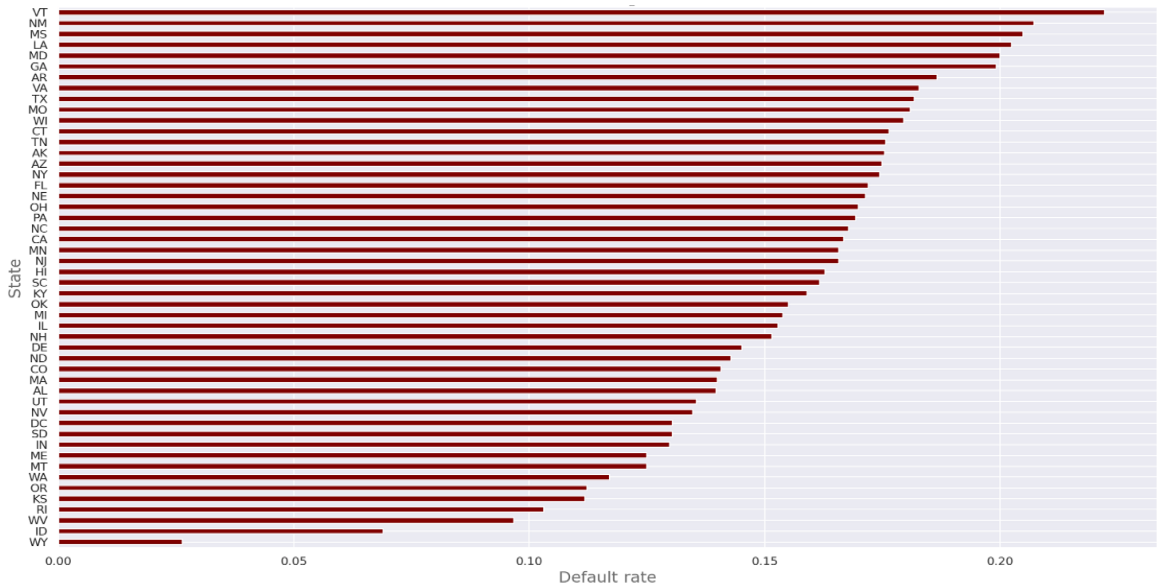
FIGURE 4. 6: Loan status based on applicants' locations



Note: This figure shows the number of borrowers locations based on their loan status.

Although the largest number of applicants was in California, it ranked 22 among all states, with a 16% charged-off rate. The highest default rate belonged to Vermont, with a 22% default rate, then New Mexico, with 20%. The lowest rate of loan default was in Wyoming, with a 2% charged-off ratio. More details are given in Figure 7.

FIGURE 4. 7:Ratio of loan default per status



Note: This figure shows the ratio of charged off per states.

4.3.3. Variable selection

The original data set contained 85 variables, but most variables had many missing values. Many variables may make multicollinearity in the regression, which can cause incorrect regression parameterization, reduced analytical power, and the elimination of significant predictor variables through model formulation (Farrar et al., 2021). Three methods were applied for variable selection and resolution of the multicollinearity problem, including LASSO (see, e.g., Meier et al., 2008; and Belloni et al., 2016), Cross-validation (see, e.g., Baumann et al., 2002; Jouan-Rimbaud et al., 1996) and BIC (see, e.g., Hocking et al., 1976; Hosmer et al., 2003). In this study, only those variables were selected that were suggested by all three methods. Thus, only 14 of the 85 variables were chosen.

Table 4- 2:selected variables:

Variable	CV	MinBIC	Adaptive	Description
Loan_amnt	✓	✓	✓	The listed amount of the loan applied for by the borrowers
term	✓	✓	✓	The number of payments on the loan. Values are in months and can be either 36 or 60.
grade	✓	✓	✓	LC assigned loan grade.
emp_length	✓	✓	✓	Employment length in years. Possible values are between 0 and 10, where 0 means less than one year and ten year means or more years.
home_ownership	✓	✓	✓	The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are RENT, Own or Mortgage
annual_inc	✓	✓	✓	The self-reported annual income provided by the borrower during registration.
dti	✓	✓	✓	Debt to income ratio.
open_acc	✓	✓	✓	The number of open credit lines in the borrower's credit file.
total_acc	✓	✓	✓	The total number of open credit lines in the borrower's credit file.
total_pymnt	✓	✓	✓	Payment received to date for total amount funded.
last_fico_range_avg	✓	✓	✓	The average score the borrower's last FICO pulled belongs to.
application_type	✓	✓	✓	Indicates whether the loan is an individual application or a joint application with two co-borrowers.
tax_lines	✓	✓	✓	Number of tax lines.
mths_since_earliest_credit_line	✓	✓	✓	The month the borrower's earliest reported credit line was opened.
Number of observations	2476			

Note: CV stand for cross validation mean prediction error, and Adaptive stand for adaptive lasso model.

Using the LASSO method has the benefit of selecting only the essential variables, shrinking regression coefficients, solving the multicollinearity problem, and hence making the results more straightforward (Meinshausen & Yu, 2009).

Table 4- 3:data description:

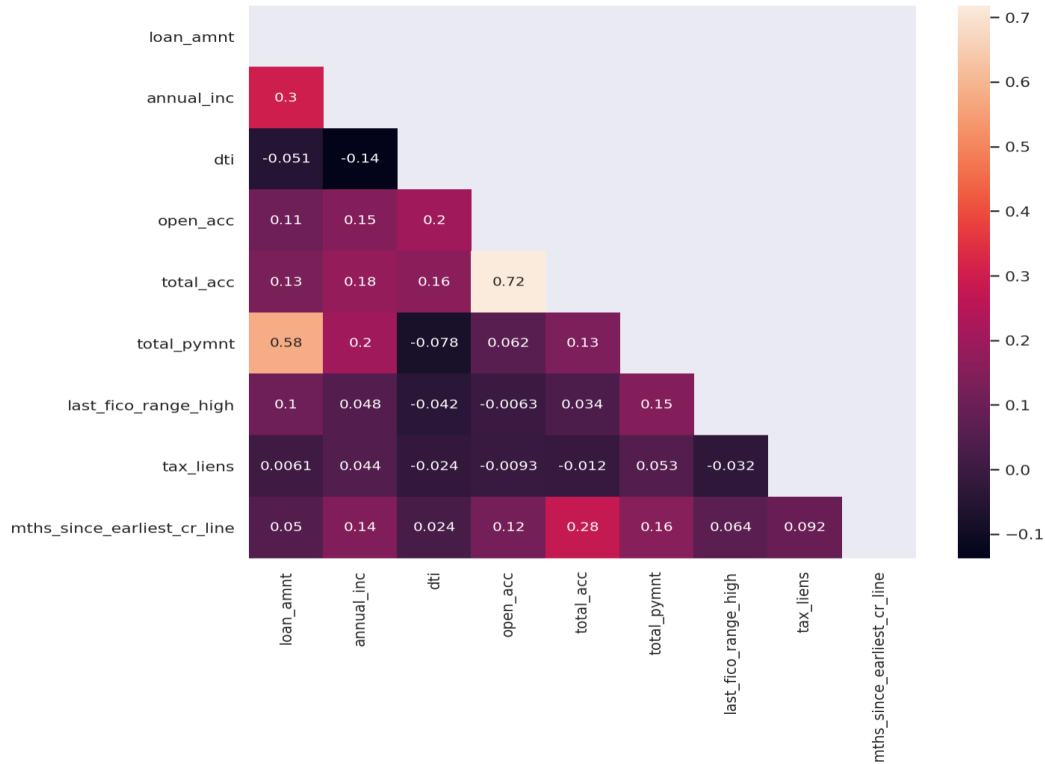
	loan_amnt	Annual_inc	dti	Open_acc	Total_acc	Fico_aveg*	Total_pymnt	Tax_line	mths**
mean	16994.06	97861.06	15.69	10.5	12132.39	677.51	12132.39	0.11	240.31
Std	10596.11	97694.66	14.53	5.4	10743.3	73.94	10743.3	0.67	96.52
min	1000	0	0	0	0	0	0	0	55
25%	8300	53000	8.43	7	4229.51	644	4229.51	0	176
50%	15000	78500	14.32	9	8634.78	689	8634.78	0	223
75%	24500	120000	13	13	16814.79	724	16814.79	0	288
max	40000	7272644	71	71	63296.88	850	63296.88	39	875

Note: * last_fico_range_aveg. , ** mths_since_earliest_cr_line

Note: This table shows summary statistic of some selected variables

According to Table 3, loans ranging from \$1,000 to \$40,000 were granted to small business applicants. On average, the applicants' annual income was around \$97,000, and their average debt-to-income ratio was 10.50. The average FICO score for loan applicants was 677, which is considered a "good" score based on the FICO category scoring system. Figure 8 shows the correlation among the selected variables, so there is no concern about the multicollinearity problem in the regression.

FIGURE 4. 8: Selected variables correlation



4.4. Empirical Method and Results

To establish a baseline for more tests, the probability of default by small businesses was estimated. The dependent variable is a dummy variable in the sense that it is one if the borrowers fully pay off the loan, and otherwise zero. Zero means the applicants could not make the full payment by the due date, and therefore the loan was considered charged off. Based on the Lending Club announcement, some loans might have been adjusted due to extension or early payment. Only personal loans with a clear ending date were selected for this study to avoid reducing the prediction accuracy.

Previous studies have focused on borrowers' features and loans' characteristics for estimating loan default rates; the government's role in the financial market, however, is

inevitable, as indicated in the introduction. The TCJA regulation was selected as the primary right-hand side variable in this study to examine the government’s role in addition to the borrower characteristics and loan features in estimating the risk of default. This regulation was applied in the regression as the dummy variable; it is equal to zero before December 2017 and otherwise one.

4.4.1. Logistic Regression Analysis

Logistic regression is the method of estimating the likelihood of discrete response variables. It is mainly applied when the dependent variable takes two different values: e.g., yes/no or true/false. Logistic regression does not need to apply various critical assumptions of ordinary least squares algorithms (OLS), especially normality, linearity, measurement level, or homoscedasticity (Wright,1995).

Regression 1 was applied to this study as the primary estimation method. More control variables were joined to the regression for every model (Models 2–5 in Table 4) to get more accurate results. For example, in Model 2, besides tax_cut, other control variables were added to Regression 1, including loan_amnt, dti, and total_acc. This caused the adjusted R-squared measure to increase by 7%.

$$\text{Loan Status}_{it} = \beta_0 + \beta_1 \text{Tax cut}_t + \delta_{it} + \text{id} * \text{issue_date}_{it} + \varepsilon_{it} \quad (1)$$

Regression 1 was applied for this study where:

Loan status: The loan’s current status equaled one for borrowers who had fully paid off the loan; otherwise, zero.

Tax cut: Tax cut equals zero if the loan’s issue date was before December 22, 2017; otherwise, it equals one.

Id: A unique Lending Club-assigned ID for the loan listing.

Issue_date: The month the loan was funded.

β_0 : the intercept; the regression coefficient.

δ_{it} : the symbol for control variables, which is equal to zero for Models 1–3 in Table 4.

For Regressions 4 and 5, in addition to the indicated variables in Table 4, annual_inc, open_acc, tax_lines, mths_since_earliest_cr_line, emp_length, and applications_type were also added to the regression; doing so caused an increase of around 9% in the R-squared model compared to Regression 3. In Regression 5, total_pymnt was also added compared to Regression 4, and R squared increased by 4%.

Table 4- 4:Regression results:

	1	2	3	4	5
	Loan Status	Loan Status	Loan Status	Loan Status	Loan Status
tax_cut	0.313*** -0.0113	0.313*** -0.0113	0.295*** -0.0111	0.191*** -0.00901	0.137*** -0.00868
loan_amnt		-0.243*** -0.000361	-0.207*** -0.000355	-0.368*** -0.000285	-0.126*** -0.000385
dti		-0.141*** -0.000235	-0.108*** -0.000231	-0.0682*** -0.000185	-0.0449* -0.000175
total_acc		0.127*** -0.000342	0.0434* -0.000339	-0.0261* -0.000271	-0.0496* -0.000257
grade			-0.513*** -0.0029	-0.0204 -0.00242	-0.128*** -0.00231
last_fico_range_avg				0.0242*** -0.0000339	0.0208*** -0.0000338
total_pymnt					0.150*** -0.000463
_cons	0.597*** -0.00886	0.634*** -0.013	0.816*** -0.0164	-0.845*** -0.0268	-0.563*** -0.0268
N	24576	24576	24576	24576	24576
R-sq	0.42	0.49	0.56	0.645	0.682
Std errors in parentheses	***p<0.001 ** p<0.01 * p<0.05				

Note: in model 4 and 5, annual_inc, open_acc, tax_lines, mths_since_earliest_cr_line, emp_length and applications_type (Y/N) were applidd as control variables.

Based on the R-squared value, Model 5 shows the most reliable results. The results of this model show that the TCJA helped small businesses avoid charged-off by

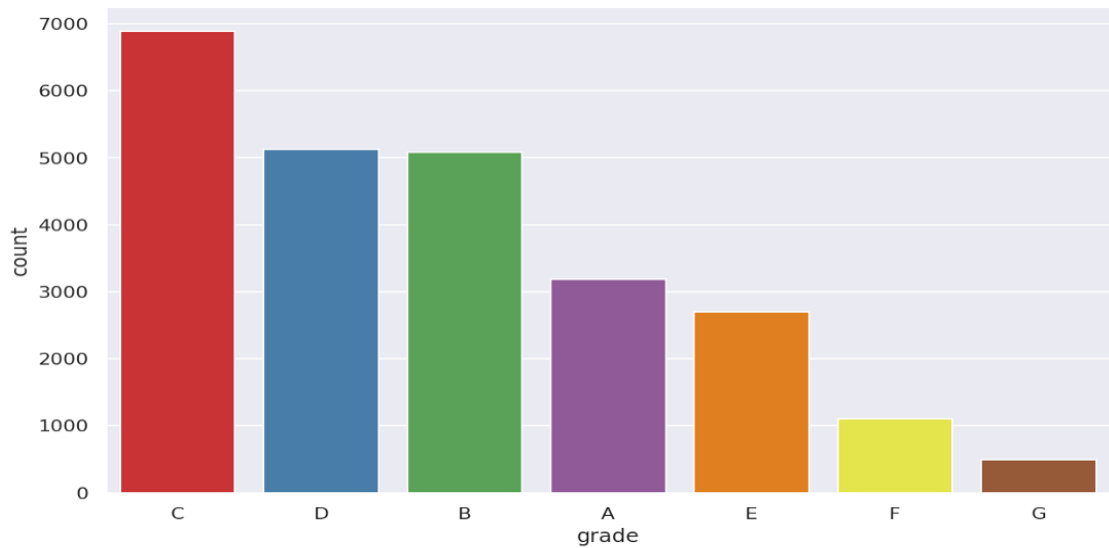
13.7%, and it is statistically significant. Higher loan amounts increased the risk of default for lenders by 12%. The results indicate that the chance of paying off the loan declines by 4% with a higher debt-to-income ratio. Granting a loan to applicants with a greater number of credit lines (`total_acc`) makes loans 4.9% riskier.

Based on candidates' applications, Lending Club categorized them into seven categories (A to G), with Group A having the lowest risk of charged-off and Group G having the highest risk. Figure 9 shows more details on this. The results prove that a higher grade, from A to G, raised the chance of default by 12%. So, Group B had a 12% higher chance of default compared to Group A.

A higher FICO score indicates a 2% lower chance of default. The regression outcome indicates 15% percent less probability of default for borrowers with more payments received to date for the total amount funded. Furthermore, `annual_inc` shows a 9.7% lower chance of `charged_off`. Still, the outcome is not statistically significant, as `open_acc` has a positive correlation with the dependent variable value of 0.11 and is statistically significant, with P-values of less than 0.01.

The correlation between `tax_lines` and dependent variables is -0.00489, but it is not statistically significant, and `mths_since_earliest_cr_line` has a value of -0.0553, with P-values of less than 0.01. Higher `emp_length` caused a 6.5% lower chance of charged-off. Still, it is not statistically significant, and the `applications_type` (Y/N) outcome exposes how individual borrowers have a 16.% higher probability rate of default, although this outcome is not statistically significant.

FIGURE 4. 9: Number of borrowers based on loan grades



The number of applicants in each grade category, borrowers in Group G, have the highest risk of default, and applicants in Group A show the lowest risk of charged off.

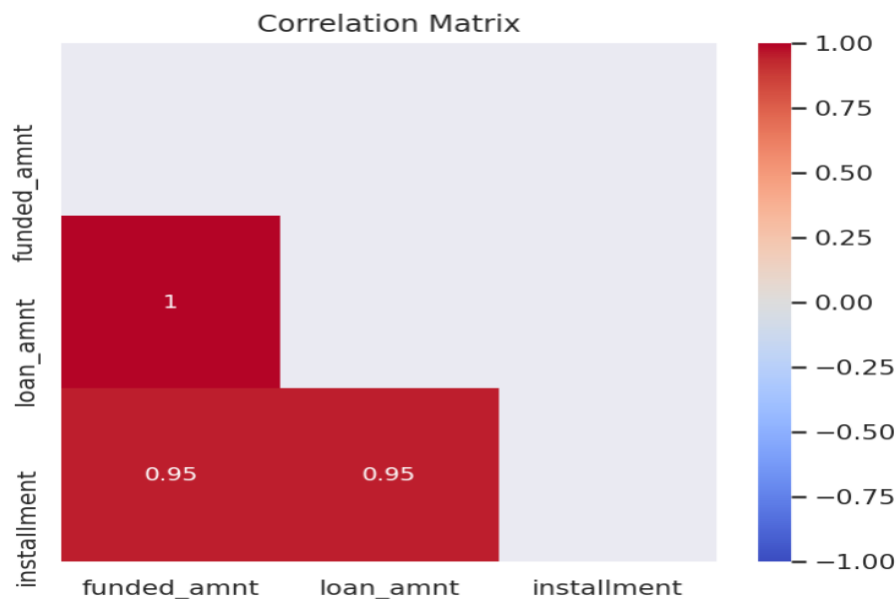
4.4.2. Robustness checks:

To validate the assumptions of this study, several regressions with new variables and different data sets were applied. All the models supported the main results, and there was confirmation of structural validity in this study. The robustness section has two separate parts: first, some variables in Model 5 were replaced with the highly correlated variables, and the regression was re-run. Second, the data set was divided into a smaller number of observations, and Model 5 was run to see whether the new outcomes supported the main finding.

4.4.2.1. Applied the model 5 with new variables:

In the data set, some variables were highly correlated; thus, to avoid the multicollinearity issue, only one was applied simultaneously in the analysis. Outcomes showed a noble correlation among loan_amnt, funded_amnt, and installment. The main regression was implemented each time with one of these variables and results reported in Table 5. Figure 10 also shows this correlation.

FIGURE 4. 10: Correlation matrix



Note: The figure shows the correlation matrix among these three variables.

Note: installment indicates the borrower's monthly payment if the loan originates. funded_amnt defines the total amount committed to that loan no

Funded_amnt had 24,576 observations, with an average of 16,981.13. The minimum value of this variable was 1,000, and the highest was 40,000. The number of observations for the installment is equal to Funded_amnt outcomes with a minimum amount of 30.75 and a maximum value of 1,717.63.

Table 4- 5:Results for robustness model 1.

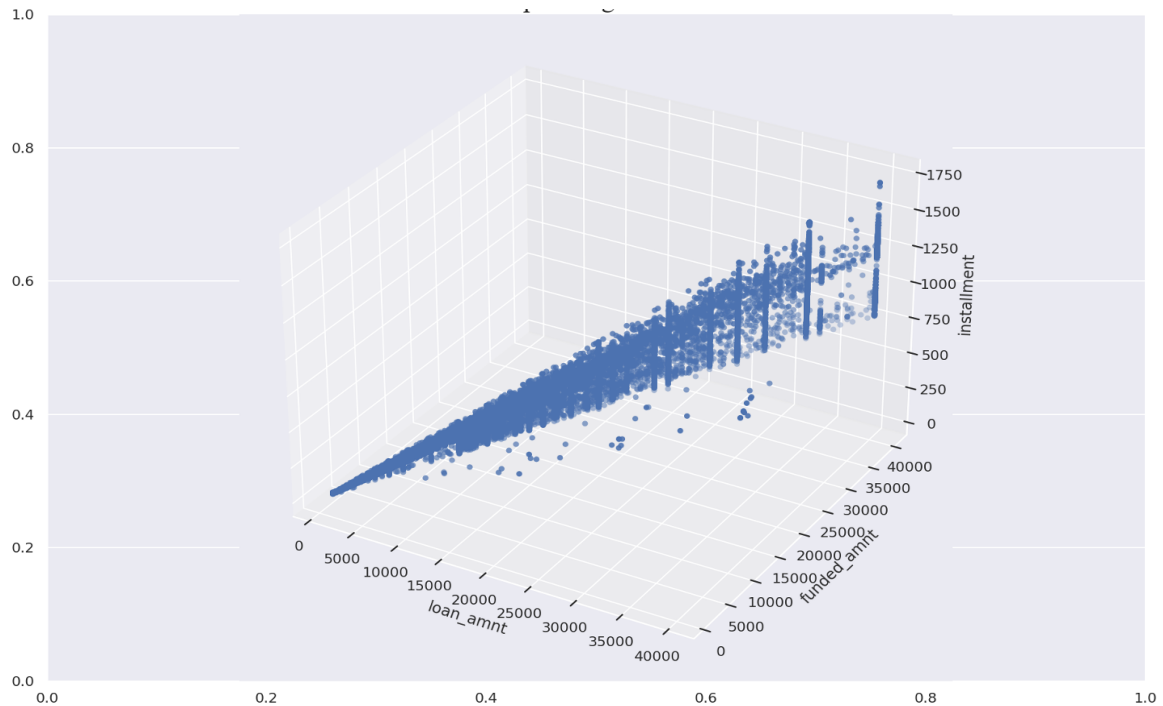
Variables	(1)	(2)	(3)
	Loan Status	Loan Status	Loan Status
tax_cut	0.137*** (0.00868)	0.136*** (0.00994)	0.138*** (0.00991)
loan_amnt	-0.126*** (0.000385)		
funded_amnt		-0.112*** (0.000432)	
installment			-0.379*** (0.000135)
grade	-0.128*** (0.00231)	-0.118*** (0.000121)	-0.127*** (0.000261)
dti	-0.0449* (0.000175)	-0.0466** (0.000179)	-0.0465** (0.000177)
total_acc	-0.0496* (0.000257)	-0.0454* (0.000375)	-0.0464* (0.000372)
total_pymnt	0.121*** (0.000375)	0.130*** (0.000591)	0.143*** (0.000600)
last_fico_range_avg	0.0208*** (0.0000338)	0.0252*** (0.0000905)	0.0207*** (0.0000898)
_cons	-0.0848 (0.0835)	-0.0844 (0.119)	-0.0820 (0.118)
N	24576	24576	24576
R-sq	0.682	0.648	0.612
Std errors in parentheses	***p<0.001	** p<0.01	* p<0.05

Note: for all three-regression annual_inc, open_acc, tax_lines, mths_since_earliest_cr_line, emp_length and applications_type (Y/N) were applied as control variables.

The highest R square with the value of 0.682 belongs to the regression with loan_amnt variable, followed by funded_amnt. There is only a slight difference among the outcomes of these three regressions. All three variables show a negative relationship with the dependent variables, and the results are statistically significant for all models.

Figure 11 shows the relationship among these three variables.

FIGURE 4. 11:Relationship three correlated variables



Note: The figure shows the relationship among funded_amnt, installment and loan_amnt

There is also a strong correlation, with coefficients of 0.96 or more, among total_rec_prncp, total_pymnt, and total_pymnt_inv in the data set. The same process was applied to these variables, and results were reported in Table 6.

FIGURE 4. 12: Relationships among three highly correlated variables



The `total_pymnt_inv` variable refers to payments received to date for a portion of the total amount funded by investors, and `total_rec_prncp` is the principal received to date in the data set. The number of observations for `total_rec_prncp` is equal to 24,576, with a minimum value of 0 and maximum value of 40,000. The maximum value for `total_pymnt_inv` is 63,296.88, and the average number of this variable is 12,122.82.

Table 4- 6:Results for robustness model 2.

Variables	1 Loan Status	2 Loan Status	3 Loan Status
tax_cut	0.137*** -0.00868	0.139*** -0.00598	0.134*** -0.0057
loan_amnt	-0.126*** -0.000385	-0.123*** -0.000282	-0.127*** -0.000271
dti	-0.0449* -0.000175	-0.0377* -0.000134	-0.0440* -0.000137
total_acc	-0.0496* -0.000257	-0.0421*** -0.000234	-0.0349*** -0.000238
total_pymnt	0.150*** -0.000463		
total_rec_prncp		0.174*** -0.00000103	
total_pymnt_inv			0.167* -0.0000233
last_fico_range_avg	0.0208*** -0.0000338	0.0201*** -0.000058	0.0255*** -0.0000589
_cons	-0.0848 -0.0835	-0.0664 -0.0809	-0.0569 -0.0823
N	24576	24576	24576
R-sq	0.682	0.612	0.679
	Std errors in parentheses ***p<0.001 ** p<0.01 * p<0.05		

Note: for all three-regression annual_inc, open_acc, tax_lines, mths_since_earliest_cr_line, emp_length and applications_type (Y/N) were applied as control variables

Among these three models, the highest R square, with the value of 0.682, belongs to the regression applied to total_pymnt (Model 1). There is only a slight difference among the results of these three regressions. All three variables present a positive relationship with the dependent variables, and the results are statistically significant for all regressions.

4.4.2.2. Applied the Model 5 With New Datasets:

As shown in Figure 13, most applicants were in California; therefore, Model 5 was applied only to California for the robustness check. The same regression was then employed on California, New York, Florida, and Texas since 42% of candidates were from these states. The last regression was run on the whole data set except for California, New York, Florida, and Texas. The results are reported in Table 7. All the three regressions had similar results. The variable of interest (Tax_cut) showed a positive relationship with the independent variables, and the outcome was statistically significant for all three models.

FIGURE 4. 13: Number of applicant’s location around the country

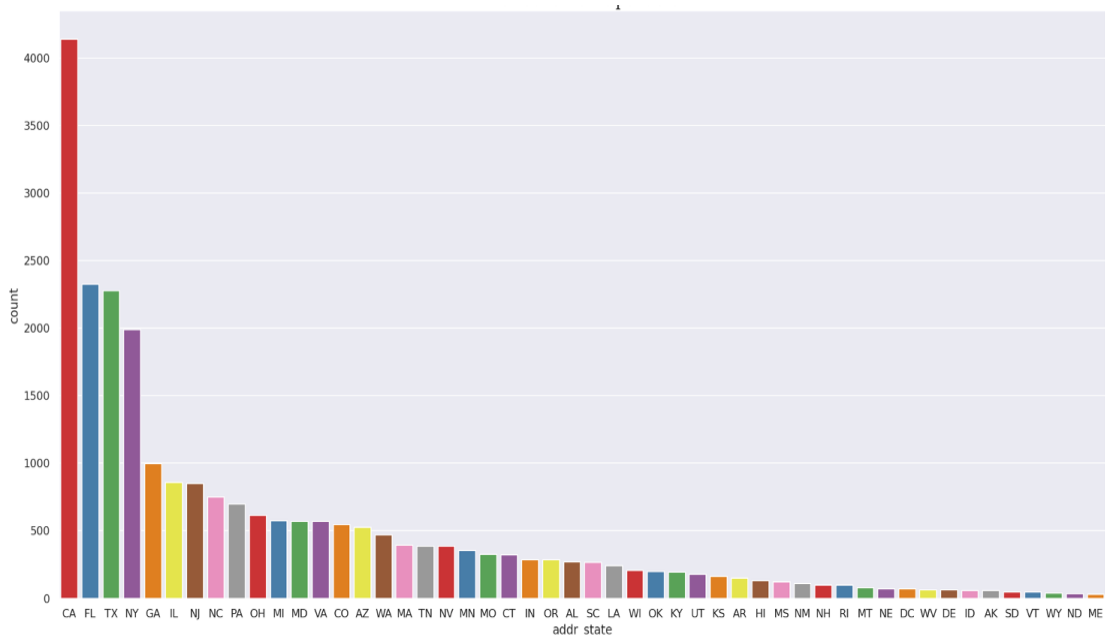


Table 4- 7:Results for robustness model 3.

	Only California	CA, TX, FL, NY	Except for (CA, TX, FL, NY)
Variables	Loan Status	Loan Status	Loan Status
tax_cut	0.313*** (0.0113)	0.137*** (0.00630)	0.140*** (0.00555)
loan_amnt	-0.126*** (0.000385)	-0.185*** (0.000339)	-0.128*** (0.000307)
dti	-0.0449* (0.000175)	-0.0444** (0.00199)	-0.0194 (0.00177)
grade	-0.128*** (0.00231)	-0.126*** (0.000710)	-0.151** (0.000601)
total_acc	-0.0496* (0.000257)	-0.0324*** (0.000349)	-0.0151*** (0.000289)
total_pymnt	0.150*** (0.000463)	0.149*** (0.000368)	0.177*** (0.000325)
last_fico_range_avg	0.0208*** (0.0000338)	0.0476*** (0.000531)	0.0517*** (0.000460)
_cons	-0.0402 (0.164)	-0.044 (0.118)	-0.070 (0.507)
N	4035	10435	13385
R-sq	0.540	0.525	0.507
Std errors in parentheses	*** p<0.001	** p<0.01	* p<0.05

Note: for all three-regression annual_inc, open_acc, tax_lines, mths_since_earliest_cr_line, emp_length and applications_type (Y/N) were applied as control variables

4.5. Discussion of Results:

This research intends to estimate the likelihood of default rate for small businesses by considering financial characteristics, such as loan status, debt-to-income ratio, FICO score, and borrowers' length of employment. The logistic regression model was employed to identify good borrowers using the Lending Club data set.

This study also tried to determine how TCJA affected the default rate by small businesses. Some necessary factors for determining loan defaults were considered in this study, but TCJA is the variable of interest. The outcomes indicated that this regulation has a negative effect on the chance of default by a value of 0.137 and is statistically significant, with a p-value of 0.001. Other studies showed that government regulations, such as tax cuts, could influence loan repayment. For example, Schrag and Pruet (2011) showed that loan repayment assistance programs with new federal legislation in 2011 helped law schools forgive student loans of those graduates who entered public service.

Additionally, the results indicate that the loan amount, debt-to-income ratio, and Lending Club scores played an essential role in mortgage defaults. All had a negative relationship with the dependent variable. Loans with lower credit grades and higher amounts were associated with a higher likelihood of default. This finding is consistent with previous studies in this area (see, e.g., Li et al., 2020; Zhang et al., 2019; Mamonov et al., 2019).

On the other hand, the last FICO average and total payment have a positive relationship with the dependent variable. A higher FICO score and higher payments

received to date for a portion of the total amount funded led to less chance of default or delinquency (see, e.g., Jagtiani et al., 2017; Mach et al., 2014).

The results show that the outcomes of all selected variables for this study were statistically significant except for the application type variable, which indicates whether the borrowers applied for loans individually or jointly.

4.6. Conclusions

This study considers the effects of TCJA on the probability of default (PD) from 2010 to 2019 using the logistic regression method. The data set includes more than 20,000 applicants from across the United States. The analysis began with variable selection, since the original data set comprised more than 80 variables. Then, other validation methods and robustness checks were employed to get more precise results. The findings indicate that government regulations like the TCJA can have a notable effect on the economy.

Previous research explained how important government regulations are to the financial market, especially the mortgage market (see, e.g., Barth, 1983). Government intervention in the financial market is not a new phenomenon. Fannie, Freddie, FHA, VA, HAMP, and HARP are just a few examples of the U.S. government's intervention in the mortgage market. Government regulations are essential in financial markets since average customers are not as financially knowledgeable as their lenders are, leading lenders to take advantage of borrowers (Poon et al., 1997).

Some studies in this area have tried to show some of the positive effects of these laws, as well as some of these regulations' downsides (see, e.g., Nowak et al., 2015). In this case, however, the results show that one of the government's goals in passing this law, to support small businesses, was achieved. This study also tried to introduce some critical

features to the loan market to help lenders decrease the likelihood of default and help policymakers understand the effects of these kind of regulations on small business growth.

Tax policy influences all members of the economy. Tax changes affect the performance of both businesses and individuals, and empirical research is crucial to recognizing the results of such a policy. This type of study can assist policymakers in making more informed decisions on tax regulations. This research adds to a growing field of literature on the effects of the TCJA by studying its effects on small business default rates. The TCJA significantly adjusted the U.S. corporate tax policy by decreasing the federal corporate tax rate from 35% to 21%. Our conclusions confirm that the average U.S. small business targeted took advantage of this law and reduced their loan default rate.

A decent technique is essential for the long-term gain of all sorts of financial institutions, including P2P lending programs. Borrowing money with the classic means would be an excellent opportunity for small businesses, since this type of loan requires less paperwork than conventional loans, which might be considered an attraction. Although this type of loan is sometimes offered with a higher interest rate, easy accessibility makes it more popular than traditional loans. P2P lending has become more prevalent recently; conducting research in this area and examining this market and its influence on small business loans is essential.

This essay measures only one small possible effect of the corporate tax adjustments in the TCJA. Many other topics could be further examined in this field to improve our conclusions. Another fascinating perspective would be how the tax reform has affected foreign firms or small business loan defaults. Future work should consider how these online platforms influence the nature of financing small businesses. It would also be

interesting for future research to analyze the time of default during the life of loan. Using default time is beneficial for policymakers since it may be an indicator of fraud or similar situations (Luo et al., 2020). Using alternative data also would help uncover more details about this research area and decrease the cost of default for lenders. In 2019, Jagtiani and Lemieux determined that many “invisible prime” borrowers with poor FICO scores have a higher possibility of default compared to standard browsers with good FICO scores.

The outcomes of this study are essential for both policymakers and financial institutions. Policymakers need to determine the consequences of the TCJA and use these results to make future regulations more efficient. The outcomes are essential since they provide crucial information about borrower features to save lenders capital. Although these outcomes are based on the Lending Club data set, applying the same model and techniques to other financial institutions’ data in this area would be useful.

Small businesses play a vital role in the economy and financial market. As shown by the results of this study, a valid policy can support the economy and financial markets. There are numerous ambiguities around P2P lending programs. This is the right time for researchers to further examine the influence of P2P on consumers and the financial operation overall as a future study.

Overall, this study shows that the TCJA helps small businesses avoid defaulting on their loans. All the selected variables show a statistically significant outcome as well, so it can be concluded that all are essential features for lenders to consider when granting loans.

4.7. Reference:

Wagner, A. F., Zeckhauser, R. J., & Ziegler, A. (2018, May). Unequal rewards to firms: Stock market responses to the Trump election and the 2017 corporate tax reform. In *AEA Papers and Proceedings* (Vol. 108, pp. 590-96).

Bernstein, W. J. (2021). *The Delusions Of Crowds: Why People Go Mad in Groups*. Grove Press. EARLE, G., NASHUA, C., HAMPSHIRE, N., & GOODIN, E. *Free Trade*.

Divino, J. A., & Orrillo, J. (2021). Robust effects of a debt-financed tax cut in an economy with incomplete markets. *International Tax and Public Finance*, 1-10.

Divino, J. A., & Orrillo, J. (2021). Robust effects of a debt-financed tax cut in an economy with incomplete markets. *International Tax and Public Finance*, 1-10.

Mach, T., Carter, C., & Slattery, C. (2014). Peer-to-peer lending to small businesses.

Nowak, A., Ross, A., & Yench, C. (2018). Small business borrowing and peer-to-peer lending: Evidence from lending club. *Contemporary Economic Policy*, 36(2), 318-336.

Chen, Y. (2017). Research on the credit risk assessment of chinese online peer-to-peer lending borrower on logistic regression model. *DEStech Transactions on Environment, Energy and Earth Sciences*, (apeesd).

Croux, C., Jagtiani, J., Korivi, T., & Vulcanovic, M. (2020). Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior & Organization*, 173, 270-296.

Đurović, A. (2017). Estimating probability of default on peer to peer market—survival analysis approach. *Journal of Central Banking Theory and Practice*, 6(2), 149-167.

Jagtiani, J., Vermilyea, T., & Wall, L. D. (2018). The roles of big data and machine learning in bank supervision. *Forthcoming, Banking Perspectives*.

Bequé, A., & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications*, 86, 42-53.

Serrano-Cinca, C., & Gutiérrez-Nieto, B. et López-Palacios L. (2015). "Determinants of Default in P2P Lending". *PLoS ONE*, 10(10).

Jin, Y., & Zhu, Y. (2015, April). A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending. In *2015 Fifth International Conference on Communication Systems and Network Technologies* (pp. 609-613). IEEE.

- Farrar, D. E., & Robert, R. Glauber. 1967. "Multicollinearity in Regression Analysis: The Problem Revisited." *review of Economics and Statistics*, 49(1), 92-107.
- Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1), 53-71.
- Belloni, A., Chernozhukov, V., & Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34(4), 606-619.
- Baumann, K., Albert, H., & Von Korff, M. (2002). A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part I. Search algorithm, theory and simulations. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 16(7), 339-350.
- Jouan-Rimbaud, D., Massart, D. L., & De Noord, O. E. (1996). Random correlation in variable selection for multivariate calibration with a genetic algorithm. *Chemometrics and Intelligent Laboratory Systems*, 35(2), 213-220.
- Mallows, C. L. (2000). Some comments on Cp. *Technometrics*, 42(1), 87-94.
- May, S., & Hosmer, D. W. (2003). Hosmer and Lemeshow type goodness-of-fit statistics for the Cox proportional hazards model. *Handbook of statistics*, 23, 383-394.
- Meinshausen, N., & Yu, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *The annals of statistics*, 37(1), 246-270.
- Wright, R. E. (1995). Logistic regression. Brock, W. A., & Evans, D. S. (1989). Small business economics. *Small business economics*, 1(1), 7-20.
- Ismoilov, R. B., Mullabayev, B. B., Abdulkakimov, Z. T., & Bakhriddino, J. R. O. (2020). The Essence Of Small Business And Private Entrepreneurship And The Theoretical Basis Of Its Development. *The American Journal of Applied sciences*, 2(08), 45-50.
- Baker, H. K., Kumar, S., & Pandey, N. (2021). Thirty years of Small Business Economics: A bibliometric overview. *Small Business Economics*, 56(1), 487-517.
- Vismara, S. (2016). Equity retention and social network theory in equity crowdfunding. *Small Business Economics*, 46(4), 579-590.
- Schrag, P. G., & Pruett, C. W. (2011). Coordinating loan repayment assistance programs with new federal legislation. *Journal of Legal Education*, 60(4), 583-615.
- Li, X., Deng, Y., & Li, S. (2020). Gender differences in self-risk evaluation: evidence from the Renrendai online lending platform. *Journal of Applied Economics*, 23(1), 485-496.

- Zhang, N., & Wang, W. (2019). Research on Balance Strategy of Supervision and Incentive of P2P Lending Platform. *Emerging Markets Finance and Trade*, 55(13), 3039-3057.
- Mamonov, S., & Malaga, R. (2019). Success factors in Title II equity crowdfunding in the United States. *Venture Capital*, 21(2-3), 223-241.
- Jagtiani, J., & Lemieux, C. (2017). Fintech lending: Financial inclusion, risk pricing, and alternative information.
- Mach, T., Carter, C., & Slattery, C. (2014). Peer-to-peer lending to small businesses.
- Barth, J. R., Gotur, P., Manage, N., & Yezer, A. M. (1983). The effect of government regulations on personal loan markets: a Tobit estimation of a microeconomic model. *The Journal of Finance*, 38(4), 1233-1251.
- Luo, G., Li, W., & Peng, Y. (2020). Overview of Intelligent Online Banking System Based on HERCULES Architecture. *IEEE Access*, 8, 107685-107699.
- Poon, S., & Swatman, P. M. (1997). Small business use of the Internet. *International Marketing Review*.
- Spence, L. J., & Rutherford, R. (2003). Small business and empirical perspectives in business ethics. *Journal of Business Ethics*, 47(1), 1-5.
- Graham, P. G. (1999). Small business participation in the global economy. *European journal of marketing*.
- Coleman, S. (2000). Access to capital and terms of credit: A comparison of men-and women-owned small businesses. *Journal of small business management*, 38(3), 37.
- Watson, K., Hogarth-Scott, S., & Wilson, N. (1998). Small business start-ups: success factors and support implications. *International Journal of Entrepreneurial Behavior & Research*.
- Chu, C. C., Tsai, S. B., Chen, Y., Li, X., Zhai, Y., Chen, Q., ... & Li, B. (2017). An empirical study on the relationship between investor protection, government behavior, and financial development. *Sustainability*, 9(12), 2199.
- Grody, A. D. (2018). Rebuilding financial industry infrastructure. *Journal of Risk Management in Financial Institutions*, 11(1), 34-46.
- Sanfelici, D., & Halbert, L. (2019). Financial market actors as urban policymakers: the case of real estate investment trusts in Brazil. *Urban Geography*, 40(1), 83-103.
- Davis, K. (2016). Peer-to-peer lending: structures, risks, and regulation. *JASSA*, (3), 37-44.

Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: evidence from the Lending Club consumer platform. *Financial Management*, 48(4), 1009-1029.

Kumar, V., Natarajan, S., Keerthana, S., Chinmayi, K. M., & Lakshmi, N. (2016, September). Credit risk analysis in peer-to-peer lending system. In *2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)* (pp. 193-196). IEEE.

Sleem, K. (2010). A Theoretical Role for Government in the Financial Markets. Available at SSRN 1354989.

VITA

MOHAMMAD JAVAD MEHREGAN

Born, Tehran, Iran

2007-2012	B.Sc. Mechanical Engineering Azad University
2013-2015	M.A. Financial Management University of Tehran
2016-2019	M.A. Economics Florida International University
2019-2021	Ph.D. Candidate Florida International University
2017-2021	Graduate Teaching Assistant Florida International University