Allegheny College
Department of Computer Science

Senior Thesis

Quest for an Acceptable Level of Income Inequality: A Comparative Analysis of Implications of Skill-Biased Technological Change and Globalization between Developed and Developing Countries

by

Sweta Rauniyar

Project Supervisor: **Oliver Bonham Carter**
Co-Supervisor: **Janyl Jumadinova**

December 11, 2020

Abstract

Quest for an Acceptable Level of Income Inequality: A Comparative Analysis of Implications of Skill-Biased Technological Change and Globalization between Developed and Developing Countries

Automation today is no longer a luxury but rather a necessity for economies to thrive. Through globalization, it has helped integrate economies around the world. Some workers reap significant benefits from technological progress, while some are displaced and struggle to find new ways of survival. Although the world's economic progress is growing tremendously, this Skill-Biased Technological Change (SBTC) has contributed significantly to wage inequality in developed and developing economies. It raises the question: What level of inequality is acceptable, and what is the effect of this widening gap between the rich and low across the globe has on economic growth and prosperity? This project aims to answer these questions by creating a panel of forty developed and developing economies to assess the relative contributions of Skill-Biased Technological Change (SBTC) and globalization towards income inequality. We implement this assessment by creating a LASSO-based regression analysis where we use the GDP per capita (measure for Income Inequality) as our dependent variable and factors influencing SBTC and Globalization as our independent variables. With our proposed research, we wish to understand if we have been blindfolded by technology advancements and what the current trends could cost to the future of the global economy.

Contents

List of Figures

List of Tables

Template Overview

You should first modify the documents in the preamble, things that appear before the main text as detailed below.

Front page: use the one provided in this template, after changing the values like names in the file `preamble/mydefinitions.tex`.

Abstract: There should be a single paragraph of about 250 words, which concisely summarizes the entire proposal, written in the file `preamble/abstract.tex`.

Acknowledgments, Abbreviations, Glossary, Dedication preamble pages are optional and can be used at the author's discretion.

The main text of the proposal should be stored in the "SeniorThesis.tex" document. The following descriptions are sections that must be included in the thesis document.

Bibliography: The bibliography should include all references cited in the text (as [**?**]) and it should not include references that have not been cited. ACM referencing style should be used when preparing the bibliography. We recommend using BibTeX or BibLaTeX and using the file `preamble/bibliography.bib`.

Chapter 1

Introduction

"It's difficult to cross from one economic class to another. You'll drown in champagne."

In today's time, technology has embedded in our day-to-day activities and businesses. As it changes the way we access information, consumer products, travel, and communicate, it helps us achieve the so-called impossible and makes our lives convenient. It has also changed the model of work. Both developed and developing economies have made transitions from agricultural to industrial and from service to digitalization. The essence of all of these revolutions that humankind has seen lies within the realms of automation. Emerging technologies, including artificial intelligence, machine learning, and advanced robotics, can automate many tasks currently performed by workers, leading us to a quiet revolution, the Automation Revolution [72].

There is much debate about where workplace automation will lead the economy, but observers tend to agree on one thing: The trend is only gaining momentum [74]. Automation in infrastructure and development is no longer a luxury but rather a necessity for humankind to thrive. As firms adopt new methods of production, markets expand, and societies evolve. Economies have continued to grow, and while many will benefit, that growth will not be costless. Some arguments say automation will take away jobs, but the advancements in automation do not eliminate jobs but instead create a "shift" reducing particular job functions at which humans are inefficient or inconsistent or are exposed to risk [72]. Workers in some sectors benefit handsomely from technological progress, whereas those in others are displaced and have to retool to survive [68]. One of the dramatic changes in society because of such technological progress is Skill-Biased Technological Change (SBTC).

1.1 The Human Effect of Income Inequality

At the beginning of the 1970s, economic growth slowed, and the income gap widened. Today income inequality is known as the "defining challenges of our time" [2]. Humankind today may be proud of all the technological advancements made, but besides all this progress, we might be failing to address a pressing issue that adversely affects human rights. Technology claims to make the world more connected than ever, but I find it amusing that we tend to grow further and further apart on the spectrum of wage and inequality. The world might be reaching its peak of development, but some communities are left in the same spot they were in years ago. When we say progress, we have been blindfolded by the top 10 percent of the world's speedy progress while the rest of the world takes small steps. Coming from an underdeveloped country, Nepal, to one of the world's biggest economies, the United States of America, I have seen, felt, and experienced the drastic difference that lies within these economies. The rich in the Nepalese community is nowhere close to the wealthy in the U.S. The situation is worse for the poor. As Nepal takes baby steps towards technological development, the U.S. boasts of its immense progress and so is the case in other economies and hence it is important to understand the value of global prosperity. There is no doubt that technology has transformed the ways we live, but why the question is, why does it have to be at the stake of human labor? Humankind has already established enough differences based on race, religion, and social norms, and one could say that the essence of these categories is somewhat based on income inequality. It is high time that we break these barriers and unite to create a better world that values peace and prosperity.

Inequality is growing for more than 70 percent of the global population, exacerbating divisions' risks and hampering economic and social development [3]. The World Social Report 2020, published by the United Nations, shows that the richest one percent of the population are the big winners in the changing global economy. It is a mind-boggling insight that the bottom 40 percent of the world have earned less than a quarter of the income in all countries surveyed [3]. The wide disparity in wages spreads to health and education areas and is one of the primary reasons people still lie within poverty walls. Technological advancements that complement the skilled workforce make it even worse. It is more likely to be profitable to people with jobs that require high social and analytical skills, and these are jobs that have high earnings already. Workers in mid-to-low-skill roles who rely on physical labor or analytical skills vulnerable to automation are at higher risk of losing their jobs or facing pressure on wages. If recent history is a guide, those who lose their jobs may face lower incomes throughout their career after being reabsorbed into the workforce, and some may choose to drop out entirely [32]. The results of job loss and wage suppression will ultimately lead to income inequality and disrupt the economy.

If the current trend continues, one could wonder what the future of the world's economy will look like. Within communities and beyond, what does it look like if the trend gains momentum. It is high time that we treat this issue at the national and international level. Whether it be within countries or beyond, leaders of all communities need to implement policies that can help bridge the gap between the poor and the rich. To tackle within country inequalities, it is required to increase fiscal policy space at the national level to enact the country-specific mix of policies needed to lift all boats and, in particular, to increase the income of those at the bottom. From an international economic system perspective, these imbalances can be addressed by requiring global financial, investment, trade, monetary, and fiscal reforms to reduce volatility [19]. There should be a just system that treats everyone equally, and this should be done before income inequality becomes the sole reason for political and social inability. And I aim to use my knowledge within the areas of Economics and Computer Science to spark light on the intensity of the issue through this senior project. Besides, work is not just a means of earning; for some people it is a purpose in life. We, humans, have the tool of intelligence that makes us a dominant species of the planet. But how inhumane is it for us to build a progressive economy on one end of the spectrum while on the other end, we are failing to address the silent voices of the people whose lives are being affected. Are we failing to ignore that technology might be robbing humanity? It is important to demand a sustainable balance and that starts by understanding the consequences that have led to people losing their jobs and the increasing wage gap between the rich and the poor.

1.2 Skill-Biased Technological Change and Globalization within Economies

The recent consensus is that skill-biased technological change favors more skilled workers, replaces tasks previously performed by the unskilled, and exacerbates inequality [5]. Countries worldwide have undergone structural changes and although these changes contribute to rapid growth, leaning towards highly skilled workers puts the low-skilled workers at risk. The early life of automation begins, then, with the industrial revolution and industrial machinery between 1790 and 1840 [4]. Even the people were in fear of having their jobs taken away however, things ended well as there was a rise in the productivity of lower-skilled workers, and when technology evolved during the 20th century, the rise of the computers benefited the higher-skilled workers. Many economists say that there should not be any fear of technology. They point to how past major transformations in work tasks and labor markets – specifically the Industrial Revolution during the 18th and 19th centuries – did not lead to major social upheaval or widespread suffering. These economists say that when technology destroys jobs, people find other jobs [75]. Switching to a new job might be easier for those who are sophisticated enough to learn the skills; the high-income bearers. The lower-income bearers barely make enough to support their basic needs and expecting them to be on track with their skill set is amusing to me.

Another reason for the increase in demand for high-skilled labor is trade [37]. The growing interdependence between economies resulting from globalization has created a worldwide market for companies and consumers to benefit. However, the general complaint about globalization is that it has made the rich become richer while making the poor get poorer [23]. With all these changes that foster the world's economic progress and contribute to the wage gap between the rich and the poor at the same time, the question becomes, "What level of inequality is acceptable, and when does it start doing harm?" It is not acceptable to turn a blind eye to these changes, and this project aims to spark light on this issue by doing a comparative analysis between the developing countries and the developed countries and aims to understand the future of the world's economy.

The economic growth of a country is influenced by advanced technological changes and the rise in free trade. However, "Is the rise of the machine, the fall of the worker?." As businesses lean towards automation, the machines become more sophisticated, and such a change requires firms to hire high-skilled workers. Since this biased direction of technological change reduces the working capital's stock by substituting capital and unskilled labor with skilled workers, is this fair progress? Countries are using technologies to produce more in less time to match their increasing demands domestically and internationally. Advanced economies like the U.S and China who have a competitive advantage among many goods, have an increasing demand globally. As nations follow this incentive and produce more goods and services on a larger scale, they have a better chance of decreasing their costs and achieving economies of scale. Investments follow this trend of reduced costs that automate tasks as the goal is to "produce more in less time." Developing nations follow the same steps and integrate technological advancements to meet their local demands and open a gateway to foreign trade.

On the bright side, SBTC and globalization bring economic prosperity to all the countries; however, the more countries experience such changes, the greater is the incentive and the potential to eliminate lower-level jobs and decrease the relative wages of the less-skilled workers. Another reason why this happens is that countries aim to achieve economies of scale by reducing their input costs. As a result, they have accumulated enough to invest this sum into a new resource; hence, the cycle continues. Critics argue that this is more of a "shift" in jobs rather than elimination. However, we fail to address that it is easier for the high-skilled workers to adjust to this change as they have access to resources to help them "catch up." On the other hand, the low-skilled do not have the privilege to make a swift switch mostly because they do not have the resources like sufficient income to invest in their education, primarily why they had to work at a low-skilled job in the first place.

1.3 Theoretical Framework

According to research by McKinsey, it is estimated that between 400 million and 800 million individuals globally could be displaced by automation and need to find new work [48]. However, finding a new job is easier said than done. Research suggests that jobs that are threatened by technological changes are highly concentrated among lower-paid and lower-skilled workers who usually have limited access to education. This shows how automation will continue to create pressure on workers at the lower end of the spectrum as they struggle to learn skills that could help them adjust to technological changes. So people with lower wages not only end up losing their jobs but are challenged to find a new work that goes beyond the scope of their skills and they do not have the resources to upgrade the skills. We are told to welcome automation rather than fear it but the 'we' in this is merely focused on the wealthy folks who are fortunate to make choices and chances to change their life swiftly. However, the rest dread as there is nothing barely anything being done to make things better.

Developed economies are insanely rich, and they could invest in their human capital to make their workers prepared for the transitions. Developing economies, who are still beginners in this game, do not seem to have the resources, and hence their workers are at a loss. Hence we could reiterate that "the rich keep getting richer, and the poor get poorer." Now the question is, what does this trend look like in the long-term for the world's economy? The purpose of this senior thesis is (1) to create an analytical model to understand what are the most important variables influencing SBTC and Globalization that drive income inequality (2) to derive insights from the existing patterns and use them to create a predictive model for the future, and (3) to finally use the insights that could be used to create better future models that emphasize how the current rate of income inequality is unacceptable, and it is high time that we shake the people in power to take actions against the injustice going on. Most importantly the goal is to use the advancements in technology (such as Machine Learning) to solve a real-world problem that has been influenced by technology itself. Technological advancements are there to support the progress of humanity; hence the rise of the machine cannot be the fall of the worker [78].

Chapter 2

Related Work

The related work section will focus on the various scholarly research done to support the argument that Skill-Biased Technological Changes (SBTC) and Globalization contribute to an increasing wage inequality within developed countries and developing economies. These analyses help us discuss the similar approaches taken by these authors to study the factors affecting income inequality, and the aim is to use their findings to create a basis for this senior project.

2.1 Essence of Income Inequality

The rising gap between the rich and the poor has fueled unrest around the world. During the most important events in history, the Industrial Revolution, there was a rapid transition to the modern age resulting in a sustained rise in the real income per person in the Western World [53]. The long formative process of the western capitalist economies rose income inequality during the initial stages of the industrial revolution, primarily in rich countries like Britain and the United States [62]. Low-income countries were not even a part of the game as the process of industrialization was not even uniformly introduced within their economies. This difference in the introduction and adoption has produced inequities among nations and among people on a scale that was never experienced, hence resulting in a spike in economic growth in more prosperous economies while the poor ones are trying to catch up.

The most influential modeling of the relationship between economic growth and inequality was presented first by Simon Kuznets, where he proposed an association between levels of income per capita and inequality in the shape of an inverted U-curve. In the empirical work by Jeffrey and Williamson, the "upswing" in the Western world was associated with two processes: unbalanced technological change (causing growing income differentials between different economic sectors and driving up the skill premium for human capital) and demographic growth [39]. For the "downswing" of inequality in the developed world, globalization was said to have played a significant role. Many countries have followed this up and down trend; however, with time, they are various changes, including government policies, laws, demographics, and most importantly, technology, that prompt us to think well beyond the curve.

Whether it is the agricultural revolution or the industrial revolution, what remains common is innovation. Currently, we lay on the edge of a new revolution- the technological revolution that has fundamentally changed the way we live, work, and connect. On the one hand, technology has become the engine of growth, leading to a more globalized economy where people have become successful, more affluent, and educated. On the other hand, the benefits of rising incomes and aggregate GDP growth rates have not been shared equally across all segments of the population. The advent of new technologies leads to more job and income opportunities, but the skill-biased nature of the new jobs complements the higher-skilled workers whereas the low-skilled are bound to be replaced. There is no doubt that these changes have adverse effects on the economy.

2.2 Evidence of Skill-Biased Technological Changes in Developing Countries

When economies are introduced to new technologies, there is a change in production methods or modifications in the organization of work. Hence, to have a smooth transition, firms turn towards more skilled labor.

In the research paper titled 'Skill-Biased Technology Transfer,' authors Berman and Machin investigate the skill-bias of technological change in developing countries using a global sample of manufacturing industries [13]. The manufacturing industries undergo drastic structural changes as they transition to new forms of technology. In this paper, we turn back to the 1980s and find strong evidence of increased demand for skills in the 1980s in middle-income countries' manufacturing sectors. This analysis links this demand shift due to the skill-upgrading within industries. A cross-country correlation analysis was conducted for 37 countries with data (primarily from the United Nations General Industrial Statistics Database) on employment, wages, and manufacturing industries. The results showed that the demand for skills accelerated in the manufacturing industries of middle-income countries in the 1980s to a rate matching even that of high-income countries, concluding that skill-biased technological change exists, thus supporting the basis of this thesis [13].

In addition to an analysis focusing on multiple countries, it is also helpful to understand the situation between individual economies. Numerous studies focus on individual countries that help create an in-depth analysis of the impact of SBTC changes under various circumstances. India is preparing itself to compete with the leaders of industrialization and technological development. With this economic prosperity, there are also rising concerns about the increasing inequality within this fast-track nation.

Kijima Yoko conducted a study to identify the major causes of changes in the wage structure in urban India during the 1980s and 1990 by conducting surveys that include information on an individual's earnings and the labor market characteristics [38]. The results showed that the demand for skilled workers grew faster than the supply of skilled labor, which raised the returns to skills, resulting in accelerated inequality in the 1990s. Another study by Esposito and Stehrer explains relative demands for skills in Central and Eastern European countries. It presents the hypothesis that the sector bias of SBTC is essential in defining the rising skill premium in three transition economies: Czech Republic, Hungary, and Poland [26]. Another study investigating the existence of SBTC uses Ethiopia to study the effects of imported technology on manufacturing employment and concludes with similar remarks of biased outcomes [31].

In areas such as the MENA region, political instability is already a factor that worsens income inequality. A study was conducted to determine the effects of technological advancements measured by the demand of patents on skilled and unskilled labor through a dynamic panel of countries within the MENA region. The sample data composed of six countries, including Egypt, Algeria, Iran, Malta, Morocco, Tunisia, for the period 1965-2010. Patents, expenditure on Research and Development, and the value of imported technologies were the key variables used in this analysis. There was a positive effect on skilled labor, while the unskilled labor force experienced a negative effect. Overall, the political and economic instability within the MENA region makes income inequality a huge issue, and these findings show the importance of addressing this pressing issue. Within the MENA region, Tunisia is taking steps towards becoming a technology-consumer country. However, there are many obstacles to implementing technological innovations within the country, like high costs, lack of human capital, poor infrastructure, etc. Such reduced technological potential within most developing economies like Tunisia enforces them to depend on foreign technologies. These organizational changes within the industry have resulted in high wage disparities, thus proving biased due to technological changes [7]. Emerging economies are taking initiatives to catch up with developing nations. In a research paper focusing on the big question of whether Skill-Biased Technological Change exists in Vietnam or not, data was collected from the Vietnam Household Living Standard Survey from 2004 to 2014 to measure the relative skill productivity. The research involved developing a regression model for estimating the wages of skilled and unskilled workers. The results concluded an increasing wage gap that is bound to increase wage inequality. An important insight is that Vietnam is an attractive destination for outsourcing due to its cheap labor. The country has been actively investing in high-tech machines and equipment, leading to the inevitable consequence of high-tech capital investment. Such tactics led by developing economies is taken to give them a competitive edge and boost economic growth. This results in the favoring of skilled workers and eventually increasing the wage differentials [54].

2.3 Skill-Biased Technological Change in Developed Countries

In most economically advanced countries, the trend towards income-inequality remains the same. Historically speaking, the Kuznets curve speaks for the relationship between economic development and income inequality. It follows a “U-shape” from 1935 to the mid-1950s and then stability until the 1970s. Then, inequality took off and has increased ever since. The rich economies: U.S. and U.K. have followed this trend quite strictly, and the results are obvious. In 2018, the top 20% of the population earned 52% of all U.S. income [17]. Including the United States, most of the high-income countries are a part of OECD. OECD stands for Organization for Economic Cooperation and Development. It’s an association of 37 nations in Europe, the Americas, and the Pacific to promote economic welfare. Following the U-shaped curve trend, income inequality has been on the rise in most of the OECD and in many emerging economies since the 1980s, and skill-biased technological changes give it momentum [28].

The Economic Department of the OECD presents a study that estimates skill-biased technological changes as the key driver of the increasing earning differentials. The research shows a model that learns from the historical trends to construct scenarios for the years leading up to 2060. The model predicts that if the current trend of SBTC observed during the past 25 years prevails, the earnings differentials will, on average, increase by almost 30 percent by 2060. The study further supports the ideology that income inequality patterns differ not only across countries but also within countries and educational attainment and globalization exert upward pressure on income inequality [15]. Another study explains how skilled refers to the educated and explains the concept using the shift in the supply-demand curve. At any point, the demand for skilled workers is considered to be fixed. Technological changes create an upward shift in the demand for skilled workers, and since the demand is greater than the supply, it leads to an increase in the relative wage of the skilled worker, thus increasing the gap between the high-skilled and the low-skilled. This economic phenomenon is explained through the trends in recent decades within multiple OECD countries, including Germany, France, etcetera [11]. While many empirical studies focus on a specific industry (for instance, manufacturing) over individual economies, this debate was widened by cross-country analysis. The study focused on the impact of SBTC and International Trade (IT) on skill-based wage inequality carried across a panel of 25 OECD countries from 1997 to 2006. The sample was further split into two groups: developed countries and developing countries. In the study, the expenditure on Research and Development was used to proxy the effects of SBTC, while Foreign Direct Investment (FDI) played a crucial role in analyzing International Trade (IT). The results suggested that SBTC was the main reason behind inequality in developed countries, while in developing countries, IT dominated that role [9]. Openness to trade allows emerging economies to implement technological advancements similar to those within the developed economies, so is it fair to see that soon the developing economies will find themselves in the shoes of the developed economies.

The situation in some of the world's largest economies is not different. In an empirical analysis conducted by Cristiano and Franceso, the rate and direction of technological change in a significant sample of 12 major OECD countries in the years 1970–2003 confirm the strong bias of new technologies [10]. The study further supports another study's findings that the same country experiences different levels of economic growth at different periods, and these can be identified by the changes in the direction of technological change [25]. New and convincing evidence has been provided recently about the strong skill-bias of the gales of technological change based upon information and communication technologies introduced in the last decades of the twenty-first century [28]. It has been a key driver in increasing the differences between people's earnings. Another model decomposes historical changes in earning differentials and uses it to construct forward-looking scenarios up to 2060. If the common cross-country trend of skill-biased technological change observed during the last 25 years prevails, earning differentials will increase by almost 30% on average in the OECD (Organisation for Economic Co-operation and Development) [15]. Skill-biased technological changes can have a dramatic effect on wages and the labor market. An examination of the impact of the SBTC on the demand of productions and non-production (unskilled and skilled) workers of the Japanese manufacturing industries was implemented through cross-sectional regressions. It showed that increased investments in computers significantly impacted the share of the wage-bill held by non-production workers []. The study defined production workers as people who are engaged in production at manufacturing establishments. While on the other hand, the non-production workers included people working at higher positions like supervisors, technical employees, and other office work [63]. The study predicts that demand for non-production or skilled workers will continue to accelerate in the years to come due to the direction of SBTC.

2.4 Combining Skill-Biased Technological Change with Globalization

The recent accelerations in the advancements of technology have also increased the globalization of production. Technology has revolutionized the global economy and has become a critical competitive strategy [40]. Through globalization, companies benefit from free trade, and consumers enjoy low prices. However, it has also created structural changes in the economy. The National Bureau of Economic Research states that trade influences what kind of technologies are more profitable to develop and tends to increase the price of skill-intensive products. It provides firms with incentives to introduce new skill-biased technologies within their manufacturing processes. In other words, trade and globalization induce skill-biased technological change [6].

Technology creates a sense of competitiveness that triggers a race to imitation and innovation for both firms and countries. Both firms can feel obligated to catch up with the technological trends in their production processes to ensure they are not lagging. However, they do so at the cost of a larger share of unskilled labor in their workforce. This argument is supported by a study conducted by Thoenig and Verdier, where they show that when globalization triggers an increased threat of technological leapfrogging or imitation, firms tend to respond to that threat by biasing the direction of their innovations towards skilled labor-intensive technologies [70]. Another finding is that there is a growing international trade integration between advanced economies and low wage countries. According to standard Heckscher-Ohlin theory, it has shifted labor demand away from unskilled workers in high wage economies and eventually create a bias towards high education workers [41]. To understand the consequences of this integration, they conduct their analysis in two regions, North-North and North-South, and conclude an increase of wage premium in both areas.

Another study focuses on how lower-middle-income countries (LMIC) rely on international technology transfer to upgrade their technological sectors. In this analysis, a Generalized method of moments (GMM) technique is applied to a panel data constituting 0f 20 manufacturing sectors for 23 countries over a decade. The results provide evidence of widening employment differentials. They introduce a new variable to overcome one of the challenges mentioned in their study: the absence of innovation and employment data in low- and middle-income countries. The results indicated a significant increase in skilled workers' demand for imports of industrial machinery, equipment, and ICT (Information and Communication Technology) capital goods. Hence, the findings support this senior thesis's arguments that technological advancement and trade directly or indirectly impact the inequality within workers [24]. It is assumed that due to their pace, developed countries are generally abundant in skilled-labor. When they increase their trade with developing countries where unskilled labor is abundant, it raises the skilled workers' wage relative to the unskilled workers. According to the Stolper-Samuelson theorem, the increasing trade should lead to a decline in wage inequality; however, the scenario is different [29]. A study using plant-level data in Mexico by Hanson and Harrison found that although trade barriers were reduced, there was evidence of rising wage inequality [same as above]. Another study studied the impact of technological changes across various firms in Mexico, Columbia, and Taiwan and found that the employer-size-wage effect is higher for skilled workers than for unskilled workers in technology investing firms. Such firms invest in exporting and conducting RD and provide training to their employees. The argument presented was that technological changes are skill-biased and that larger firms are more technology-intensive [29].

Lower Middle Income Countries (LMIC) are not the only ones struggling. The results are prominent within developed economies where globalization drives economic growth. Innovation in such economies requires a significant amount of skilled labor, and this is an attractive feature for multinational corporations (MNC) who are willing to invest in these sectors. According to the 2019 World Investment Report, China was ranked the world's second-largest Foreign Direct Investment (FDI) recipient after the United States. In a study that performs regression analysis to firm-level data from the Chinese electronic industry, the findings show that despite the high adoption cost of technology, multinational firms' investments lead to higher productivity of skilled labor in developed economy subsidiaries than in the emerging economies [43]. The study concludes by reiterating the point mentioned above that innovation in developed economies favors skilled labor. It is more costly for MNCs of developed economies to invest in innovation techniques of unskill-biased technologies and customize their production facilities within developed economies [43].

2.5 Statistical Models within Machine Learning: LASSO Regression

Most works in Economics focus on the application of traditional statistical methods. Today, innovative techniques within Artificial Intelligence (AI) like Machine Learning (ML) have transformed the way people live. Netflix's show recommendations, Instagram's likable posts, Apple smartwatch that tracks health, and Siri's voice-recognition system that responds to user's commands are a few examples of how ML has made our lives easier. Nevertheless, many economists have not embraced this new and innovative technology. Unlike the standard econometric models, the predictive algorithms that make Siri smarter over time cannot answer questions about correlation and causation. With the bulk amount of data that we have today, we must turn our heads to the innovative methods in Machine Learning that provide accurate trends. When it comes to making predictions, the existing techniques tend to "over-fit" the data sample and make flawed generalizations of new and unseen data. Hence, this focus on the predictions' accuracy is where machine learning algorithms play an essential role. Materials scientists seek to use these learning algorithms to easily and efficiently apply to their data to obtain quantitative property prediction mode [47].

Machine Learning models can minimize the forecasting error by trading off bias and variance and handling vast amounts of data focusing on prediction problems. In addition to over-fitting, the traditional methods also create an overestimation of how well the model performs using the included variables to explain the observed variability (optimism bias). Such problems can be addressed by various penalized and regularization regression models, and some of these models are Ridge and LASSO (Least Absolute Shrinkage and Selection Operator) [59]. In this senior thesis, I plan to use LASSO regression since it leads to improved model accuracy. Tibshirani introduced this regularization method to subset the variables from the original list based on the extent that they impact the model [71]. The model has been applied in a wide range of disciplines.

In Bioinformatics, recent technological advancements have been highly captivated by the computational techniques and algorithms. In this study, they propose the development of a fused lasso logistic regression (FLLR) to differentiate the patients of early Alzheimer's disease (AD) from normal controls based on the corpus callosum (CC) thickness profiles. The callosal thicknesses are spatially correlated because the thickness at one point is correlated to the thicknesses of its neighboring point. The study focuses on using fused lasso regression to select continuous regions rather than individual thickness points to differentiate the groups. The accuracy of the classifier was estimated to be 84 percent based on five-fold cross-validation [42]. Recent scientific and technological advancements show the emergence of MicroRNAs (miRNA) as a significant class of regulatory molecules involved in a broad range of biological processes and diseases ranging from Alzheimer's to Diabetes. With the increasing amount of miRNA and gene expression data, a study conducted to accurately identify miRNA-mRNA pairs proposed the LASSO regression model. It concluded it to be a robust tool as it deals well with sensitivity and specificity when used for the diagnoses and the treatment of complex diseases [45]. The LASSO method is also widely used within high-dimensional data. When combined with penalized logistic regression, LASSO was used in high-dimensional cancer classification. It led to a relatively efficient and feasible classification by selecting fewer genes with a high area under the curve and a low misclassification rate [8].

A study conducted on predicting corporate bankruptcy implemented Lasso and Ridge regression to design a model for an extensive data set of 2032 non-bankrupt firms and 401 bankrupt firms belonging to the hospitality industry over the period 2010-2012. The algorithms were used because they deal better with multicollinearity and display the ideal properties to minimize the numerical instability due to overfitting [58]. The model used plenty of variables, and it became essential to identify the most critical variables in a data set. Using Lasso regression allows us to shrink the parameter estimates close to zero or even zero, excluding some of the model variables. Such implementations improve the model's prediction accuracy, leading to Lasso's outperformance against existing statistical methods [76]. Numerous studies widely use the logistic regression method in economics. However, Hosmer and Lemeshow report that although logistic regression copes with many restrictive assumptions of ordinary least squares regression that include linearity, normality, and heteroscedasticity, it cannot handle multicollinearity [33]. With this concept, Lasso was also used to create a model that ensures economic sustainability by predicting franchisor success or failure based on numerous financial and contractual variables [18]. The model correctly predicted 82.47% of franchisor survival in the training sample and 81.82% in the test sample, supporting the model's accuracy. In addition to this, the LASSO regression has also been an attention-grabber in the stock-market alongside investors. To compare its efficiency against the traditional linear regression techniques within the stock market, a study was carried out that was composed of daily closing stock price in Indonesian Stock Market over the years 2000 to 2014. The study focused on analyzing the effect of the stock market of G7 Nations and the Association of Southeast Asian Nations (ASEAN) on the Indonesian stock market. There was high correlation and multicollinearity within the variables; hence the solution is geared towards shrinking the parameter coefficients, aka LASSO regression. LASSO beat the linear regression techniques using the least-squares method and became the best model in the Indonesian Stock Market [65].

LASSO has also made some great appearances in the field of social science. In the heterogeneous world of social sciences, the common method uses a multiplicative interaction term between the treatment and a hypothesized effect modifier to evaluate theoretical arguments. These standard methods use a "base-line" regression model to identify the relationship between interest variables and the hypothesized moderator. However, biased estimates have been critical problems, eventually leading to overfitting and unstable estimations. Data-driven approaches within Machine Learning can be used to address the direct and indirect regularization bias as conducted in a study that uses a post-double selection approach that uses several Lasso estimators to select the interactions to include in the final model [14]. According to the evidence from the simulation, the LASSO-based model showed better performance. In another study conducted on US midterm elections, it was found that due to low survey response rates, there has been greater use of non-probability samples, which may lead to selection bias. The study estimated the voting preference for 19 elections in the US 2014 elections using non-probability surveys from SurveyMonkey. The model was adjusted to estimated control totals using a model-assisted calibration combined with adaptive LASSO regression (also called estimated controlled LASSO, ECLASSO). This methodology was shown to be powerful compared to the traditional calibration methods and has been proposed to be applied across various social science and health research disciplines as they struggle to receive probability samples [21].

What makes the LASSO technique highly attractive is that it improves predictions by shrinking the regression coefficients. This feature has helped accurately forecast the variability in solar irradiance that reaches the ground by moving clouds. The study was conducted in Hawaii to a second irradiance time series data, and LASSO was used because its bias-variance trade-off leads to better predictions. LASSO considers the sum of l1-norms of the regression coefficients as a penalty, which shrinks the regression coefficient. This feature suits the application in this study where one can control the down-wind stations' influence on the up-wind stations in a highly correlated data set [77]. Few more studies provide evidence of LASSO's prediction techniques; one of the models is used in remote sensing. The model carries out a biomass estimation using satellite data in a region with extremely low vegetation cover. Multiple empirical models are applied to test the ability to deal with a large dataset, and LASSO was shown to outperform all the models. Working with an extensive data set of satellite-based predictors shows how efficient the Lasso model can be. Another climatological research uses LASSO regression to identify systemic biases in decadal precipitation predictions from a high-resolution regional climate model (CCLM) for Europe. The algorithm was applied to observe precipitation and numerous predictors related to precipitation derived from a training simulation. This trained Lasso regression model was then transferred to a virtual forecast simulation for testing. LASSO outperformed the local predictors, and hence his bias adjustment has been said to contribute to improving the seasonal cycle of precipitation prediction [44].

The feature of variable selection is crucial in some sensitive areas such as medicine. For instance, in epidemiology, it has been used to detect collective exposure from a pool of candidate variables diagnosed with Hepatitis B. Implementing the LASSO regression technique in this study helped detect the crucial factors that resulted in the infection, and intensive simulations were used to compare this method with previously researched methods[30]. LASSO can also be used to implement models with high-dimensional data and hence reduce the amounts of computation. To identify the risk of disease recurrence and the benefit of adjuvant chemotherapy for patients who have had surgery for stage II colon cancer, Zhang built a model using the LASSO regression from the high-dimensional microarray data [79]. Another model utilized the LASSO regression for selecting variables from multiple factors that influence energy consumption in residential buildings [64]. Another crucial application aligns with climate change, which is currently one of the most critical global concerns. Household Carbon Emissions (HCEs) have contributed significantly to the rise in CO2 levels in the atmosphere. There are many factors of HCEs; however, a clear understanding of the driving factors achieved through LASSO regression would be critical for policymakers to rank the factors according to their importance and make required modifications to their policies [66]. With the amount of data and problems increasing globally, it is worth noting that setting our aims high and resolving everything might not be a great tactic. However, what could help is identifying the pressing issues and finding solutions to them, and Lasso regression can be useful for such implementations.

Chapter 3

Method of Approach

3.1 Brief Overview

In this project, the objective is to create an eight-year panel-data analysis for 12 countries. These countries come under the category of developed, emerging, and low-income economies. To have a global perspective, these countries are chosen in a way that resembles regions worldwide, including East Asia Pacific, Europe Central Asia, Latin America, Middle East, and Africa, North America, South Asia, and Sub-Saharan Africa.

3.2 Country Classification by Income Level and Years

The World Bank classifies the world's economies into four income groups: high-income, upper-middle-income, lower-middle-income, and low income. This classification is based on the Gross National Income (GNI) per capita, which is currently measured in US$ []. The GNI is a measure of the total income earned by the people and businesses, whether they are located within the nation or abroad [def. Investopedia]. The GNI per capita for high-income countries is greater than $12,375, whereas the GNI per capita for upper-middle-income countries is between $3,996 and $12,375. For lower-middle-income-countries, the GNI ranges from $1,026 to $3,995, and for low-income countries, it is lower than $1,026. This thesis focuses on the first three categories: high-income, upper-middle-income, and lower-middle-income.

High-Income economies are advanced with better infrastructure, mature capital markets, and most importantly, high household incomes. Multinational corporations are attracted to such economies, and they tend to invest heavily, leading to accelerated economic growth and job opportunities. Next, upper-middle-income are in rapid growth, but in comparison to high-income, they have lower capital markets and lower household incomes. The progress of an upper-middle economy allows it to engage more with the global market, leading to increased trade and foreign direct investment (FDI). In short, upper-middle-income (also called emerging economies) are transitioning towards becoming a high-income economy. The next category that has been defined in this project is lower-middle-income economies. These countries struggle to provide basic necessities such as food and water, have limited access to education, and low returns to work. Based upon this classification, the final list of countries created for this study is included in the table below.

Country	**Category**
United States	High-Income
Germany	High-Income
Australia	High-Income
Japan	High-Income
Israel	High-Income
China	Upper-Middle Income
India	Upper-Middle Income
Russia	Upper-Middle Income
South Africa	Upper-Middle Income
Mexico	Upper-Middle Income
Nepal	Lower-Middle Income
Nigeria	Lower-Middle Income

Table 3.1: Categorization of Economies Based on the Level of Income

For the panel-data analysis, the data has been extracted from 2012 to 2019. The Great Recession of the year 2007 cast a long shadow over the economic expansion that followed. The world's financial markets took the biggest hit, along with the banking and real estate industries. The recession led to an increase in home mortgage foreclosures, and millions of people were in a situation where they were losing their life savings, jobs, and homes. In 2012, the World Economic situation and Prospects mentioned that after much turmoil during 2011 global capital markets gained some stability in early 2012. During the first quarter of 2012, some high-income economies such as United States, Japan, and Germany were predicted to have some calming financial markets and were expected to recover moderately. Furthermore, most middle-income economies expected less volatility in private capital inflows, more moderated swings in exchange rates and modest stock market gains. The state of the global economy was fragile, and despite the circumstances, the world progressed slowly and steadily. Each economy in the world has transitioned differently following the years of the economic crisis. In addition to the trends of the transition, the analysis over those years will further spark light on how an economic situation of a country allows it to rise above the circumstances and how the resources available to help them get on on the right track.

3.3 Dependent and Independent Variables

It is crucial to analyze the gap between countries to understand why some countries are rich and why others are poor. In doing so, multiple factors or variables will be used to search for the cause and effect relationship to identify why things are a certain way and use those insights to get reliable predictions. We have classified our variables as dependent and independent variables. The dependent variable is the variable that is being tested and measured for this thesis is income inequality, and its value will be affected by independent variables. On the other hand, the independent variable is the variable that will be manipulated by us and is assumed to have a direct effect on income inequality. In this study, those effects are led by Skill-Biased Technological Change and Globalization.

3.3.1 Dependent Variable

Studies show that income inequality is a condition that prevails along with economic growth. Earlier, we discussed the inverted U-shaped Kuznets curve that showed that income inequality is the lowest at the initial level of low economic growth. As the nation's economic growth progresses, income inequality increases until a threshold level and decreases with increased economic growth. Our study is based on the idea that economic growth has an impact on income inequality. GDP is considered to be the single best indicator of economic growth. In a study conducted to understand the relationship between annual GDP growth and income inequality in developed and developing economies, a linear relationship was found [46]. The study further serves as an understanding of what countries could expect to happen based on their GDP.

A country's GDP accounts for the total value of goods and services produced within a country's borders during a specific period. Breaking down this output and dividing it by a country's population leads us to GDP per capita. The metric shows how much economic production value can be accounted to the citizens of the country. Furthermore, it helps with a better analysis of the average living standards and well-being when comparing countries. A citizen is supposed to be doing well financially when they have acceptable living standards, which can be accounted for through their income. Hence, this measure of economic activity will be acquired from the World Bank's website

3.3.2 Independent Variable

To analyze the impact of Skill-Biased Technological Change and Globalization on our dependent variable, GDP per Capita, the following list of factors that will be accounted as our independent variables.

1. **Expenditure on Research and Development (R&D)**:

   Research and Development (R&D) is a significant factor contributing towards technological advancement across nations. It is a process that has the incentive to create new or improved technology to provide a competitive advantage at the business, industry, or national level, hence playing a crucial role in GDP growth. A study conducted on some selected OECD countries showed that R&D expenditures positively and significantly impact economic growth [69]. R&D expenses support a lot of industrial, technological, health care, and pharmaceutical sectors. Many economists agree that innovation can lead to higher productivity for workers, leading to increased wages, improved quality of life, and greater output for the economy. These industries are also the ones to be most impacted by skill-biased technological changes. As a result of R&D, when new tools and machinery are integrated within the manufacture of products and services, the demand for high-skilled laborers who can adjust to the change increases. To gain a competitive advantage and economies of scale, the countries and firms have an incentive to provide innovative and efficient products or services that will gain recognition at the national and international markets. Hence, they tend to invest a lot within R&D. The data for R&D (as a percent of GDP) is accessible from the World Bank's database.

2. **Availability of Scientists and Engineers:**

   Development within nations has been linked with advancements in science and technology. Nations that have a large number of citizens involved in generating knowledge and creating technology do not have to rely on other countries to develop new tools. This variable is used as an indicator to calculate the availability of highly educated human resources available in a given country. People in such positions are experts in applying scientific and technical knowledge and play a vital role in the economy's sustained growth and stability. They are collectively regarded as a nation's highly qualified workforce and contribute towards understanding and solving some of the complex challenges faced by a business, industry, or country, which often involves developing new technologies. A Wall Street Journal article noted that the quality of scientist and engineers and their proximity to research centers is crucial to multinational corporations [56]. The variable will play an essential role in identifying how technological leadership affects a nation's economic strength. We will access this data from UNESCO's (United Nations Educational, Scientific and Cultural Organization) database. This variable's value will range from one to seven, one being scarce and seven being widely available

3. **Expenditure in Education:** Education is an essential identifier of economic growth and development. Educated citizens have access to a wide range of opportunities compared to those who are not, and it becomes easier for them to adapt to the changing environment. Broad availability of quality education is a foundation for future training as people transition to new jobs [55]. One of the best predictors of an individual's income is educational attainment; hence a nation's strategic policies towards investment in education could equip the

skills required for the jobs of today and tomorrow [50]. Especially in low-income nations, lack of education can lead to increased poverty and slow economic development. With access to better educational programs, citizens of a country can earn skills and apply them to their country's development. According to a recent OECD report, providing every child with access to education and the skills needed to participate fully in society would boost GDP by an average of 28% per year in lower-income countries and 16% per year in high-income countries for the next 80 years [16]. The data for a country's expenditure on education will be accessed from the World Bank's database.

4. **High Technology Exports:**

   The strong emphasis on investments in research and development, largely in manufacturing and production sectors, has led to the creation of high technology products, leading to high technology exports, which significantly impact economic well-being. High-technology exports are products that have an RD intensity and such exports include aerospace, computers, pharmaceuticals, scientific instruments, electrical machinery etcetera. Developing countries are the popular exporters of such products. A study by Connolly provides empirical evidence that high technology goods imports from developed countries not only positively affect domestic innovation, but also lead to increased GDP growth as higher quality capital goods are used in domestic production [27]. Implementation of these technologies makes the production process efficient on the one hand. On the other hand, the operation of such technologies takes away low-level jobs and instead favors more skilled jobs. We will acquire the data set for this variable from the World Bank's database.

5. **Number of Exports and Imports:** International trade, as an engine of economic growth, has gained a lot of interest. It allows a country to promote efficient allocation of resources, foster technological progress, gain competitive advantage, and achieve economies of scale. A country's imports and exports have a massive role in influencing the GDP per capita. A neoclassical growth modeling framework investigated the contribution of exports and imports to economic growth in a few European Union economies; the results suggested a positive relationship [12]. With more and more countries being open towards free trade, the consumers from all these countries can enjoy low-prices and various goods and services that encourage economies to involve more in import and export activities. However, some studies suggest based on the country's trade, and economic structure, an increase in trade could result in income disparities [61]. This variable will be used to analyze the economic progress and investigate how expensive or beneficial this shift can be for a nation's workers. The data will be accessed from the World Bank's database.

6. **Foreign Direct Investment (FDI):**

   A foreign direct investment (FDI) happens when a business firm or individual in one country shows business interests in another country. Such interest is specifically essential for developing and emerging economies that do not have sufficient funding to expand their businesses. A case study on Pakistan was conducted to analyze the impact of FDI on GDP using a regression model, and a positive relationship was established [36]. FDI brings technological expertise within various industries that influence economic growth. In 2017, developing countries received $671 billion, or 47% of total global FDI. Investments rose 9% in developing Asia, which received $476 billion [52]. The recipient countries see growth in jobs and standard of living. Few consequences of FDIs significant to this thesis include trade deficits and disruption of domestic business practices. For instance, a state-level panel data analysis conducted in the US led to the conclusion that FDI can positively affect income inequality within some states [22]. The relationship between FDI, GDP, and income inequality has been relatively ambiguous. The data for this variable will be accessed from the world bank's database.

7. **Gross Capital Formation (GCF) (% of GDP):**

   Gross Domestic Capital formation refers to the net investment in physical assets for a particular year for an economy. The asset includes factories, machinery, plants, equipment, and materials that can increase the physical capital stock of the nation. The increase in capital formation can provide a country's population with necessary production tools, stimulate technological progress, and increase labor productivity, leading to an increased output level for the nation. It determines the national capacity to produce, which in turn affects economic growth. A research paper implemented the Ordinary Least Squares Technique (OLS) to investigate the impact of capital formation on Nigeria's economic growth and showed a positive relationship [73]. It can be argued that deficiency of capital and acts as a barrier in the production of goods and services. The data for this variable will be accessed from the world bank's database.

8. **Openness to Trade:**

   Prior to defining trade relationships with a country, it is important to look at what extent is the country "open" to defining its commercial relations with the rest of the world. The degree of openness is heavily impacted by the overall political, economic, and legal situation of a country. There are different degrees of openness based on the restrictions imposed by the country on free trade. A higher degree of openness can lead to new market opportunities for domestic firms, job opportunities for people, poverty reduction, and prominent economic growth. According to the World Bank, no country has thrived in modern times without harnessing economic openness- to international trade, investment, and the movement of people. An empirical study was conducted to show the relationship between openness to trade and GDP using the General Method of Moments (GMM) on 87 countries (developing and developed). The results supported the endogenous theory that a bidirectional causal relationship exists between the two variables, hence making trade openness a prime driver of economic growth [35]. The data for openness to trade will be extracted from the World Bank's database.

9. **Access to Internet:**

According to the World Bank, broadband (or high-speed) internet access is not just a luxury but a basic necessity for economic and human development in both developed and developing countries. Access to the internet in developing countries can help improve the delivery of essential services of education and healthcare, accelerate telecommunication infrastructure, reduce poverty, and stimulate growth. McKinsey Global Institute's research shared that the internet accounted for 11 percent of GDP growth over 2005 - 2010 for emerging economies (China, Brazil, India) and almost double for advanced economies (United States, Germany) [49]. Broadband also helps expand the reach of task-based work through online outsourcing platforms, which are projected to provide millions of jobs and billions of dollars in revenue over the coming years [1]. The internet has also become a platform for international trade, leading to increased economic growth and prosperity. The data will be accessed from the World Bank's database.

10. **Number of Patents:**

    A patent is an exclusive right granted by the government to an investor interested in manufacturing, using, or selling an invention for a certain number of years. Countries tend to collaborate to influence their technological advancements. An empirical panel-data analysis conducted using global patent data to investigate the importance of quality and quantity of patents on economic growth in 58 countries indicated that countries hosting firms with a high number of patents witness significant economic growth. This globalization of innovation is a means of gaining competencies abroad lacking at home, rather than exploiting home technological strengths. The empirical findings also indicate that the intensity of globalization of innovation is higher in the multidisciplinary country–industry pairs who compete internationally in trade [20]. The data on the number of patents will be extracted from the World Intellectual Property Organization (WIPO), WIPO Patent Report.

11. **Economic Freedom:**

    In an economically free society, individuals have the freedom to produce, trade, consumer, and invest in any way they desire without any government intervention. The Economic Freedom Index, published by The Heritage Foundation, documents the positive relationship between economic freedom and various positive social and economic goals. This index is based on the four pillars of freedom (Rule of Law, Government Size, Regulatory Efficiency, and Open Markets). Hence, it will give us an in-depth analysis of a country's economic and political situation [51]. A panel-data research conducted to study the relationship between economic freedom and growth for the fifty US states established a positive relationship between the two variables [22]. The thesis will explore how economic freedom across multiple countries influences economic growth.

12. **Labor Force Participation Rate:**

    The labor force participation rate is an essential measure of the active workforce of an economy. It includes the sum of all workers (16 or older by age) that are employees or are actively seeking employment. It can be an important metric to analyze the employment and unemployment level of an economy. The growing labor force of the United States has provided a tremendous boost to the potential rate of expansion in the economy [60]. Another study found a positive correlation between the labor force and GDP within Bangladesh [34]. For this thesis, the labor force participation rate for the enlisted economies will be accessed from the World Bank's database.

3.4 Data Analysis: Tools and Techniques

The final data set will include each of the variables collected individually for the enlisted countries, ranging from 2012 - 2019. The organized data set is stored in a CSV (comma-separated values) file. Now that the data is ready to be used, the next step is to examine the data to develop meaningful insights. Given the data's size and complexity, a Machine Learning model will be implemented to identify patterns and structures within the data.

3.4.1 Machine Learning

The fundamental essence of the research lies within the realms of Machine Learning, a subset of Artificial Intelligence. It is based on the idea that systems can make inferences from an existing data set, learn the signals within the data and use it on new data sets to make decisions with minimal human intervention. There is so much data in the world, and it is beyond human capacity to analyze by themselves. Machine learning uses powerful algorithms to learn from the data set and make useful predictions based on the patterns detected. It powers so many of the services that we use today, including search engines like Google, streaming platforms such as Netflix, Youtube, Spotify, social media platforms such as Facebook, Instagram, Twitter, and even more. Within these platforms, the system is collecting a huge amount of data - including what are people searching for, what genres are they watching, what links are they clicking, what kind of posts are the posts they are interested in - all helping the system to make an educated guess on what might people want next. Beyond these platforms, many other industries, including business corporations and financial services, healthcare, government etcetera have also recognized the value of machine learning. Some of these real-world applications include image recognition, speech recognition, medical diagnosis, statistical arbitrage, prediction and forecasting, classification, information extraction, and developing regression models, and many more. Machine Learning is growing rapidly and will eventually lead to more advancements in technology.

3.4.2 Model Structure

I. Programming Language

The programming language used for the implementation of the model is Python. Python is an open-source, interpreted, and objected-oriented high-level programming language. It uses simple syntax, dynamic-typing, and high-level data structures that make it easier to use. This popular programming language also provides a large standard collection of libraries, especially machine learning libraries that gives Python an edge in data science application. For our data-oriented analysis, Python provides several libraries that assist with data handling, including $pandas$, $sci-kitlearn$, $NumPy$, and $Matplotlib$. The right combination of features and performance of Python makes it a powerful language. The most recent version of Python that is Python3, will be used for the study.

II. Libraries in Python

The various libraries that will be used in the implementation of the model are as follows:

- **Pandas:** Pandas is a popular Python library primarily used for data manipulation and analysis. It provides high-performance and easy-to-use data structures that make it convenient for data wrangling. This thesis involves importing multi-dimensional data from a CSV format and the pandas library will help in storing the data in a two-dimensional structure data frame that consists of rows and columns. The library further supports data aggregation and data visualization.
- **NumPy:** NumPy is short for Numerical Python. It is a Python library that provides scientific programming tools. Numpy allows us to present numeral data in the form of arrays. Various mathematical and statistical functions within NumPy can be used to work with small-sized arrays and large-sized multi-dimensional arrays. It is a simple yet essential library that helps perform heavy numerical computing, which can help extract important information hidden within the data.
- **Matplotlib:** Matplotlib is a multi-platform data visualization library in Python. It is used to create high-quality graphs, plots, scattergrams, and more. It is built on NumPy arrays. This multi-platform plotting library makes it convenient to visualize the important trends effectively.
- **Scikit-Learn:** Scikit-Learn is a popular robust machine learning library in Python. It provides a wide range of machine learning algorithms and statistical modeling techniques for data mining and analysis. It features methods of classification, regression, clustering, and dimensional reduction. Scikit-learn is built upon NumPy, Matpplotlib, and SciPy (an open-source library used for mathematical computing). It is an extremely versatile library that is widely used in solving real-world data problems.

3.4.3 Techniques within Machine Learning

The purpose of every statistical model is different. Machine Learning provides multiple techniques that can be implemented to accomplish the goals of the model. So, it is important to choose the right kind of estimator. Below is a flowchart designed by sci-kit that provides a rough guide to the users, including a pathway to solve their problems.

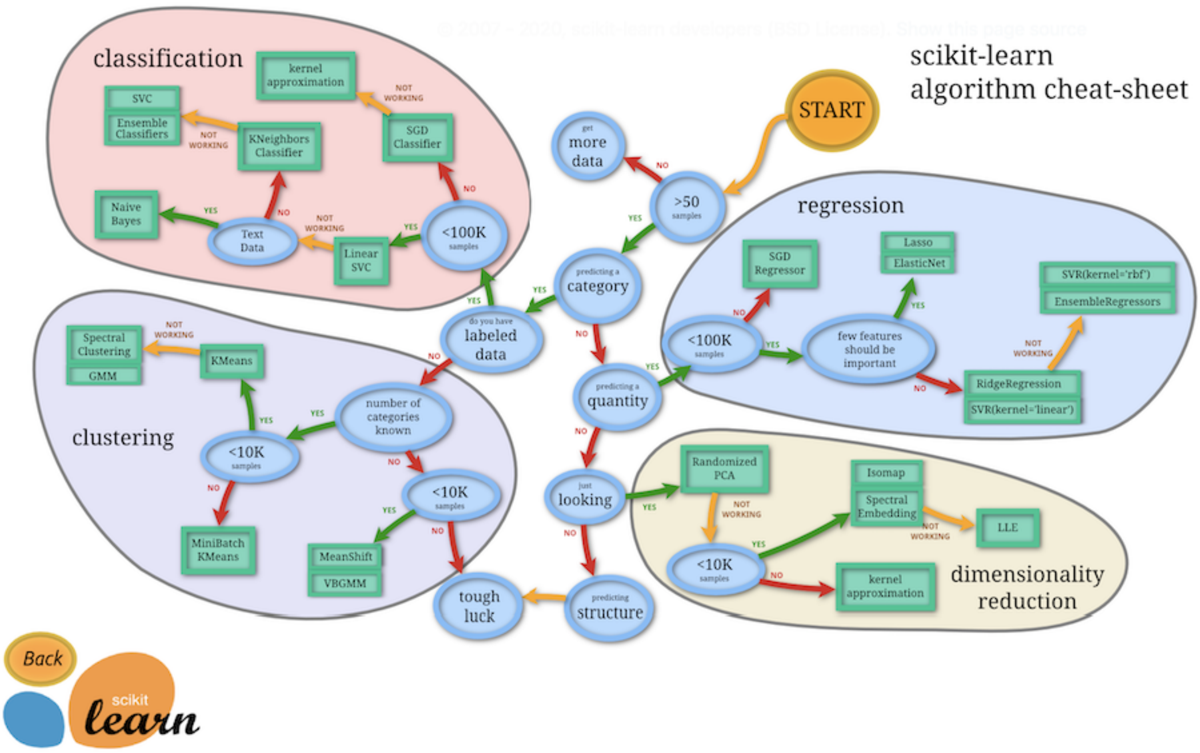

Figure 3.1: Sci-Kit Learn Algorithm Cheat Sheet [57].

According to our predictions and the flowchart, we have planned to implement the techniques of Regression within our model.

3.4.4 Regression

Regression is one of the most important data analysis methods used for data mining and predictive modeling. It aims to describe how changes within independent variables affect our dependent variable. This analysis method will help us answer some of the critical questions like which variables matter the most, which ones can be ignored, how do the variables interact with each other. In addition to this, the analysis will also help us identify whether the variables that we have chosen are going to lead to an accurate analysis or not. Hence, upon implementation, we can identify which variables we have chosen are statistically significant and what role each variable plays in affecting our dependant variable (GDP per capita). Based on our findings, we may have to control each variable to analyze the impact of the specific independent variable on our dependant variable and regression allows us to do so.

3.4.5 Special Kind of Regression Technique

Linear Regression is one of the most popular and simplest kinds of regression. In such a regression model, the relationship between the dependent variable (Y) and the independent variable (X) is assumed to be linear in nature, meaning an increase or decrease in one variable leads to an increase or decrease in another, respectively. A linear regression line has a straight line equation of the form $Y = a + bX$, where X is the independent variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when $x = 0$). The slope of the line is also called the regression coefficient. The regression coefficient represents the mean change in the dependent variable for one unit change in the independent variable while holding the other independent variables constant. Holding other variables constant allows the model to interpret each variable's effect in isolation from the others.

However, in most real-life scenarios, the relationship between the variables of a data set is not linear which is why a straight-line relationship would not fit the complex dataset properly. This is why a complex function is needed to give a more accurate representation of the data. Other limitations of the linear model are overfitting and multicollinearity, which affect the accuracy of our results. Overfitting occurs when we assume our trained model is 99% accurate, but when we feed new "unseen" data to our model, the accuracy goes down. It indicates that our model does not generalize well when we feed new data to it. On the other hand, multicollinearity is when there is a strong correlation between our independent variables. In this thesis, we argue that SBTC and Globalization go hand-in-hand; hence it is arguable that their variables are correlated. For instance, the increase in expenditure in Research Development (as a contributing factor for SBTC) would lead to a rise in high-tech exports (contributing factor for Globalization). Hence, if the correlation between the variables is high, it can cause problems to our model and lead us to incorrect conclusions. As a solution, we need a model that can learn what variables best contribute to the model's accuracy; such models are called regularization embedded models. One of such models is a subset of the linear regression model within machine learning: LASSO Regression.

3.4.6 LASSO Regression

LASSO Regression (Least Absolute Shrinkage and Selection Operator) is a regression analysis method built on the principle of a linear regression model. What sets it apart from the standardized regression model is that it performs variable selection and regularization (prevents overfitting), the primary reasons for choosing this model. One of the model's primary goals is to obtain a subset of accurate predictor variables (or independent variables) and hence enhance the prediction accuracy for the model.

I. Variable Selection and Overfitting

In this study, thirteen variables account for Skill-Biased Technological change and Globalization. There is a possibility of high collinearity, or there may be variables that have a lesser impact than others and can add unnecessary noise to the model. The variable selection feature allows us to have a subset of predictors. For prediction purposes, this method can help us save time by not measuring the redundant predictors. Hence, through variable selection, one can answer questions like: what are the variables that have the most and the least impact on GDP per capita, which of those variables account for Skill-Biased Technological change and globalization, respectively, and how closely related are these variables? It is important to have a machine learning model that will generalize well from the training data and work with any new data sets. The current analysis includes 12 countries representing different economic backgrounds from different parts of the world. Based on the country's income level, the final results could help detect a pattern followed by other countries with similar backgrounds that are not included in the analysis. The insights from the current research could possibly contribute towards creating a prediction tool or as a foundation for future research on other economies.

II. Regularization in LASSO

Overfitting occurs when a statistical or machine learning model is tailored to a specific data set and fails to generalize to other datasets, thus reducing the model's accuracy. To prevent this from happening, the process of regularization introduces additional information (a regularization term) to the model.

The straight-line equation for a linear regression is $y_1 = \theta_0 + \theta_1 x_1$. The θ_1 in this equation is the regression coefficient. The coefficients help us interpret whether the relationship between the independent and dependent variables is positive or negative. (A positive coefficient value indicates that as the independent variable increase, the dependent variable also increases. A negative coefficient value indicates that as the independent variable increases, the dependant variable decreases.) When there is more than one independent variable, the linear equation takes the form:

$$y_i = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_i x_i$$

$$y_i = \theta_0 + \sum_{i=1}^{n} (\theta_n x_n) \tag{3.1}$$

The regularization method involves penalizing the loss function, known as the residual sum of squares. A loss function is a measure of how accurately the machine learning model accurately predicts the true value. The loss function takes two items as input: the true value (y) and the model's predicted value (y_i). The function is defined as a squared error to avoid negative values). In a linear model, the most common way of calculating the loss function is the Residual Sum of Squares (RSS), and the sum of the total residuals is expressed in the form RSS $= (y_i - y)^2$. For a function with multiple variables, the equation will look like this:

$$RSS = \sum_{i=1}^{n}(y - \theta_0 - \sum_{j=1}^{n}\theta_1 x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Now adding the lasso regularization term (L1) to this will look like:

$$RSS + \lambda\sum_{j=1}^{p}|\theta_j| \tag{3.2}$$

In equation 3.2, θ_j is the regularization term or penalty and λ is the regularization strength. Hence, in LASSO, the goal of the algorithm is to minimize all coefficients' size by introducing the penalty called L1 or lambda. Lambda acts as a constraint on the parameters of the model. It shrinks the regression coefficients towards 0 (by forcing the sum of the absolute value of regression coefficients to be less than a fixed value). Variables with a regression coefficient of zero after shrinkage are excluded from the model.

III. The Regularization Penalty (λ)

The value for lambda can vary. If lambda is set to zero, it implies that all the variables are considered important for the model. If lambda is set to infinity, that implies none of the features are considered. Hence, the closer the value of lambda gets to infinity (∞), the more variables are eliminated. And, larger the penalty, the further the coefficients are shrunk towards zero.Too choose the value of lambda that best fits our model, we will be using the popular k-fold cross-validation approach. For this approach, the data set will be randomly partitioned into k sub-samples of equal size. These k-subsamples are used as a training set for developing a prediction model. The remaining sub-samples will be used to validate the model. For accuracy, this procedure is carried out k number of times, with each one of the k sub-samples, in turn being used for validation and the other ones for model development. Finally, we will be combining all the k-separate validation results for a range of lambda values and choose the most appropriate lambda value. This technique will reduce overfitting as we are not restricted to a single subset of the data set for internal validation.

Overall, regression coefficients have an important role in a statistical model as they help us identify the relationship between variables. Standard regression models (such as linear regression) usually have regression coefficients that are of larger value, and hence they are not compatible with working with highly correlated variables. In our thesis, we are looking at the impact of skill-biased technological change and globalization, which tend to have strong ties with each other. For example, when developing countries undergo structural changes to increase their industrial productivity, they trade with other countries, resulting in an increase in technological imports. However, these productivity gains are coupled with a growing gap between the employment of skilled and unskilled workers [67]. LASSO will help us better understand such relationships between individual variables and provide accuracy to our analysis.

3.4.7 Implementation of LASSO in Python

For the implementation of LASSO regression, we have used the Python3 environment and have pre-installed the $scikit-learn$ library. Next, we have to import some packages into our environment (file Lasso.py), including $pandas$, $NumPy$, $scikit-learn$, and $Matplotlib$. The following import statements will be used for this process:

```
import pandas as pd
import numpy as np
from sklearn.linear_model import Lasso
import matplotlib.pyplot as plot
```

In the same directory where the python process (Lasso.py) is based, we have stored the dataset 'Economies.csv'. The rows include the list of the countries alongside the years 2012 - 2019. The columns include all the dependent and independent variables. A snippet of what the data set look like is given below:

To read the data from the file 'Economies.csv,' we will be using the pandas data frame, which is a two-dimensional data structure with labeled rows and columns. Next, the *iloc*() function within *pandas*, will help us retrieve a particular value or set of values belonging to a row and column using the index values assigned to it. The desired values of X (rows) and Y (columns) can be selected through this manner:

```
X = economies_pd.iloc[:, :-1]
Y = economies_pd.iloc[:, -1]
```

The iloc() function and the sklearn model in $scikit-learn$ will be used to split the datqset into two parts: training set and testing set (use X_train, X_test, y_train, y_test). 25% of the data set will be used for training. The snippet of the code that will be used for this process is given below:

```
x_train, x_test, y_train, y_test = train_test_split(economies_pd.iloc[:, :-1],
economies_pd.iloc[:, -1],test_size = 0.25)
```

Year	Country	GDP per Capita	R&D	ExpEd	EngSciAvl	HighTech	Exports	Imports
2012	US	2.2495458509774	2.68166		5.42	172387240523	13.528918971166	17.039264973664
2013	US	1.8420810701195	2.70972	4.931049823761	5.35	172145368079	13.54452443065	16.468482783025
2014	US	2.5259734504071	2.71924	4.9617400169373	5.32	179263970137	13.531590400314	16.427997422204
2015	US	2.9080218612834	2.71742		5.42	178349526944	12.43291493401	15.322240254668
2016	US	1.6378384570389	2.76145		5.53	176668196714	11.900500647478	14.639015639051
2017	US	2.3698007844042	2.81741		5.75	156937052150	12.165156887204	15.011219358154
2018	US	2.9273257323297	2.83766			156365524736	12.287096055245	15.248491917248
2019	US	2.1611762882992				156361791913	11.73295611216	14.581197436168
2012	Germany	0.4184975942176	2.86813	4.9565801620483	4.53	204071158697	46.30712014308	40.206934735968
2013	Germany	0.43759130314467	2.82105	4.9611101150513	4.92	210129818072	45.418677859391	39.660198836858
2014	Germany	2.2095434313487	2.86691	4.9396200180054	4.92	216297040017	45.619263312872	39.000830079626
2015	Germany	1.4919315276077	2.91197	4.8349800109863	4.98	199797306303	46.920738356608	39.325486256601
2016	Germany	2.2299998678201	2.91712	4.8402199745178	5.03	206133813845	46.073262854336	38.696383113113
2017	Germany	2.601976004626	3.03763	4.9051198959351	5.15	195752362801	47.204174412398	40.207524249508
2018	Germany	1.2679952322507	3.09415			210082307180	47.372549837475	41.223837373861
2019	Germany	0.55545088794567				208677809287	46.892912541134	41.096214899755

Figure 3.2: Dataset for Economies

I. Training Set and Testing Set

Once the dataset has been split into training and testing data set, the lasso model will be imported from sklearn. The code snippet for this is as follows:

```
#import Lasso regression from sklearn library
from sklearn.linear_model import Lasso

#train the model
lasso = Lasso(alpha = 1)
lasso.fit(x_train, y_train)
y_pred1 = lasso.predict(x_test)
```

As discussed earlier, the lasso model uses the L1 regularization by introducing a penalty called λ. To determine the accurate value of lambda, we will be using the k-fold cross-validation method (module: from sklearn.model_selection import cross_val_score) and plot the cross-validation (CV) R^2 scores of the training and test data as a function of λ. The chosen lambda will have the highest R^2.

```
from sklearn.model_selection import cross_val_score
```

```
for ind, i in enumerate(lambdas):
    reg = Lasso(alpha = i)
    reg.fit(X_train, y_train)
    results = cross_val_score(reg, X, y, cv=5, scoring="r2")

    train_r_squared[ind] = reg.score(X_train, y_train)
    test_r_squared[ind] = reg.score(X_test, y_test)
```

II. Plotting

A good visualization will help tell an easy-to-understand story. With our visualization tools, we aim to look at the trends and outliers that will further help derive insights crucial to the topic. There are multiple libraries in Python that will help us visualize our data. In this thesis, we will be using Matplotlib and Plotly. Matplotlib is the most popular data visualization library of Python and is a 2D plotting library. It has a versatile visualization library and is easy to use. The library enables us to create plots, bar charts, histograms, scatter plots, pie charts, etcetera. Plotly, on the other hand, is a web-based data visualization toolkit. It has a great Application Programming Interface (API) that makes it convenient to use. Some unique functionalities of Plotly include dendrograms and 3D charts along with scattered plots, contour plots, line charts, bar charts, etcetera. Using Plotly will help us visualize our results from the implementation of the lasso regression and the various values chosen for lambda to identify which value is the most accurate. Based on the results that we get, we can make inferences from our data regarding how SBTC and Globalization play a crucial role in affecting income inequality.

Chapter 4

Experimental Results

This chapter should describe your experimental set up and evaluation. It should also produce and describe the results of your study. Possible section titles are given below.

4.1 Experimental Design

4.2 Evaluation

4.3 Threats to Validity

Chapter 5

Discussion and Future Work

This is the conclusion. You might want to leave it unnumbered, as it is now. If you want to number it, treat it like any other chapter.

This chapter usually contains the following items, although not necessarily in this order or sectioned this way in particular.

5.1 Summary of Results

A discussion of the significance of the results and a review of claims and contributions.

5.2 Future Work

5.3 Conclusion

Bibliography

[1] Connecting for inclusion: Broadband access for all.

[2] Remarks by the president on economic mobility.

[3] Rising inequality affecting more than two-thirds of the globe, but it's not inevitable: new un report | | un news.

[4] A history of automation: The rise of robots and ai. *ThinkAutomation* (Feb 2020).

[5] Acemoglu, D. Technical change, inequality, and the labor market. *Journal of economic literature 40*, 1 (2002), 7–72.

[6] Acemoglu, D. Technology and inequality, 2003.

[7] Aissaoui, N., and Ben Hassen, L. Skill-biased technological change, e-skills and wage inequality: Evidence from tunisia.

[8] Algamal, Z. Y., and Lee, M. H. Penalized logistic regression with the adaptive lasso for gene selection in high-dimensional cancer classification. *Expert Systems with Applications 42*, 23 (2015), 9326–9332.

[9] Almeida, A., and Afonso, Ó. Sbtc versus trade: testing skill-premia evidence across 25 oecd countries. *Applied Economics Letters 17*, 15 (2010), 1497–1501.

[10] Antonelli, C., and Quatraro, F. The effects of biased technological change on total factor productivity: empirical evidence from a sample of oecd countries. *The Journal of Technology Transfer 35*, 4 (2010), 361–383.

[11] Atkinson, A. B. The distribution of earnings in oecd countries. *International labour review 146*, 1-2 (2007), 41–60.

[12] Awokuse, T. O. Causality between exports, imports, and economic growth: Evidence from transition economies. *Economics letters 94*, 3 (2007), 389–395.

[13] Berman, E., and Machin, S. Skill-biased technology transfer around the world. *Oxford review of economic policy 16*, 3 (2000), 12–22.

[14] Blackwell, M., and Olson, M. Reducing model misspecification and bias in the estimation of interactions. Tech. rep., Working Paper available at https://mattblackwell. org/files/papers/lasso . . . , 2020.

[15] Braconier, H., and Ruiz-Valenzuela, J. Gross earning inequalities in oecd countries and major non-member economies: Determinants and future scenarios, Jul 2014.

[16] Brende, B. Why education is the key to development.

[17] Bureau, U. C. Data.

[18] Calderon-Monge, E., Pastor-Sanz, I., and Huerta-Zavala, P. Economic sustainability in franchising: a model to predict franchisor success or failure. *Sustainability 9*, 8 (2017), 1419.

[19] Carpentier, C. L., Wright, R. K., and Passos, F. D. Goal 10-why addressing inequality matters.

[20] Causa, O., Serres, A. d., and Ruiz, N. Growth and inequality: A close relationship?, 2014.

[21] Chen, J. K. T., Valliant, R. L., and Elliott, M. R. Calibrating nonprobability surveys to estimated control totals using lasso, with an application to political polling. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 68*, 3 (2019), 657–681.

[22] Chintrakarn, P., Herzer, D., and Nunnenkamp, P. Fdi and income inequality: Evidence from a panel of us states. *Economic inquiry 50*, 3 (2012), 788–801.

[23] Collins, M. The pros and cons of globalization, May 2015.

[24] Conte, A., and Vivarelli, M. Imported skill-biased technological change in developing countries. *The Developing Economies 49*, 1 (2011), 36–65.

[25] David, P. A., et al. The tale of two traverses: Innovation and accumulation in the first two centuries of us economic growth. Tech. rep., 2005.

[26] Esposito, P., and Stehrer, R. The sector bias of skill-biased technical change and the rising skill premium in transition economies. *Empirica 36*, 3 (2009), 351–364.

[27] Gani, A. Technological achievement, high technology exports and growth. *Journal of Comparative International Management* (2009).

[28] Goldin, C. D., and Katz, L. F. *The race between education and technology.* harvard university press, 2009.

[29] Görg, H., and Strobl, E. Relative wages, openness and skill-biased technological change.

[30] Guo, P., Zeng, F., Hu, X., Zhang, D., Zhu, S., Deng, Y., and Hao, Y. Improved variable selection algorithm using a lasso-type penalty, with an application to assessing hepatitis b infection relevant factors in community residents. *PLoS One 10*, 7 (2015), e0134151.

[31] Haile, G., Srour, I., and Vivarelli, M. Imported technology and manufacturing employment in ethiopia. *Eurasian Business Review 7*, 1 (2017), 1–23.

[32] Harris, K., Kimson, A., and Schwedel, A. Labor 2030: The collision of demographics, automation and inequality, Mar 2019.

[33] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. *Applied logistic regression*, vol. 398. John Wiley & Sons, 2013.

[34] Hossain, M. I. Total labor force and gdp of bangladesh: an analysis in between the year 2002-2009.

[35] Idris, J., Yusop, Z., and Habibullah, M. S. Trade openness and economic growth: A causality test in panel perspective. *International Journal of Business and Society 17*, 2 (2016).

[36] Iqbal, N., Ahmad, N., Haider, Z., and Anwar, S. Impact of foreign direct investment (fdi) on gdp: A case study from pakistan. *International Letters of Social and Humanistic Sciences 16*, 5 (2014), 73–80.

[37] Kang, S., and Hong, D.-P. Technological change and demand for skills in developing countries: an empirical investigation of the republic of korea's case. *The Developing Economies 40*, 2 (2002), 188–207.

[38] Kijima, Y. Why did wage inequality increase? evidence from urban india 1983–99. *Journal of Development Economics 81*, 1 (2006), 97–117.

[39] Korzeniewicz, R. P., and Moran, T. P. Theorizing the relationship between inequality and economic growth. *Theory and Society 34*, 3 (2005), 277–316.

[40] Lamba, T., and Malhotra, H. Role of technology in globalization with reference to business continuity. *Global Journal of Enterprise Information System 1*, 2 (2009), 70–77.

[41] Leamer, E. E., et al. The heckscher-ohlin model in theory and practice, 1995.

[42] Lee, S. H., Yu, D., Bachman, A. H., Lim, J., and Ardekani, B. A. Application of fused lasso logistic regression to the study of corpus callosum thickness in early alzheimer's disease. *Journal of neuroscience methods 221* (2014), 78–84.

[43] Li, B. Multinational production and choice of technologies: New evidence on skill-biased technological change from china. *Economics Letters 108*, 2 (2010), 181–183.

[44] Li, J., Pollinger, F., and Paeth, H. Comparing the lasso predictor-selection and regression method with classical approaches of precipitation bias adjustment in decadal climate predictions. *Monthly Weather Review 148*, 10 (2020), 4339–4351.

[45] Lu, Y., Zhou, Y., Qu, W., Deng, M., and Zhang, C. A lasso regression model for the construction of microrna-target regulatory networks. *Bioinformatics 27*, 17 (2011), 2406–2413.

[46] Luan, Z., and Zhou, Z. The relationship between annual gdp growth and income inequality: Developed and undeveloped countries.

[47] Mannodi-Kanakkithodi, A., Pilania, G., and Ramprasad, R. Critical assessment of regression-based machine learning methods for polymer dielectrics. *Computational Materials Science 125* (2016), 123–135.

[48] Manyika, J. Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages, Nov 2017.

[49] Manyika, J., and Roxburgh, C. The great transformer: The impact of the internet on economic growth and prosperity, Feb 2019.

[50] Mayer, S. E. The relationship between income inequality and inequality in schooling. *Theory and Research in Education 8*, 1 (2010), 5–20.

[51] Miller, T., Holmes, K. R., and Feulner, E. J. *2010 index of economic freedom.* Wall Street Journal, 2010.

[52] Nachum, L. United nations conference on trade and development (unctad): World investment report 2000: Cross-border mergers and acquisitions and development united nations, 2001.

[53] Nardinelli, C. Industrial revolution and the standard of living.

[54] Nguyen, T. Skill-biased technological change in vietnam.

[55] Office, I. L. A skilled workforce for strong, sustainable and balanced growth: a g20 training strategy, 2010.

[56] Onyeiwu, S. Does investment in knowledge and technology spur "optimal" fdi in the mena region? evidence from logit and cross-country regressions. In *African Economic Development Conference, Tunis* (2008).

[57] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[58] Pereira, J. M., Basto, M., and da Silva, A. F. The logistic lasso and ridge regression in predicting corporate failure. *Procedia Economics and Finance 39* (2016), 634–641.

[59] Ranstam, J., and Cook, J. Lasso regression.

[60] Regev, M. D., and Tali. Labor force participation and the prospects for u.s. growth, Nov 2007.

[61] Rosenfeld, T. R. International trade and income inequality: The case of latin american countries.

[62] Ryckbosch, W. Economic inequality and growth before the industrial revolution: the case of the low countries (fourteenth to nineteenth centuries). *European Review of Economic History 20*, 1 (2016), 1–22.

[63] Sakurai, K. Biased technological change and japanese manufacturing employment. *Journal of the Japanese and International Economies 15*, 3 (2001), 298–322.

[64] Satre-Meloy, A. Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models. *Energy 174* (2019), 148–168.

[65] Setiawan, B. Lasso technique application in stock market modelling: An empirical evidence in indonesia. *Sriwijaya International Journal of Dynamic Economics and Business 2*, 1 (2018), 51–62.

[66] Shi, X., Wang, K., Cheong, T. S., and Zhang, H. Prioritizing driving factors of household carbon emissions: An application of the lasso model with survey data. *Energy Economics 92* (2020), 104942.

[67] Srour, I., Taymaz, E., and Vivarelli, M. Skill-biased technological change and skill-enhancing trade in turkey: evidence from longitudinal microdata.

[68] Sumagaysay, L. World bank: Robots are taking some jobs, but not all, Jan 2019.

[69] Surani, S., Gendron, W., and Maredia, S. The economic impact of research and development.

[70] Thoenig, M., and Verdier, T. A theory of defensive skill-biased innovation and globalization. *American Economic Review 93*, 3 (2003), 709–728.

[71] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*, 1 (1996), 267–288.

[72] Tobenkin, D. The automation revolution, Aug 2019.

[73] Ugochukwu, U. S., and Chinyere, U. P. The impact of capital formation on the growth of nigerian economy. *Research journal of finance and accounting 4*, 9 (2013), 36–42.

[74] Uzialko, A. Workplace automation is everywhere, and it's not just about robots, 2019.

[75] Vardi, M. Y. What the industrial revolution really tells us about the future of automation and work. *The Conversation* (Aug 2019).

[76] Wang, S., Ji, B., Zhao, J., Liu, W., and Xu, T. Predicting ship fuel consumption based on lasso regression. *Transportation Research Part D: Transport and Environment 65* (2018), 817–824.

[77] Yang, D., Ye, Z., Lim, L. H. I., and Dong, Z. Very short term irradiance forecasting using the lasso. *Solar Energy 114* (2015), 314–326.

[78] Yglesias, M. The automation myth. *Vox* (Jul 2015).

[79] Zhang, J.-X., Song, W., Chen, Z.-H., Wei, J.-H., Liao, Y.-J., Lei, J., Hu, M., Chen, G.-Z., Liao, B., Lu, J., et al. Prognostic and predictive value of a microrna signature in stage ii colon cancer: a microrna expression analysis. *The lancet oncology 14*, 13 (2013), 1295–1306.