C O V E N T R Y
U N I V E R S I T Y

Faculty of Engineering, Environment and Computing

School of Computing, Mathematics and Data Science

MSc Data Science
7150CEM Data Science Project (December 2023)

Economic Dynamics in China: A Comprehensive Time-Series Analysis and Forecasting of Key Indicators

Author: Alexandra Espialidou

SID: 13047171

Supervisor: Dr Mark Johnston

Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Data Science

Academic Year: 2023/24

Declaration of Originality

I declare that this project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists. I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

Statement of copyright

I acknowledge that the copyright of this project report, and any product developed as part of the project, belong to Coventry University. Support, including funding, is available to commercialise products and services developed by staff and students. Any revenue that is generated is split with the inventor/s of the product or service. For further information please see www.coventry.ac.uk/ipr or contact ipr@coventry.ac.uk.

Statement of ethical engagement

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website (https://ethics.coventry.ac.uk/) and that the application number is listed below (Note: Projects without an ethical application number will be rejected for marking)

Signed:  Date: 08/12/2021

Please complete all fields.

First Name:	Alexandra
Last Name:	Espialidou
Student ID number	13047171
Ethics Application Number	P164785
1st Supervisor Name	Dr. Mark Johnston
2nd Supervisor Name	Dr. Ganiat Kazeem

This form must be completed, scanned and included with your project submission to Turnitin. Failure to append these declarations may result in your project being rejected for marking.

Abstract

In recent times, China has emerged as an economic superpower and a world factory. However, its journey to economic prominence has not been a continuous trajectory throughout history. This report provides a comprehensive analysis of China's economic indicators, including GDP, inflation, trade values, and foreign direct investment (FDI), and their implications on a global scale. It covers China's economic history, from the Ming Dynasty to the present day, and examines the impact of key events, including the economic reforms of 1978, which marked China's transition from a closed, state-centric economy to an open, capitalistic market.

Furthermore, the report addresses the development and evaluation of ARIMA models for forecasting GDP, Inflation, trade values and FDI inflows in China, providing detailed insights into the process and the resulting predictions. The process involves manual model development with the assistance of the Dickey Fuller test, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Subsequently, the optimal models are identified utilizing the stats.model library in Python. The forecasting models are refined, and 10-year annual forecasts are presented for each dataset. Notably, the forecasted values suggest a positive trend for China’s economy with reduced inflation. Lastly, the SARIMAX (2, 2, 1) model for GDP emerges as the most accurate, achieving a 5% error.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to sincerely thank my supervisor, Dr Mark Johnston for their support and guidance, throughout this project and the entire course as a lecturer and supervisor. Their engagement, constructive feedback and insightful ideas have been instrumental in shaping the success of this report.

Furthermore, I would like to thank my parents, Cathrine and John for their unconditional support and belief in my capabilities throughout my academic journey. Without them, I wouldn't have been able to complete this project.

1 Introduction

China, officially known as the People's Republic of China, is a vast and diverse country situated in East Asia (Britannica, 2023). Encompassing a land area of approximately 9.6 million square kilometers, it ranks as the third-largest country in the world by total land area. As of 15th of November 2023, China stands as the second most populated nation on the planet, with a population exceeding 1.4 billion people, as per recent estimates (Worldometers.info, 2023). China's recorded history spans over 4,000 years, fostering a rich cultural heritage that has significantly influenced various aspects of global civilization. It stands out for its longevity and resilience as a discrete politico-cultural nation, as well as one of the few nations that was prosperous both culturally and economically during the dawn of human civilization. Its cultural tapestry is woven with many diverse ethnic groups (such as Han, Zhuang and many more), each contributing to the country's unique identity. From the Great Wall to the Terracotta Army, China is renowned for its historical marvels that reflect its ancient civilization's ingenuity and architectural prowess.

Undoubtedly, China has emerged as an economic superpower, as highlighted by Cordesman (2023). The rapid growth of China's economy has not only been a national phenomenon but has also exerted a profound influence on the global economic landscape, accelerating pre-existing trends (Overholt, 2016). In a pivotal transition in 1978, China embraced a global market economy, marking a significant shift in its economic ways. This transition was accompanied by a newfound moral consciousness concerning the global supply chain. Over the years, China's success in implementing market reforms has been transformative, playing a crucial role in lifting billions of people out of poverty and serving as a catalyst for similar reforms in other nations. Chen (2018) emphasizes the significant transformation undergone by China's economy, attributing it to the liberalization of foreign direct investment (FDI) policies and their consequential impact on economic development. Remarkably, China's ability to transition to a service economy and assert dominance in finance within a short span is noteworthy (Overholt, 2016). The evolution from a traditional manufacturing-focused economy to one centred on services showcases China's adaptability and dynamism. Looking ahead, it looks like China's pivotal role in the global monetary order and its influence on supply chain dynamics will be instrumental in shaping the future of the global economy.

In essence, China's economic journey, marked by strategic reforms and global integration, underscores its profound impact on the global economic stage, with far-reaching implications for the future.

1.1 Problem Statement and Motivation

The motivation behind this dissertation stems from the critical need to unravel the complex indicators of China's economic dynamics and its implications on a global scale. China's economic metrics (for example GDP, trade values, inflation rate and incoming FDI) are not only of great importance to scholars and researchers but also hold practical significance for policymakers, businesses and investors operating worldwide. Understanding the factors that drive China's GDP growth, its patterns of inflation, its trade relationships and its foreign direct investment is vital for making informed decisions in an era where China plays a leading role in the international economic system.

Moreover, as China continues to undergo rapid economic transformations, there is a pressing need for a comprehensive analysis that investigates these economic indicators and the events that led to their trends. This research aspires to contribute to the existing body of knowledge by providing a nuanced understanding of how these metrics are interconnected and how changes in one aspect may reverberate across the entire economic spectrum. This research will also involve historic events that have influenced China's economy, either domestically or internationally.

The findings of this dissertation are expected to offer valuable insights to scholars, policymakers and businesses, aiding them in understanding and navigating the complexities of China's economic landscape and making informed decisions in an era of heightened global economic interdependence.

Lastly, incorporating a forecast for the next 10 years using the Autoregressive Integrated Moving Average (ARIMA) model would significantly enhance the practicality of this research (Eissa, 2020). By employing the ARIMA model the study will extend beyond the historical and economic analysis and offer insights into the projected trajectory of these key economic metrics for China. This forecast will be indeed useful to stakeholders and businesses with a forward-looking perspective of investing in China. In this way they can anticipate potential economic trends, trade dynamics and market fluctuations. This component will add a proactive dimension to the research, providing a more practical application and relevance into the coming decade.

All in all, this research seeks to bridge the gap between theoretical understanding and practical implications, shedding a light on the nature of China's economic metrics and their broader significance.

1.2 Project Objectives

The main aim of this report is to provide an explanation of China's economy and unravel the political, social and global factors than have influenced them. This will be done through a thorough analysis of various indicators, an exploration of the trends they reveal and a literature review of the reasons behind these patterns.

The analytical approach that will be followed is an uncommon methodology, with the literature review included in the analysis part. It intertwines graph-based insights with specialised literature. We intend to construct graphs, based on these metrics and subsequently interpret them. This process will be enriched by referencing relevant academic articles and journals authored by specialists in the field, where we will draw insights from existing literature in order to understand the underlying factors contributing to these economic trends. This approach not only ensures a data-driven analysis but at the same time ensures that the interpretations are informed by the expertise of scholars who have explored similar economic dynamics.

The indicators will be analysed separately and all together (in Section 2.7) offering a complete comparative economic analysis based on economic time series data.

The main questions that this research report will address are:

Evolution of the Chinese Economy:

1) How has the Chinese economy evolved over time? How have the macroeconomic trends evolved?

Long-Term Economic Trends and Correlations:

2) Are there any long-term economic trends in China and do they correlate with significant global events? Are there any trends that cannot be explained? (In case of unexplained by known events economic disruptions, this project will delve deeper into the political and social context of those periods).

Recurring Patterns and Cycles:

3) Are there any recurring patterns or cycles in the Chinese economy, and if so, what insights do they offer into economic dynamics?

Predictive Capabilities of Economic Analysis:

4) Can time series forecasting through the ARIMA model, trade analysis and advanced modeling techniques contribute to predicting future economic trends, trade relationships and financial market disruptions in China? What are their predictions and can they be trusted?

1.3 Overview of This Report

The project is divided into two parts. The first part will cover the economic analysis of five key economic indicators: GDP, inflation, imports, exports and FDI, by examining the graphs derived from them. The main objective is to interpret these visual representations in light of insights provided by other authors in the field. In this way, economic trends will be paired with historic facts, providing a comprehensive understanding. The second part will include 10-year forecasts for all five indicators.

Part one includes Chapter 2 (2.1 – 2.7). Section 2.1 initiates the analysis by exploring the economic history of China before the 21st century with a comparative GDP analysis from 1500 until today. In section 2.2 the Chinese reforms, which have played a transformative role in China's economic evolution, will be analysed. Sections 2.3, 2.4, 2.5 and 2.6 assess key economic indicators, namely GDP, inflation, trade and Foreign Direct Investment (FDI), respectively, and break down any trends and correlations with important events. These analyses are complemented by insights drawn from existing literature. In the end, in section 2.7, a comparative analysis of all metrics is provided.

In Part two, which includes Chapter 3 (3.1 - 3.5), the focus shifts to forecasting values for all metrics through the application of ARIMA models. In the initial phase of the analysis, a manually constructed model will be developed based on the results of the Augmented Dickey-Fuller test and an exploration of the Autoregression and Partial Autoregression Functions for each metric. The dataset will then be split into 0.88 training datasets and automated models with help from the auto_model() function will be built, in order to provide a thorough evaluation of the use of ARIMA models for forecasting economic values. In the end, an evaluation of all models will be conducted, which can be found in section 4.

Section 5, includes the project management, section 6 discusses the Social, Legal, Ethical and Professional Considerations of the report and section 7 presents the conclusions and future works.

2 Economic Analysis

2.1 The Chinese economy through its GDP

This section will analyse China's economy through its Gross Domestic Product (GDP), as GDP is the mostly used measure for economic growth (Logubayom et al., 2013).

This timeline provides information about key events that have influenced China's GDP value over the years and are analysed through this section.

Gross Domestic Product

Date	*Key event*
1368- 1644	Ming Dynasty
1644-1912	Qing Dynasty
1700-1830	Population growth
1840-1860	Opium Wars
1851-1864	Taiping Rebellion
1861-1895	Self-Strengthening Movement
1912	Republic of China
1919	The May forth movement
1949	Declaration of People's Republic of China
1958-1962	The Great Leap Forward
1966-1976	The Cultural Revolution
1978	Chinese economic reforms
1979	Reform on People's Bank of China
1989	The Tiananmen Square incident
2007-2008	Financial Crisis

2.1.1 Economic history of China through GDP

Over the past few decades, China has undergone a dramatic economic transformation, becoming one of the world's leading economies (Maddison, 2006). As per Maddison: "in world perspective, China's performance has been exceptional". With its rapid industrialization and technological advancement, China has secured a prominent position on the global stage, profoundly impacting international trade, investments and geopolitical dynamics. The dynamic interplay between its rich historical legacy and its contemporary economic and technological advancements have established China as a key player in shaping the world's future.

The matter of Gross Domestic Product (GDP) has emerged as a primary concern among macroeconomic variables. Data pertaining to GDP is considered a crucial indicator for evaluating national economic progress and for assessing the overall operational status of the macroeconomy (Ning et al, 2010). It is widely considered as the premier metric for measuring the overall performance of the economy. Data on GDP is considered a crucial indicator for evaluating national economic progress and for assessing the overall operational status of the macroeconomy. There are many variations of GDP, all counting different aspects of a country's economy, with different focuses that have a substantial impact on the outcome (OECD, 2023b). Before starting this analysis, it is beneficial to differentiate between the terms.

Nominal GDP is good for measuring current prices, without adjusting for inflation, representing the total value of goods and services manufactured during the present reporting period (OECD 2023b).

On the other hand, *GDP Per Capita* calculates a country's gross domestic product and then divides it by its population (Investopedia, 2022). In this way the country's standard of living is shown.

A different method, called *GDP at Purchasing Power Parity (PPP)* accounts for differences in the cost of living and inflation that exist between countries. In this way, the purchasing power of different countries is equalised (OECD, 2023a). Calculating based on PPP has a large influence in the comparison of wealthy to developing counties. Figure 1 shows a comparison of Nominal GDP vs GDP at PPP for USA, China, Japan, Germany, India, UK and France (World Bank, 2022) in trillions of US$.

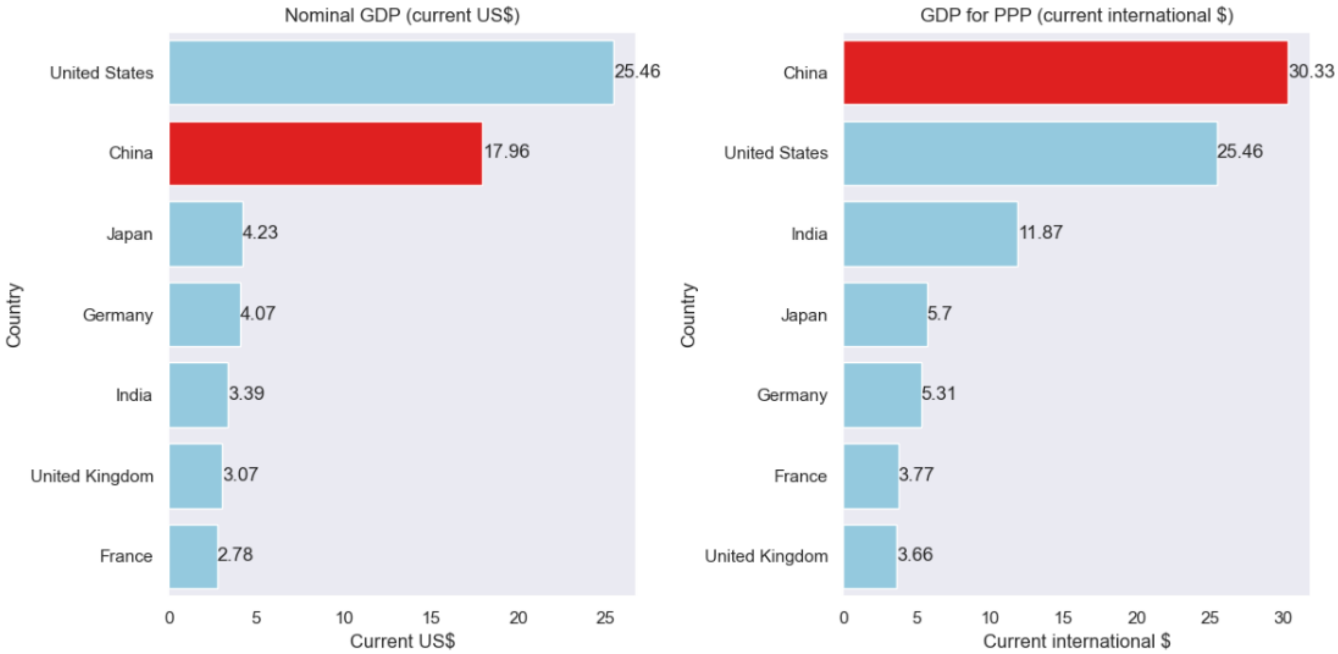

Figure 1: Nominal GDP and GDP at PPP for China (2022)

According to Figure 1, China had the second-largest nominal GDP, trailing the US in 2022. However, when expressed in terms of purchasing power parity, China's GDP surpasses that of the United States by a margin nearing $5 trillion in current international valuation. Consequently, this discrepancy indicates that the selection of the metric significantly shapes the narrative of any analysis. Now that the difference between GDP variations has been established, the Global Share of GDP at PPP from 1500 will analyses in section 2.1.2.

2.1.2 Global Share of GDP at PPP from 1500

As a starting point, it would be beneficial to get a first look at the trajectory of GDP for some of the biggest economies in the world. Table 1 presents GDP adjusted for purchasing power parity (PPP) of China, USA, India, Japan, Germany, Russia, UK, France and the rest of the world as a percentage of the global GDP (World Bank, 2020). The data spans from 1500 to 2020, providing insights into the economic performance and growth trajectories of these countries.

Table 1: Percentage of share of GDP at PPP globally (1500-2020)

Year	China	USA	India	Japan	Germany	Russia/USSR	UK	France	Rest of World
1500	24.89	0.32	24.36	3.10	3.32	3.41	1.13	4.39	35.07
1600	28.97	0.18	22.41	2.90	3.82	3.45	1.81	4.70	31.76
1700	22.31	0.14	24.46	4.15	3.68	4.36	2.89	5.27	32.74
1820	32.96	1.81	16.07	2.99	3.87	5.43	5.22	5.11	26.53
1870	17.10	8.87	12.15	2.29	6.50	7.54	9.03	6.50	30.03
1900	11.06	15.85	8.64	2.64	8.23	7.81	9.37	5.92	30.47
1913	8.83	18.93	7.47	2.62	8.68	8.50	8.22	5.29	31.45
1950	4.59	27.29	4.16	3.02	4.97	9.56	6.52	4.13	35.75
1960	5.24	24.27	3.88	4.45	6.62	10.00	5.37	4.09	36.09
1970	4.63	22.39	3.41	7.36	6.12	9.82	4.35	4.30	37.61
1980	5.20	21.12	3.18	7.83	5.52	8.53	3.64	4.06	40.92
1990	7.83	21.39	4.05	8.55	4.66	7.33	3.48	3.78	38.93
2000	11.77	21.89	5.18	2.11	4.24	3.51	3.30	3.40	44.59
2010	13.88	16.71	6.00	5.00	3.64	3.59	2.53	2.61	46.04
2020	18.21	15.68	6.76	3.99	3.39	3.08	2.27	2.26	44.36

Based on GDP adjusted for purchasing power parity

From 1500 up until 1900, China held the position as the largest economy globally, closely trailed by India, as seen in Table 1. In 1500, China's GDP accounted for an impressive 24.89%, while the combined GDP of the rest of the world stood at 35.07%. However, a pivotal shift occurred after 1820, as China's GDP began to witness a downward trend, ultimately leading to the United States surpassing China by 1900, claiming a share of 15.85% compared to China's 11.06%. Strikingly, India's economy also experienced a decline during the same period, mirroring China's economic struggles. Throughout the 20th century, China's economic trajectory appeared to be marked by challenges, in stark contrast to the continuous rise of the United States' economy, as Table 1 suggests. Notably, from 1980 onward, China's economy exhibited a rapid upward climb, reflecting a significant shift in its economic landscape. By 2020, China's GDP surged to 18.21%, indicating a remarkable resurgence, with the United States coming in second position with a GDP share of 15.68%. Interestingly, data suggests that the US economy has been undergoing a negative shift since 2010, adding a new layer of complexity to the global economic dynamics.

In light of this information provided, it is interesting to understand the reasons behind China's former status as an economic powerhouse, starting from the Dynasties that governed China and commanded the economy, as examined in sections 2.1.3, 2.14 and 2.1.5. Figure 2 showcases the percentage share of GDP mentioned in Table 1 in a graph format, offering a comprehensive visual representation of GDP at PPP distribution. The code for Figure 2 and all other figures of this report can be found in Appendix C.

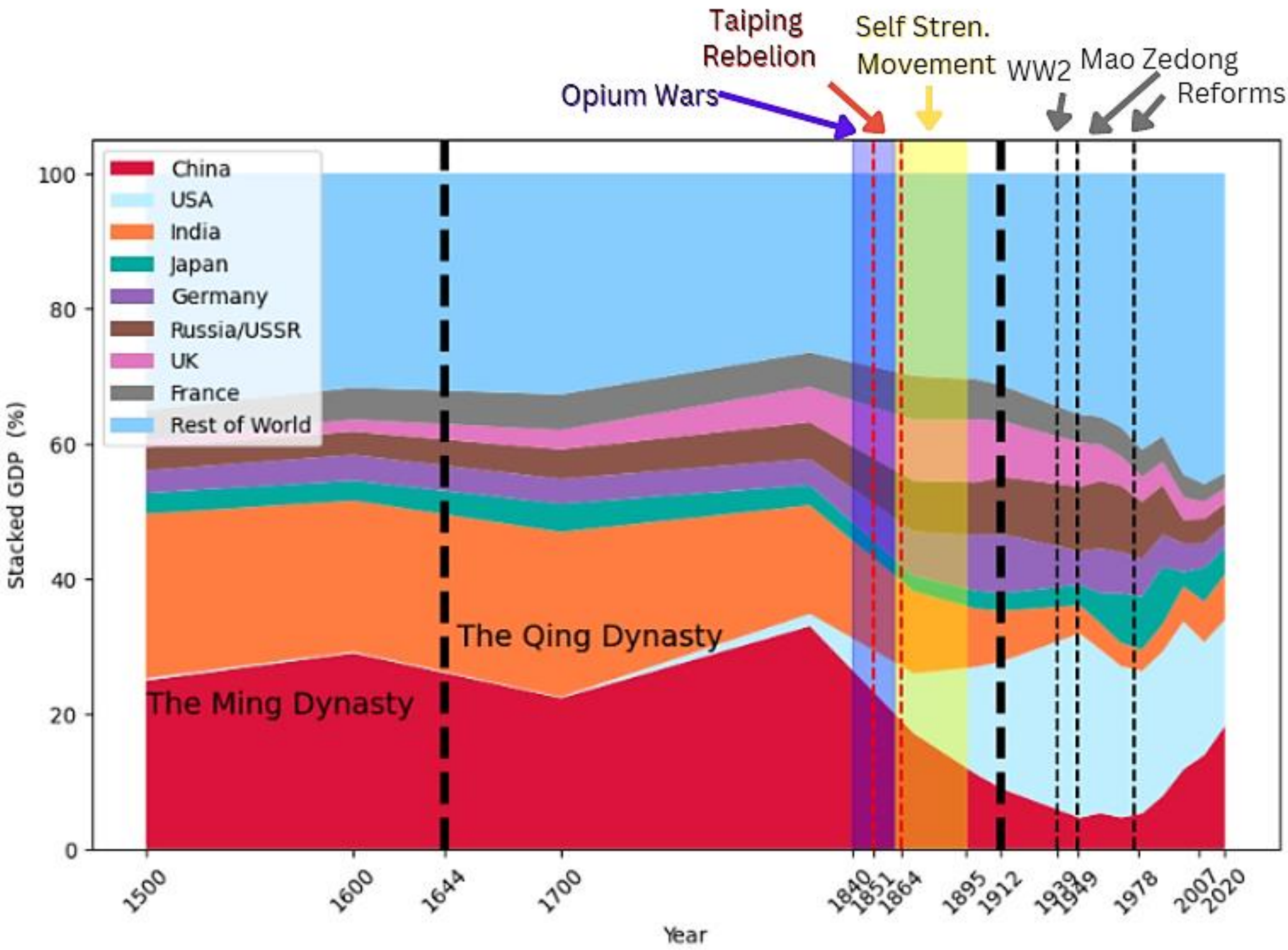

Figure 2: Share of Global GDP at PPP (1500-2020)

2.1.3 The Ming Dynasty (1368-1644)

During the period from 1368 to 1644, the Ming Dynasty reigned over China presiding over an era characterized by notable economic prosperity (History.com Editors, 2023), culminating in China's GDP leading the world in GDP value with a share of 28.97% in 1600, as seen in Figure 2. During this time, agriculture served as the primary livelihood for the population and the discovery and integration of new crops, such as sweet potatoes, peanuts, tobacco and maize, significantly augmented the nutritional landscape and the economy of the time. Moreover, this decade introduced advanced agricultural innovations, such as irrigation systems and land reclamation projects (Maddison, 2006). These innovations played a pivotal role in bolstering China's agricultural output, facilitating increased production of essential goods and reinforcing the foundation of the flourishing economy.

Another sector that benefitted China’s economy during that time was handicrafts and manufacturing. Silk weaving, porcelain and other traditional crafts were indeed popular (Maddison, 2006). Chinese porcelains became widely known during that period, a good that facilitated the growth of exports. Furthermore, railways, industrial production, commerce and banking contributed significantly to China’s modern sector’s growth at that period. It is noteworthy that China’s naval fleet comprised 2,700 patrol vessels and combat ships, along with 400 large warships, 400 grain transport freighters and nearly 300 exceptionally large "treasure-ships" designated for expeditions to the Western Oceans. These colossal vessels, intended for maritime exploration, were five times the size of any ships commanded by Vasco da Gama, the Portuguese admiral who initiated European trade with Asia by circumnavigating Africa in the late 15th century.

2.1.4 The Qing (Chi'ing) Dynasty (1644-1912)

The next dynasty that ruled over China and impacted their economy was the Qing Dynasty, between 1644 and 1912 (History.com Editors, 2023). Even though they continued many of the previous economic practices, they also encountered several wars and conflicts, both internal and external, including the population growth, natural disasters and the Opium wars.

China population has experienced two growth spurts, one during 960-1127 and one during 1700-1830 under the Qing dynasty. During the second growth spurt, the Chinese population rose from 55 to 400 million people, a 1.50 percent annual growth rate (Perkins, D.H., 2013). This growth rate demonstrated greater sustainability, fundamentally altering the demographic trajectory of China. Many researchers, such as Maddison, Gang and Zhongyi (2006), have taken China's population size during the Qing dynasty as a surrogate indicator for the scale and vitality of the economy (Deng & Shengmin, 2019). This leads to a circular argumentation that a large population was sustained by a robust economy and a large economy supported a substantial population.

Moreover, The Opium wars (1840-1860) and the Taiping Rebellion (1851-1864) took a substantial toll on the Chinese economy (History.com Editors, 2023). After an economic exploitation from the British and a big rebellion, China's agricultural and trading sectors were profoundly affected. From that period on, China's economy took a big plunge as illustrated in Figure 2 with these events marked in blue highlighting and in red vertical lines. However, around 1850 all countries seem to have faces difficulties leading to reduced economic power.

Finally, it is noteworthy that between 1861 and 1895, the Qing government launched a self-strengthening movement (marked in yellow on Figure 2), endeavoring to integrate Western practices into Chinese governance (Britannica, 2022). However, the movement encountered significant setbacks attributed to opposition and resistance from the Chinese population. The Chinese economy can be seen to be shrinking and losing the big momentum that it had previously sustained until the late 1900s and the implementation of the reforms leading to a renewed and rapid growth trajectory.

In summary, as illustrated in Figure 2, these significant events during the Qing Dynasty appeared to have intensified the decline of the Chinese economy, contributing to a more noticeable downward trajectory.

2.1.5 The Republic of China and the Declaration of the People's Republic of China

The fall of the Qing Dynasty in 1912 lead to the establishment of the Republic of China, signaling the end of the millennia-old imperial system. However, the events that took place in 1911 were merely the beginning of a long revolutionary struggle that would span several decades, ultimately concluding in 1949. During this period, the leaders of the newly formed Republic of China faced significant challenges in their attempts to unify the vast and diverse nation. Despite their efforts, they were able to assert control effectively only in one province, Nanjing, highlighting the complexity of the political landscape and the persistent regional divisions (Office of the Historian, 2019).

As the years unfolded, the political landscape in China continued to evolve, while the Republic of China struggled to consolidate power and faced opposition from various factions within the country. These struggles could be the reasons behind China's sharp decline in GDP during that period, as illustrated in Figure 2. In 1949, amidst the backdrop of a protracted civil war, the Communist Party, led by Mao Zedong, declared the establishment of the People's Republic of China, according to Philips (2016). This declaration marked the definitive end of the Republic of China's rule and the beginning of a new era under communist leadership.

The establishment of the People's Republic of China in 1949 not only marked a significant shift in China's political landscape but also had profound implications for the global balance of power during the Cold War (Britannica, 2023c). The events leading up to and following the 1911 revolution were instrumental in shaping modern China, with enduring repercussions that continue to influence the nation's trajectory in the 21st century.

While examining Figure 2, an upward trend becomes evident around the year 2000, indicating a period of notable economic growth across various countries. China, in particular, appears to be experiencing a rapid and sustained expansion during this timeframe. However, a significant inflection point is observed in 2007, closely aligning with the onset of the Financial Crisis that affected the globe (Weinberg, 2013). Interestingly, China's GDP continues to rise as the other countries in the Figure appear to be impacted by the crisis.

Overall, China emerged as a leading force in global economic growth across the second millennium, primarily driven by its remarkable productivity in agriculture. By looking at Figure 2 one last time, it is evident that China was maintaining a leadership position until the 19th century, where industrialization gave the US the lead. Various historic events led China to experience a downward trend faster than the rest of the countries. Nonetheless, after 1978, while the rest of the countries show a reduced GDP, China's seems to be rapidly ascending, reclaiming the leading position once again by 2020.

Lastly, it is important to note that GDP is not a definitive indicator of economic size and power. It provides a snapshot of the total value of goods and services produced within a country's borders, serving as a fundamental tool for comparing economic performance between different nations. Yet, it does not capture factors such as income distribution, quality of life or other key socio-economic aspects that contribute to a comprehensive understanding of a country's overall well-being and development

2.2 Chinese reforms

The Chinese reforms of 1978 marked China's pivotal transition from a closed, state-centric economy to an open, capitalistic economic framework. This transformative initiative stands as a paramount historical event, as it effectively opened China's economy to global engagement, thus laying the foundation for its resurgence as a formidable economic power.

Before delving into the intricacies of the Chinese economic reforms that facilitated its ascent into a dynamic economy, it is important to comprehend the contextual circumstances that were taking place during that period. These circumstances underscore the necessity and significance of the undertaking of these reforms by China.

2.2.1 Mao Zedong and the need for reforms in China

Mao Zedong, Chairman of the Chinese Communist Party (CCP), was brought to power in 1949 up until his death in 1976 (Britannica, 2023). On 1 October 1949, Mao declared the foundation of People's Republic of China (PRC) (Philips, 2016) with views heavily influenced by Marxist-Leninist ideologies, but he adapted them to suit the specific conditions of China. His policies included the Great Leap Forward (1958-1962) and the Cultural Revolution (1966-1976), which had significant and far-reaching consequences for China's social, political, and economic landscape. His leadership style and policies led to both significant advancements and catastrophic setbacks for China.

As Zhang and Clovis (2023) recount, prior to 1978, China was a central economy, where a big part of the country's economic output was largely controlled by the government. The CCP would set production goals, control the prices of goods and services, set limited trade and in general wanted to achieve a relatively self-sufficient economy. To achieve this goal, the government invested heavily into the rapid industrialization of China. In order to reach their set targets, the majority of industrial production was generated from centrally controlled, state-owned enterprises.

As for Foreign Direct Investments, they were generally frowned upon over this period. Foreign trade was also limited. The government played a dominant role in economic decision-making, while the focus was on achieving specific production targets rather than responding to market forces and only goods that they were unable to be produce or obtained were allowed (Zhang and Clovis, 2023).

The majority of economic structures in China during that time (1949-1978) were predominantly overseen by the central government as market mechanisms for the efficient allocation of resources were notably absent. This meant there was no motivation for productivity for workers, farmers or firms. The people's primary focus was the productivity goals set by the government, thereby concern for quality of a product's improvement was not a major one. Under these circumstances, the Chinese economy was at a level point. These practices held China back, while the rest of the world was evolving.

There were several attempts to revive the economy at that period, like the Great Leap Forward and the Cultural Revolution by Mao Zedong, as mentioned in section 2.1. However, they were unsuccessful.

2.2.2 Chinese Economic Reforms

The '*Chinese Economic Reforms'* is the term that was given to China's reforms in 1978. This was a pivotal moment in the country's economic history (Bettelheim, 1988). Their aim was the transition China from a centrally planned and closed economy to an open-market economy, open to trade and capitalism-oriented. Below are several crucial aspects of these reforms from an economic point of view.

Agricultural Reforms

A major emphasis was placed on the process of dismantling collective farming in the agricultural sector as Perry and Wong (1989) report. Before the reforms, most (if not all) of the farms were collective. This meant that farmers were not at liberty to produce their own crops. With the decollectivization, the farmers were free to cultivate freely, which led to increased agricultural productivity throughout China and reduced rural poverty. Furthermore, state-owned enterprises implemented open-market reforms, which meant they were more independent and could operate withing the open-market.

People's Bank of China

As for the financial sector, the creation of a dual-track currency system was introduced, which promoted monetary stability (Walter, 1985). One system was implemented for domestic transactions and a different one for trade and foreign investments. The People's Bank of China, founded in 1948 and reformed in 1979, played a vital role in creating and executing monetary policies, overseeing financial institutions and ensuring financial stability

International trade

After the reforms, international trade and Foreign Direct Investments were once again encouraged. China opened up to the world and to the global economy allowing for a fast expansion of China’s exports and especially the manufacturing sector. Since China was a developing economy, they offered low-cost labor, which attracted many investors (Perry and Wong, 1989). To this also contributed the introduction of Special Economic Zones, such as Shenzhen, Zhuhai and Shantou. These zones, which operated as experimental before implementing policies to the rest of China, had free trade and enabled economic activities with more flexible policies.

Overall, China's reforms established the foundation for its impressive economic transformation, resulting in significant enhancements in standards of living, decreased poverty rates and the incorporation of millions into the worldwide economy. According to David Mann, global chief economist at Standard Chartered Bank, "From the end of the 1970s onwards we've seen what is easily the most impressive economic miracle of any economy in history" (Harrison and Palumbo, 2016).

2.3 Analysis of China’s GDP

In 1978, when the reforms were introduced, China’s economy was at an all-time low (Bettelheim, 1988). Compared to the pre-reform period, China's economic growth has significantly accelerated and has generally managed to steer clear of significant economic upheavals. In Figure 3, China’s GDP values after the reforms are presented (Macrotrends, 2023). From 1978, where China’s GDP was $149.54 billion of US, to 2022, where their GDP was $17,963.17 billion of US, their annual growth averaged 9.23 %. This indicates that the Chinese economy has experienced consistent expansion and rapid economic growth. This sustained growth has likely led to improvements in living standards, poverty reduction and the overall economic well-being of the Chinese population (Zhang and Clovis, 2023). Additionally, this level of economic expansion may have positioned China as a significant player in the global market, contributing to the country's increased influence and prominence on the international stage.

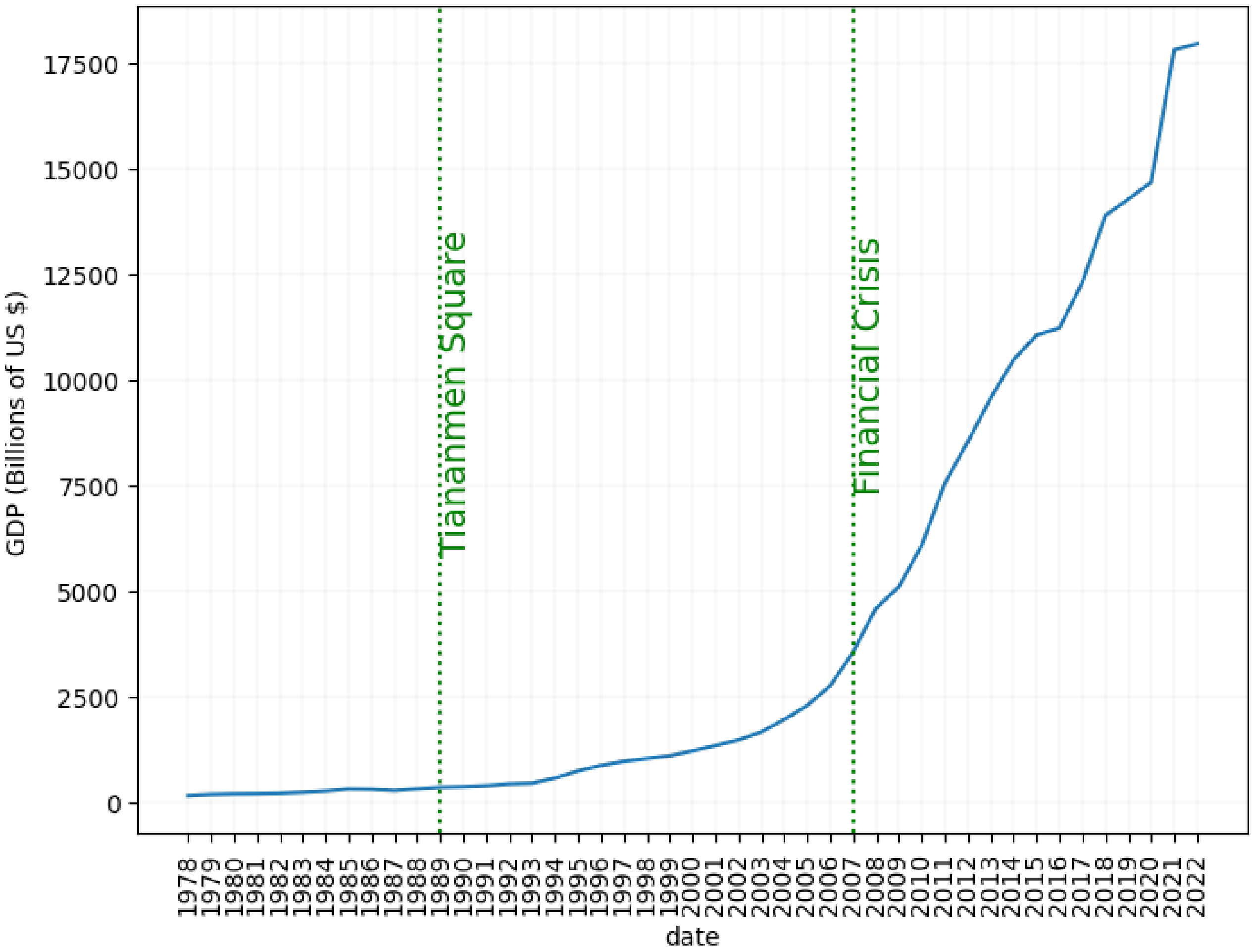

Figure 3: Chinese GDP (Billions of US $) (1978-2022)

Figure 4, with data also sourced from Macrotrends (2023), illustrates the annual percentage change in China's GDP (1978-2022). While Figure 3 exhibits a consistent and persistent upward trajectory in its GDP, Figure 4 indicates intervals of stagnant growth and fluctuations. Furthermore, there were many events that effected the Chinese economy and their influence can be observed in Figures 3 and 4.

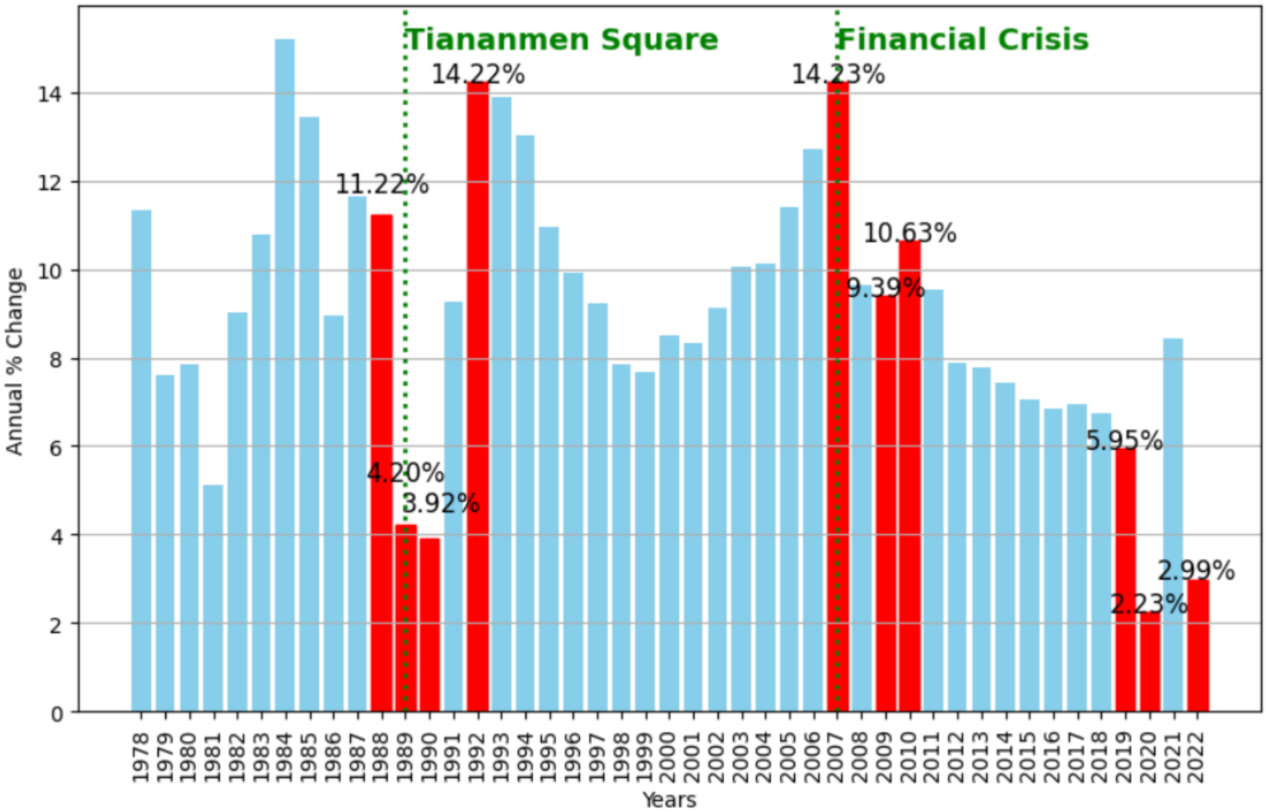

Figure 4: Annual % Change of GDP of China (1978-2022)

2.3.1 The Tiananmen Square incident

After the reforms, in the spring of 1989 (marked in green in Figures 3 and 4), a series of protests and gatherings took place in China, known as The Tiananmen Square incident or June Fourth incident. (Britannica, 2023). These protests reached a climax on the night of 3rd June 1989 when the government forcefully suppressed the demonstrators in Beijing's Tiananmen Square. While similar protests and responses unfolded in various cities across China, the focus remained on the events in Beijing, particularly in Tiananmen Square, which had historical associations with prior protests such as the May Fourth Movement of 1919. The May Fourth Movement was a similar revolution connected with national independence, rebuilding society and culture. On 4th May 1919, over 3,000 students from 13 Beijing-based colleges collectively protested the decision of the Versailles Peace Conference, responsible for drafting the treaty that formally concluded World War I, which involved the transfer of the former German concessions in Shandong province to Japan. Following the imposition of trade sanctions by various nations because of this incident, including the United States (Hu, 2016), China's progress in economic reforms was halted. This interruption led to a significant drop in China's actual GDP growth rate from 11.3% in 1988 to 4.2% in 1989 and further down to 3.9% in 1990, as seen in Figure 4. Regardless, by 1991, China restarted its economic reforms, leading to the reduction or removal of foreign sanctions. Consequently, the country experienced a revival in its real GDP growth, reaching 14.2% the next year.

2.3.2 Financial crisis

Another economic event that had significant implications on the Chinese economy was the 2007-2008 economic crisis (marked in green in Figures 3 and 4). While China's financial system was not as directly affected as those of some Western countries, it still experienced notable consequences (Li et al., 2012). The 2008 global financial crisis, one of the most significant economic incidents since the Great Depression (1929-1939), was caused by the collapse of the housing market in the United States. Unsustainable subprime mortgage lending practices, combined with the securitization of these high-risk loans, created an unstable environment that eventually led to the failure of major financial institutions, like Lehman Brothers, Bear Stearns etc. (Weinberg, 2013). The crisis swiftly spread across the globe, impacting major economies such as the United States, the European Union and several Asian countries, including Indonesia, Republic of Korea, Malaysia and China (Park et al., n.d.). Banking sectors faced severe liquidity problems, stock markets plummeted, and unemployment rates soared, causing widespread economic distress.

The 2007-2008 financial crisis had a substantial negative impact on China's exports, leading to a decline in export growth and a contraction in the country's trade surplus. The crisis primarily affected the export sector in Asian countries (Tang, 2022). China, like Japan, South Korea and India were established as major global exporters at the time. Because of the economic downturns that the US and Europe were facing, these countries saw a big decline in demand for their products, which affected China's export-oriented industries, leading to layoffs and factory closures. This is expected as more than 40% of China’s GDP relies heavily on Western demand for products (Li et al., 2012). Furthermore, the crisis caused a crash in China's stock market, starting October 2007, wiping out more than two-thirds of its value. A similar story unfolded in the real estate market, where a bubble started to grow alongside China's booming economy as people were convinced that the real estate sector was safer to invest compared to banks. This affected China’s economic growth significantly. After having reached a 14.23% growth rate in 2007, as Figure 4 shows, it fell to 9.39% in 2009. This percentage may still seem large but it disrupted China’s economic rise.

The Chinese government's response to the crisis involved the implementation of an economic stimulus package amounting to $586 billion US in 2010. Chien (2008) recounts that the primary focus was on funding infrastructure projects and implementing more relaxed monetary policies to encourage increased lending by banks. These measures effectively helped China counter the adverse impacts of the significant global decline in demand for Chinese goods. Their GDP growth gradually declined over the following years, dropping 5.95% in 2019. In 2019, amid the covid-19 pandemic, which started from Wuhan, China (AJMC Staff, 2021), it led to China’s GDP rate to fall to 2.23% in 2020. According to Reuters (2023), the after covid growth rate has been one of the lowest in the past half decade, which was less than 3% in 2022. This was caused by Chinese officials relaxing and strengthening anti-virus measures.

2.4 Analysis of China’s Inflation rates

The next key metric that will be analysed is China’s Inflation Rates. This timeline provides information about key events that have influenced China’s inflation rates over the years and are analysed through this section.

Inflation

Date	*Key event*
130 BC -1453	Silk roads
1958-1962	The Great Leap Forward
1978	Chinese economic reforms
1988	Announcement of Economic reforms being implemented by all means
1992	South China Tour Speech
2007-2008	Financial crisis

The dynamics of China’s inflation are vital to both the stability of the domestic economy and in the global economic landscape. Between its fast industrialization and urbanization, China's inflation rate has fluctuated, driven by a complex interaction of factors including domestic consumer demand, international commodity prices and governmental policy. Due to systematic monetary policy changes implemented by the People’s Bank of China, inflation rates were reduced and kept at a low rate after the beginning of the 21st century, as Zhang and Clovis (2010) mention.

As mentioned in section 2.2, before the late 1970s, China was largely a closed and centrally planned economy, with limited interactions with the global market. The Chinese government's emphasis on self-sufficiency and state control of resources probably meant that economic data, including inflation figures, were not systematically recorded or made publicly available. However, following the initiation of economic reforms in the late 1970s and the subsequent shift toward a more market-oriented economy, China seems to have begun to prioritize the collection and dissemination of economic data, including inflation rates, as more data are present in academic publications.

2.4.1 Inflation in China before the reforms

According to Zhang and Clovis (2010), China went through long periods of inflation, but also long periods of price stability. Specifically, before 1949 under the Kuomintang (the Nationalist Party of China (NPC)) and in between 1961-1962 after the Great Leap Forward, the country experienced hyperinflation[1]. As Hsiao (1971) suggests, inflation rates in China in 1953 and 1956 reached around 10-15% and in 1961 there was a surge of 16.2% in retail prices along with a substantial spike of up to 260% in prices within the free market. Burdekin (2000) recounts that elevated levels of inflation are typically linked to both rapid rates of money growth and substantial budget shortfalls. Although deficits that are not financed by the central bank may not necessarily result in significant inflation, there is often a tendency to at least partially fund such deficits through monetary means. This tendency is particularly evident in nations with underdeveloped private financial markets, such as China at that time. According to the government (Zhang and Clovis, 2010), inflation in China is highly connected with disorder and economic instability.

2.4.2 Inflation in China after the reforms

China's dynamic evolution of inflation has seen notable booms and busts due to cyclical behavior since the early 1980s. This picture is generally confirmed by Figure 5, which shows yearly data of inflation rates in China from 1987 to 2022 (Macrotrends, 2023). As illustrated, it is evident that between 1989 and 2011 China experienced significant fluctuations in their inflation rates. The Chinese inflation rate had two notable peaks in the late 1980s and in the middle of the 1990s since the start of the economic reforms in the late 1970s. During the late months of 1988, after the Chinese government announced that the economic reforms would be implemented by all means, the Chinese were shocked and hoarded activities (Ishihara, 1990), leading to an increase in inflation by over 20%. This sudden rise can be seen in Figure 5. As a result, the CPC decided to postpone price reforms indefinitely.

[1] According to Cambridge Dictionary (2023) hyperinflation is “the condition where the price of everything in a national economy goes out of control and increases very quickly”.

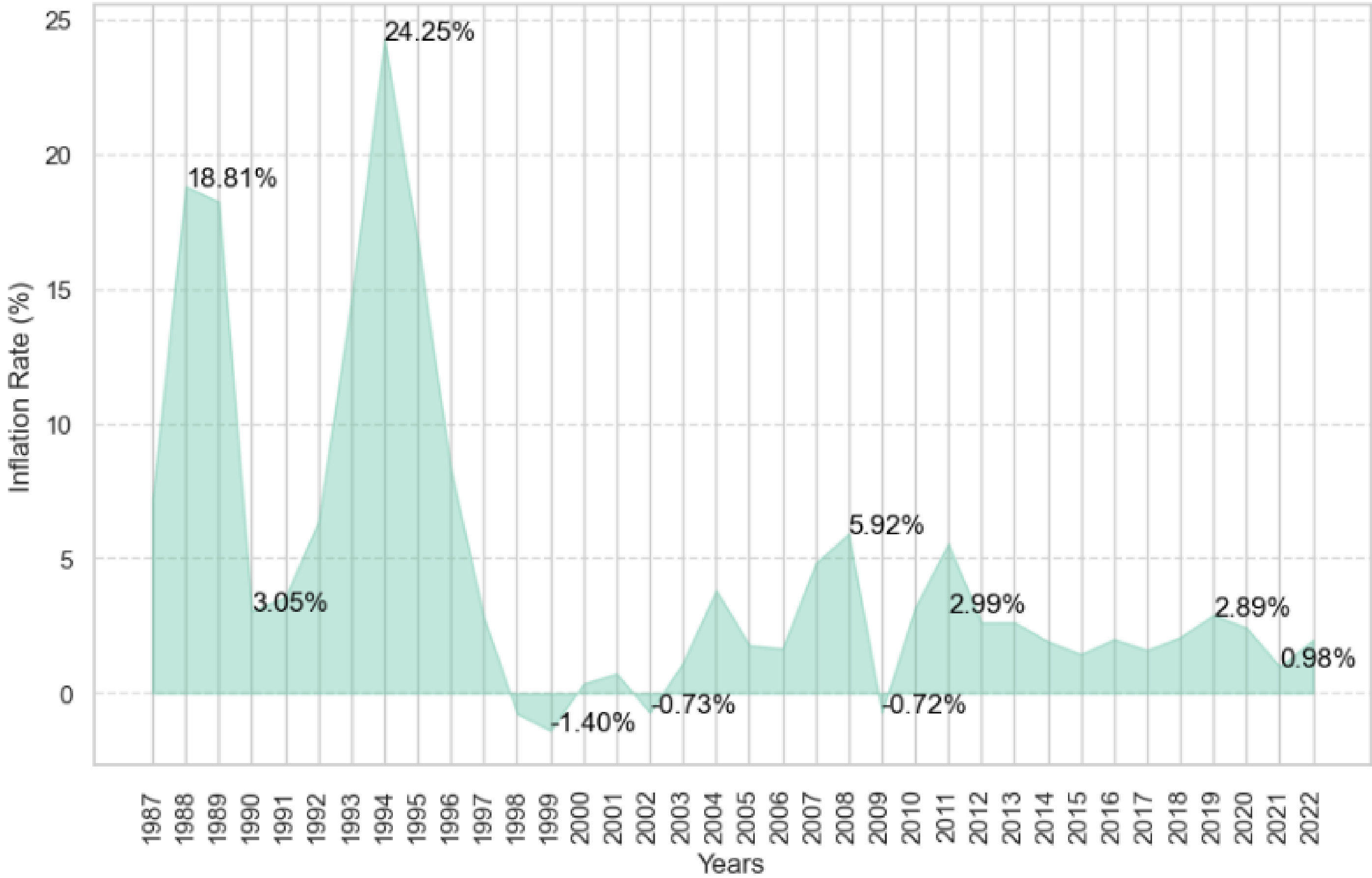

Figure 5: Inflation Rate (%) of China (1987-2022)

After the reforms, as the prices were liberated, this meant that the people were also not protected from high inflation anymore. This explains the sudden rise as people were trying to cope with the new system. Another reason why the inflation rate started rising post-reform is the Demand-pull Factor (Ishihara, 1990). As it is known, the Chinese economy had many shortages, which meant that demand was always higher that supply. When the prices were released, that extra demand, led prices up. All in all, the abolition of price controls, led to the emergence of long-suppressed inflation up to 18.81% in 1988. In response to the inflation peak, as Ishihara mentions, the central government decided to reduce fixed investments and tighten money and credit supply. This strict credit crunch generated liquidity problems for many companies in China, leading to reduced inflation (3.05%) over 1990-1992.

Another event that contributed to the fluctuation of the Chinese inflation, according to Zhang and Clovis (2010) was the landmark speech of "promoting Chinese economic development with all efforts" in 1992 (also known as South China Tour Speech) by the Chinese leader Deng Xiaoping, which encouraged investments, loosened credit controls and raised the growth rate of money supply. The proactive policy stance led Chinese inflation to increase in 1992 and peak in 1994 at 24.25%. After more tightening policies, inflation started to drop and actually reached -1.40% in 1999.

The last twenty-two years can be considered a low inflation period, with several deflation periods e.g., 1998–2000, 2001–2003, 2008-2010), even if there were two local inflation peaks in 2004 and 2007 as a result of temporary supply shocks (such as shocks to the price of food and energy) and demand shocks (such as shocks to the real estate market) (Zhang and Clovis, 2010). After the financial crisis of 2007-2008, China's inflation remained stably under 3%.

The period between 2000-2010 has witnessed accelerated globalization (Zhang, Song and Wang, 2015). China’s domestic markets were integrated with global markets regarding supply and demand conditions, which led to domestic prices and inflation to be interacted with supply-and-demand conditions in global markets, which allowed inflation rates to stay at a low point.

After 2000, China witnessed some fluctuations starting with a small deflation in 2002 at -0.73%. It swiftly rose at 5.92% 8 years after and dropped at -0.73% in 2009, during the financial crisis. When the Covid-19 pandemic started, in 2019 (AJMC Staff, 2021), inflation was at 2.89% and in the span of 2 years it declined at 0.98%. The data also suggest that there is an upward trend the following year.

2.5 Analysis of China’s trade

The next key metric that will be analysed is China’s trade dynamics, which includes the inflow of goods and services into China, as well as the outflow of goods and services to the global market. This timeline provides information about key events that have influenced China’s trade over the years and are analysed through this section.

Trade

Date	*Key event*
1960	Sino-Soviet breach
1978	Chinese economic reforms
2007-2008	Financial crisis

China's journey towards economic modernization has been closely intertwined with the evolution of its trade dynamics as Chen and Lieberthal (2019) mention. China has been described as a “world factory” (Hong et al., 2016), since Chinese companies acquire capital and intermediate goods from their trade partners and export finished products to various destinations worldwide. China’s foreign trade witnessed significant fluctuations from the 1950s onwards, with an initial dominance in trade with communist countries, such as North Korea and North Vietnam. However, after the fallout of the Sino-Soviet relationship in 1960 because of the Sino-Soviet breach, there was a marked shift towards fostering trade ties with non-communist nations, including Indonesia, Burma (Myanmar), Pakistan and Ceylon (Sri Lanka) in Asia, as well as Ghana, Algeria, Tanzania and Egypt in Africa. During this period, China actively extended substantial financial aid, grants and long-term interest-free loans to facilitate trade with these countries. Following the passing of Mao Zedong in 1976, and the economic reforms of 1978, China underwent a strategic reevaluation of its international aid initiatives, leading to an expansion of its trade with western countries. In 2021, China had been exporting 4,398 products to 214 countries and 4,391 products had been imported from 215 countries (World Integrated Trade Solution, 2018).

2.5.1 Silk Road

While discussing about China's trade, it is impossible to ignore the profound impact of the Silk Road, the historic trade routes that laid the groundwork for China's enduring role in global commerce. The silk road(s) or Silk route(s) was a multitude of short connecting routes, created from trading agreements, that started from Asia and reached all the way to Europe, to the eastern Mediterranean and the capital cities of the Roman Empire, as Berit (2017) explains. They were established during the Han Dynasty of China in 130 BC and lasted until CE 1453. They were also overland extensions to India, connecting East Asia with Africa via the eastern Mediterranean region. These routes were facilitated by maritime routes around Southeast Asia, ultimately ending in Venice or Genoa. Some of the goods that were exported were silk, paper, spices and other luxury goods. Though the silk road was created for trade purposes, religion, culture and political power also flew through them. Figure 6 shows a depiction of the Silk Roads at their full extend.

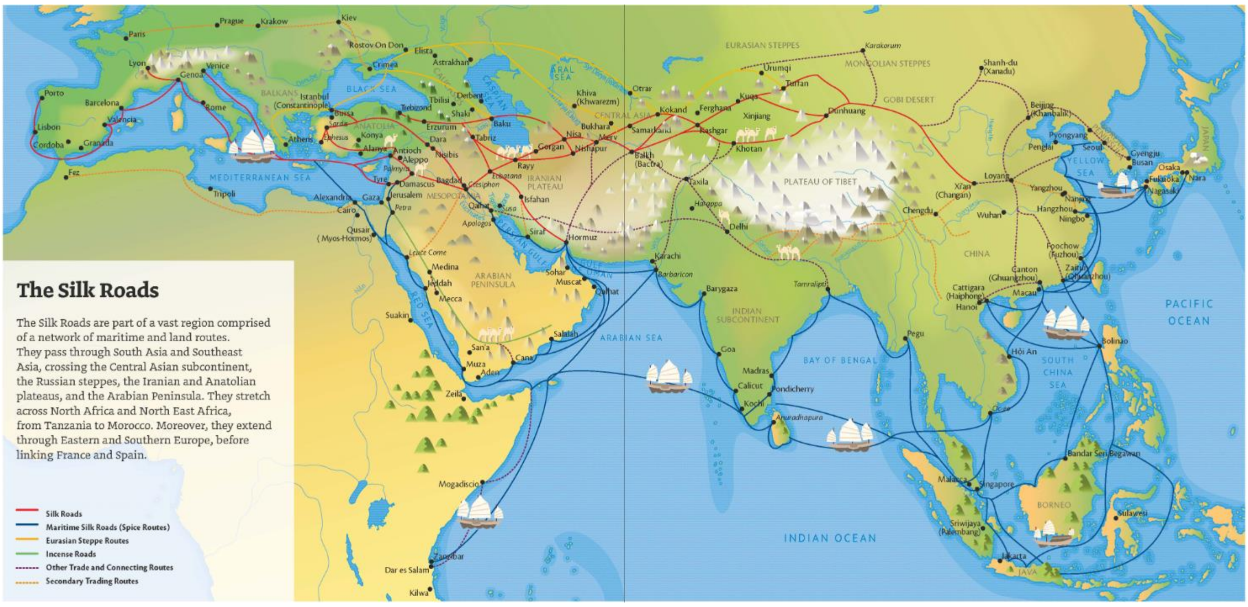

Figure 6: Map of the Silk Roads (UNESCO, n.d)

The establishment of the Silk Roads was not solely the result of amicable trade agreements (Hook, 2022). Over a span of 2000 years, the network of the Silk Roads experienced fluctuations, interruptions, adaptations, and resurgences, shaped by both periods of harmony and conflict. These shifts were influenced by various Chinese dynasties (Chinese Han dynasty (206 B.C.E.-C.E. 220), the Chinese Tang dynasty (C.E. 618-907), the Mongol Khanate (13th and 14th centuries)) (Miami University, n.d.), territorial expansions, the rise and fall of the Roman Empire and the emergence of Islamic dynasties in the Middle East and India.

During that time, there was a significant demand for Chinese silk from the Roman Empire (Hook, 2022). For this reason, Roman emperor Justinian (CE 527-565) ordered two emissaries disguised as monks to China in order to steal Chinese silkworms around CE 550. With their success the Byzantine silk industry was established. The silk was transformed into various items, including symbols, pillows and garments for the emperor, his court and wealthy Romans.

Despite the presence of native silkworms in the Mediterranean regions during the Bronze Age, the Roman upper class showed a preference for the superior quality textiles produced by the Seres, also known as the "Silk People" of China, as it was considered a luxurious symbol. Nevertheless, concerns were raised by figures like Pliny the elder (1885), who pointed out that the Chinese were eager to sell their wares to Rome but were reluctant to purchase Roman goods. This discrepancy led Pliny to estimate that Rome was losing "a hundred million sesterces" annually in transactions with India, Seres and the Arabs, with minimal financial returns to the Roman treasury.

2.5.1.1 Tea and Chay

One interesting fact about the silk road has to do with tea. The world has two words for tea: one is the English word *tea* and the other is a Hindi/Arabic variation *cha* or *chay* (The Language Nerds Team, 2019). Both words originated in China but the interestingy fact is how the were spread throughout the world. The variations of the "cha" term traveled across land through the network of the Silk Road, while the versions of "tea" were transmitted across water by Dutch traders who introduced tea to Europe.

2.5.1.2 One road one belt

Bruce-Lockhart (2017) adds that in 2013, China's president Xi Jinping announced the creation of a new double trade route, named "One Belt, One Road". According to a plan released in 2015, the roads will cover both land and maritime, reopening old channels between China and its trade partners, such as Central and North Asia and Europe. This new trade road will act as a "new era of globalization" and according to China, it will allow them to broaden their trade partners and facilitate transport.

2.5.2 China's trade during the 21st century

By the 21st century China has evolved into "the world's factory" (McKinsey Global Institute, 2019). Jahn (2021) narrates that in 2013, China surpassed all other nations globally in terms of the value of its imports and exports, a position previously held by the United States. Their exports grew dramatically after the opening of the country in 1978 under Deng Xiaoping, while their economic boom and the expansion of exports were largely attributed to the country's special economic zones, like Shenzhen, Macao and Hong Kong, where tax breaks were provided to foreign investors. One of these benefits was the tax-free importation of machinery and technology. Moreover, these trade and investment liberalization have made China a significant trading power.

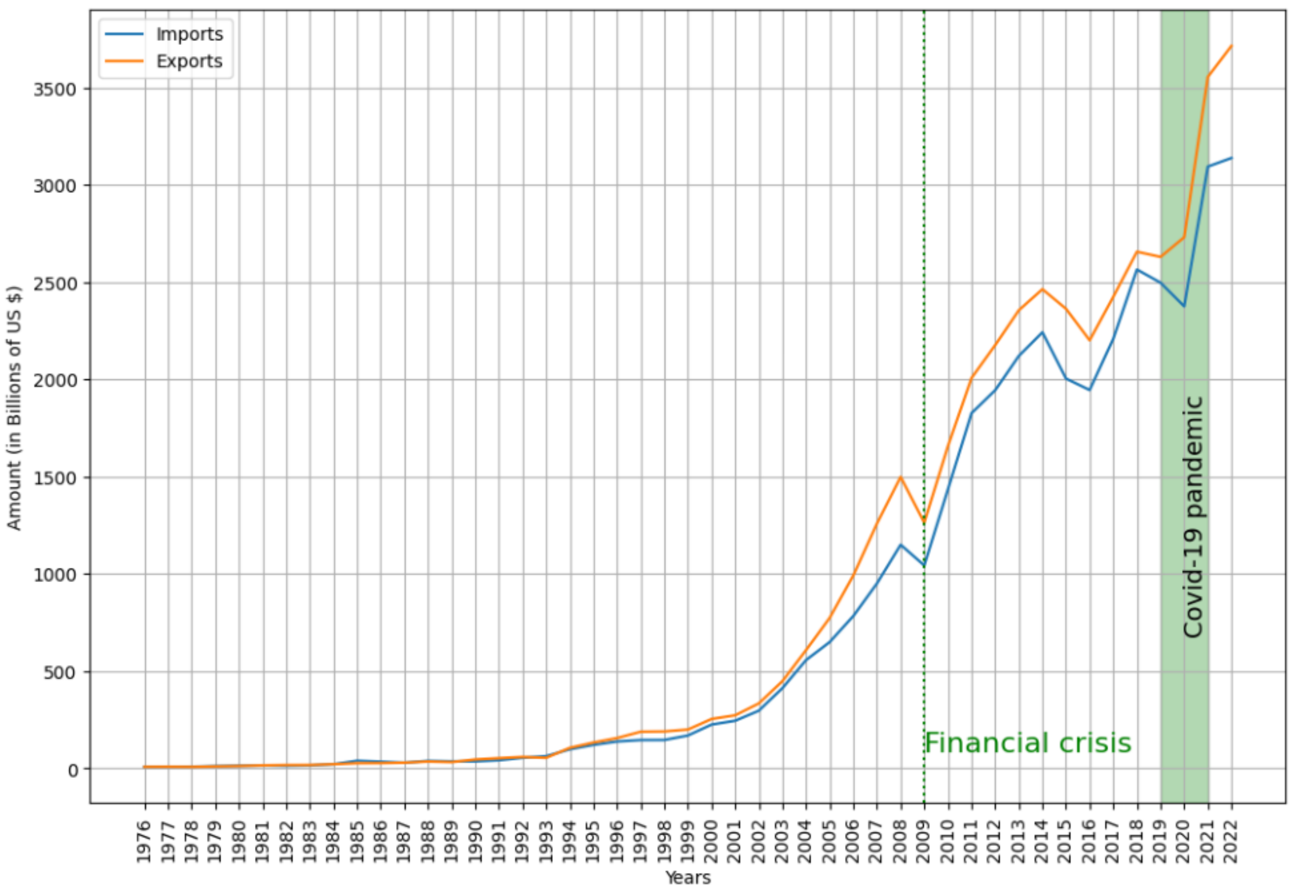

Figure 7: Trade values of China (1976-2022)

Figure 7 shows China's imports and exports in US dollars from 1976 to 2022 (Macrotrends, 2023). Figure 8 indicates the percentages of change between these values. Accorrding to Figure 7, while product imports increased from $10.53 billion to $3,137.59 billiion from 1979 to 2022, Chinese merchandise exports increased from $9.2 billion to $3.7 trillion. This means a 29,231.61% increase in imports and a 40,100% increase in exports in a span of almost half a century. Due to their rapidly expanding trade flows, China has become a key trading partner to many countries. According to China, in 2013, they were the largest trading partener for 130 countries (Zhou Mingwei, 2014).

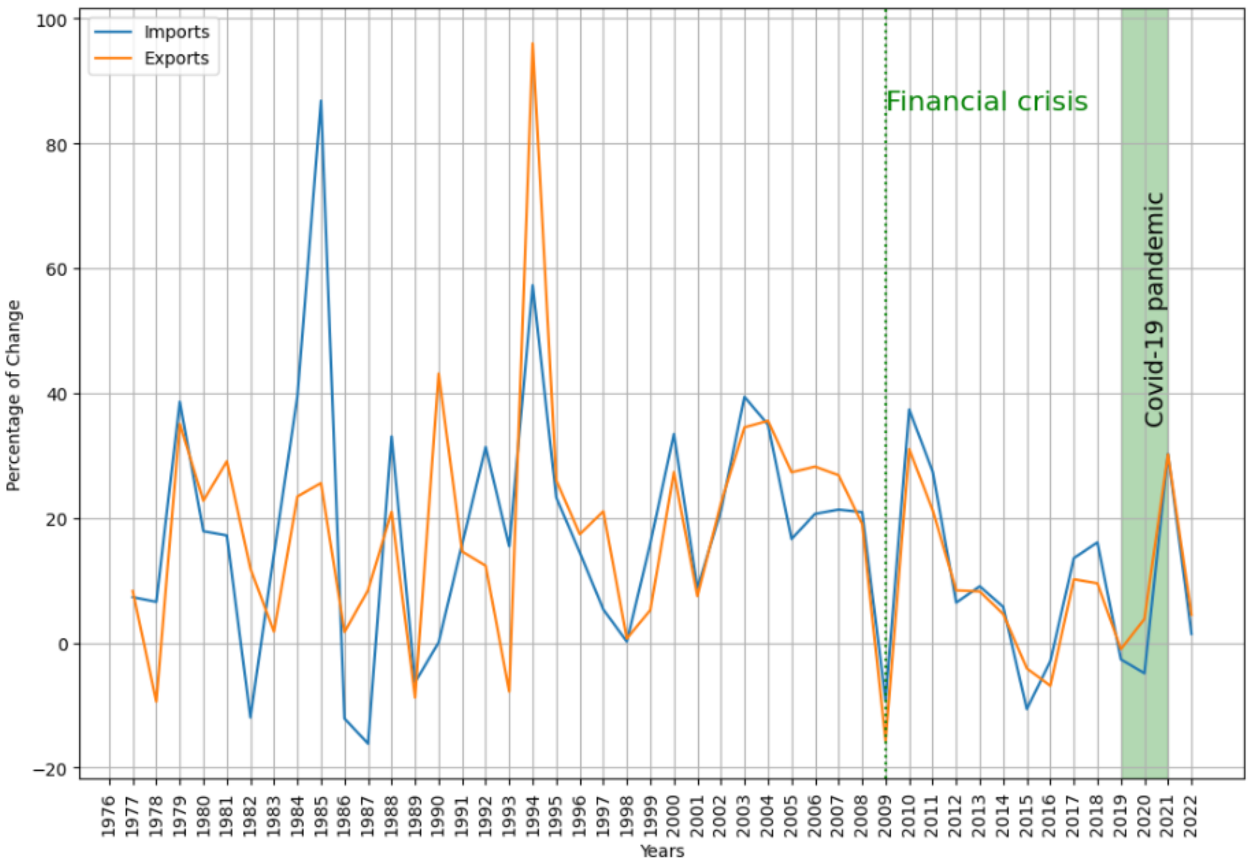

Figure 8: Change of % in China's trade values (1976-2022)

Investigating Figure 8, China's merchandise imports and exports grew at average annual rates of 19.13% and 15.93%, respectively, between 1976 and 2022. After the price liberalization in 1978, the imports and exports increased by 473.30% and 271.21%, respectively between 1976 and 1985. Between 1985 and 1987 the imports decreased by 26.34%, while exports increased by 10.34%. Between 1993 and 1994 there was a big surcharge in exports by 95.98%, as Figure 8 shows. As a result of the global financial crisis, China's imports and exports decreased by 9.27% and 15.70%, respectively between 2008 and 2009. In 2009 and 2010, China's commerce rebounded, as can be seen in Figure 8, with average growth in exports of 37.41% and average growth in imports of 31.05%. China became the world's largest merchandise exporter and second-largest merchandise importer (after the United States) in 2009, surpassing Germany in both categories. But ever since, China's trade growth has abruptly stalled. China's imports and exports increased at average annual rates of 8.59% and 7.04%, respectively, between 2011 and 2022. Exports and imports decreased by 2.93% and 6.86%, on average, between 2015 and 2016, which can be attributed to both a slowing global economy and falling commodity prices (such as ores and oil), as Morison (2019) states. On the other hand, China's imports and exports increased by 13.57% and 10.19%, respectively, in 2017. During the start of the Covid-19 pandemic (2018-2019), China’s imports decreased by average 3.75%, while exports increased by average 4.27%.

2.5.3 Top 10 exported Chinese products in 2022

Table 2 provides a summary of the top 10 export products of China in 2022, along with their market values in billions of dollars (Jahn, 2023). There are a variety of products created within China, including technology, energy, automotive and entertainment. With the highest export value of $238.08 billion, the top product is phone devices including smartphones. Ranking second at $187.9 billion are computers and optical readers. Integrated circuits/micro-assemblies (such as microchips) are third with $154.52 billion and Solar power diodes/ semi-conductors are forth with $65.88 billion. In conclusion, China's diverse export portfolio spans various sectors with phone devices, computers, and integrated circuits emerging as the top-ranking products in terms of export value.

Table 2: Top 10 exported Chinese products (2022)

Phone devices including smartphones	$238.08 billion
Computers, optical reader	$187.9 billion
Integrated circuits/micro-assemblies	$154.52 billion
Solar power diodes/ semi-conductors	$65.88 billion
Electric storage batteries	$57.22 billion
Automobile parts/accessories	$49.73 billion
Models, puzzles, other toys	$48.35 billion
Processed petroleum oils	$48.30 bilion
Electrical convertes/power unites	$47.47 billion
Lamps, lightning, illuminated signs	$46.06 billion

2.5.4 China's Major Exports Companies

The Forbes Global 2000 assesses the world's largest companies by evaluating four key metrics: sales, profits, assets and market value (Tucker, 2023). Notably, approximately 150 Chinese corporations have secured a position in the Forbes Global 2000, as reported by Jahn (2023). Table 3 details the top 11 Chinese companies within this ranking, primarily involved in machinery/electronics or gas-related products, with beverages occupying the 11th position. This Table offers valuable insights into the Chinese export sector, highlighting the notable success of these corporations across various industries.

Table 3: China Major Exporting Companies in 2022

1. Aluminum Corporation of China (aluminum)
2. BYD (cars, trucks)
3. Dongfang Electric (electrical equipment)
4. Dongfeng Motor Group (cars, trucks)
5. Gree Electric Appliances (household appliances)
6. Midea Group Co. Ltd. (household appliances)
7. PetroChina (oil, gas)
8. SAIC Motor (cars, trucks)
9. Sinopec-China Petroleum (oil, gas)
10. Sinopharm Group (pharmaceuticals)
11. Tsingtao Brewery (beverages)

It is interesting to note that the majority of these companies are related to manufacturing, a sector that China was not a dominant force in, as seen in sections 2.1.3 and 2.2.2. This shows China's determination and perseverance in transforming its economic sectors to align with current trends.

2.5.5 China’s Major Trading Patners

Having mentioned China's 2022 export Figure of $3.1 trillion, it would be beneficial to examine the top nations that served as importers for these goods. China's total export value accounts for 16.2% of the world's overall exports, which were valued at $22.144 trillion in 2022, as Workman mentions (2023b). Figure 9 shows the 15 top trading partners of China in terms of exports in 2022, along with the percentage of the total exports. Nearly two-thirds of all Chinese exports were sent to the above countries, shown in Figure 9. The United States claims the top position, accounting for $582.8 billion and constituting 16.2% of China's overall exports. This amount surpasses the export value to Hong Kong, the second largest importer of Chinese goods by almost double. The following five positions are taken by neighboring Asian countries, such as Hong Kong with 8.3% of China’s total exports, Japan with 4.8% and South Korea with 4.5%.

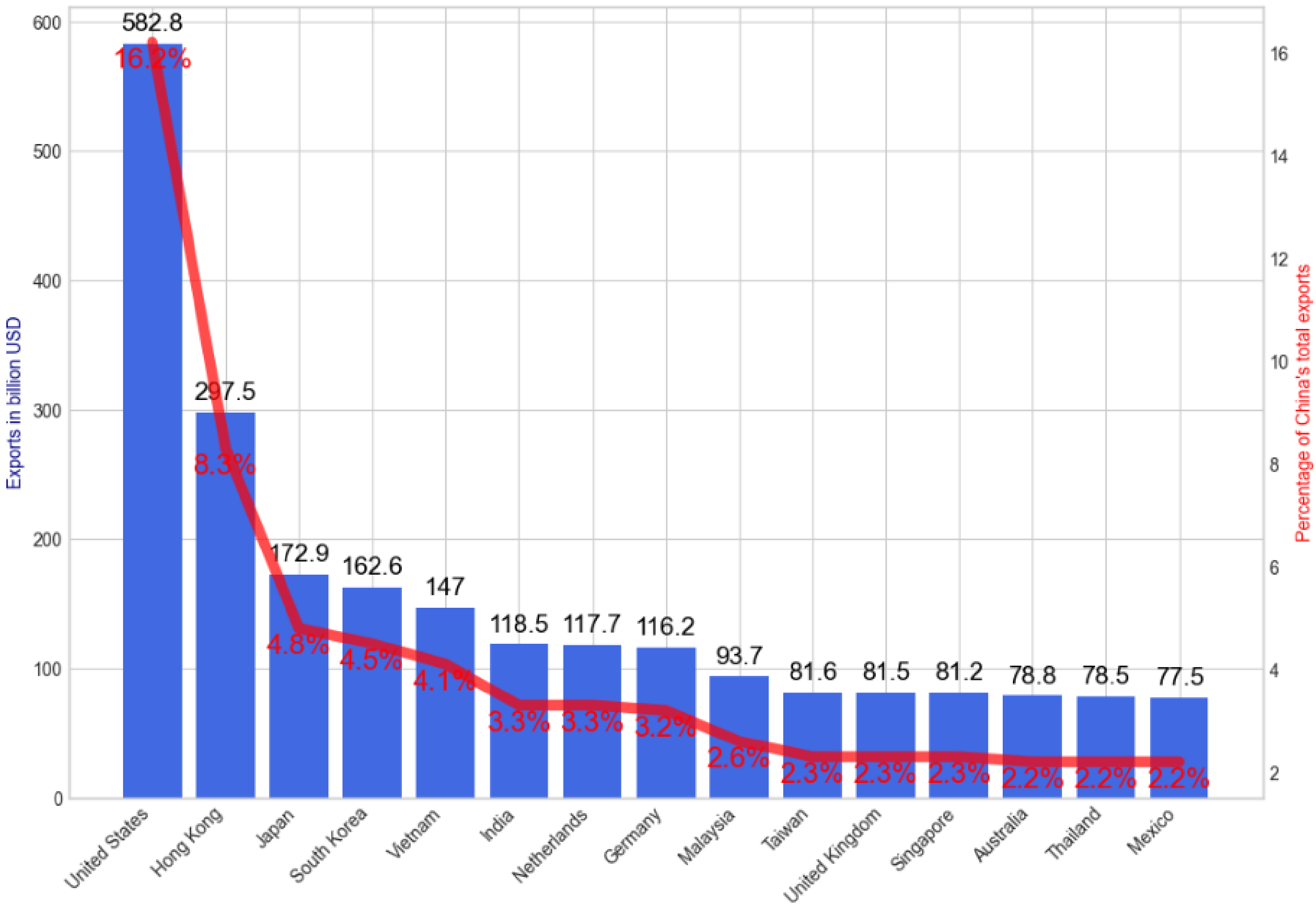

Figure 9: China's Exports by Country in 2022 (in billions of US$ and %)

Apart from the top 15 exports, when viewed through a continental perspective, almost half (47.4%) of mainland China's exports in terms of value were sent to other Asian nations, with 20.7% being sold to European importers. Additionally, 19.9% of China's exports were shipped to North American countries. Smaller portions were distributed to Latin America (4.8%, excluding Mexico but including the Caribbean), Africa (4.6%) and Oceania (2.6%, primarily Australia and New Zealand)[2].

[2] The data is calculated from the totality of China’s exports and not just the 15 countries shown in Figure 9.

2.5.6 China's Biggest Trading Surpluses

The top 10 countries contributing to China's significant trading surpluses[3] in 2022 are diverse and shown in Figure 10 (Workman, 2023b). The leading contributor is the United States, with a substantial surplus of $403.8 billion, emphasizing the robust economic ties between the two countries. Hong Kong follows with almost three fourths of the American value, contributing $289.7 billion, while the Netherlands and India have the third and fourth positions with surpluses of $105.2 billion and $101 billion, respectively. Notably, the list encompasses a mix of developed and emerging economies, such as Mexico, the United Kingdom, Vietnam, Singapore, the Philippines and Poland, all playing pivotal roles in China's trade dynamics. These surpluses are indicative of the strong export performance in various sectors, including machinery, electronics and other key industries, as explained in section 2.5.3, exhibiting China's strategic positioning in the global trade landscape.

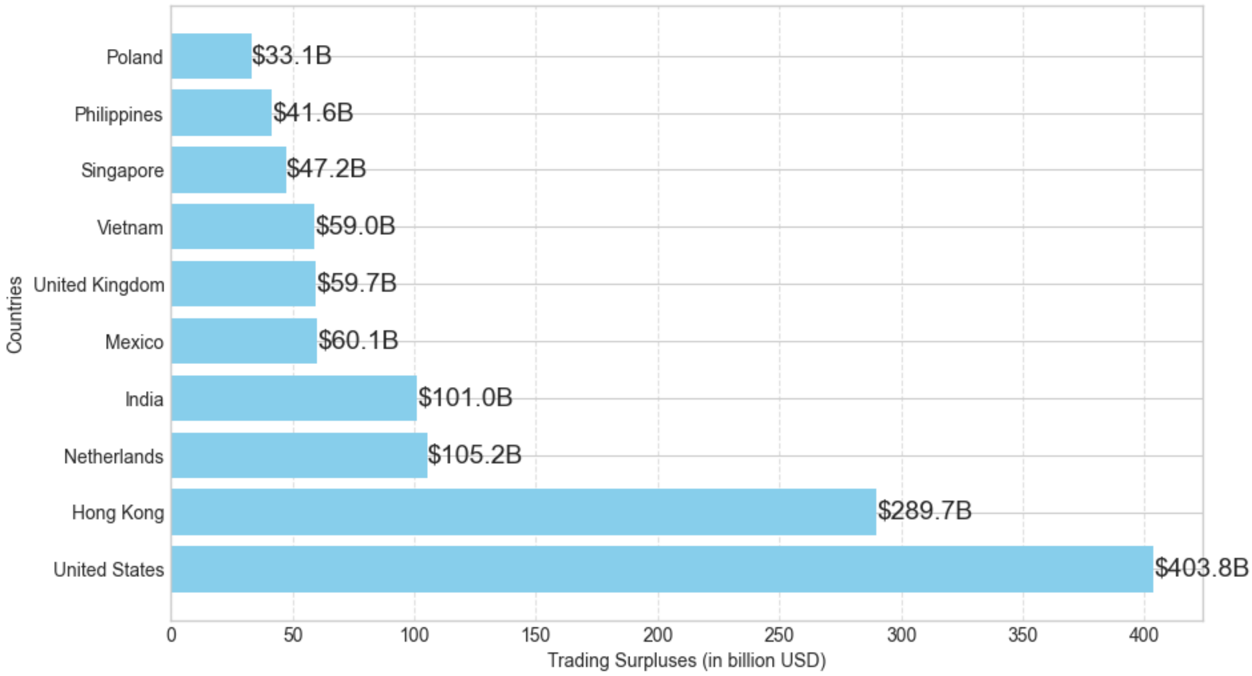

Figure 10: Countries Generating China's Biggest Trading Surpluses (2022)

2.5.7 China's Greatest Trading Deficits

Figure 11, as reported by Workman (2023a), provides insights into the ten countries that recorded the highest trade deficits[4] with China in 2022. With a deficit of -$156.5 billion, first on the list is Taiwan. Australia follows with a deficit less than one third of Taiwan's with -$63.3 billion, highlighting challenges in the trade dynamics between the two countries. Brazil, Switzerland, and Saudi Arabia follow, with deficits of -$47.6 billion, -$42.2 billion and -$40.1 billion, respectively. The list further includes Russia, South Korea, Oman, Iraq and Chile, all contributing to China's trading deficits. Nevertheless, these deficits are rationalized by the elevated import levels depicted in Figure 7, indicating that while China's exports remain the predominant metric, the disparity between exports and imports amounted to just $576.65 billion in 2022.

---

[3] A trade surplus is defined by the Britannica Dictionary (2023b) as: "a situation in which a country sells more to other countries than it buys from other countries: the amount of money by which a country's exports are greater than its imports".

[4] A trade deficit is defined by the Britannica Dictionary (2023a) as: "a situation in which a country buys more from other countries than it sells to other countries".

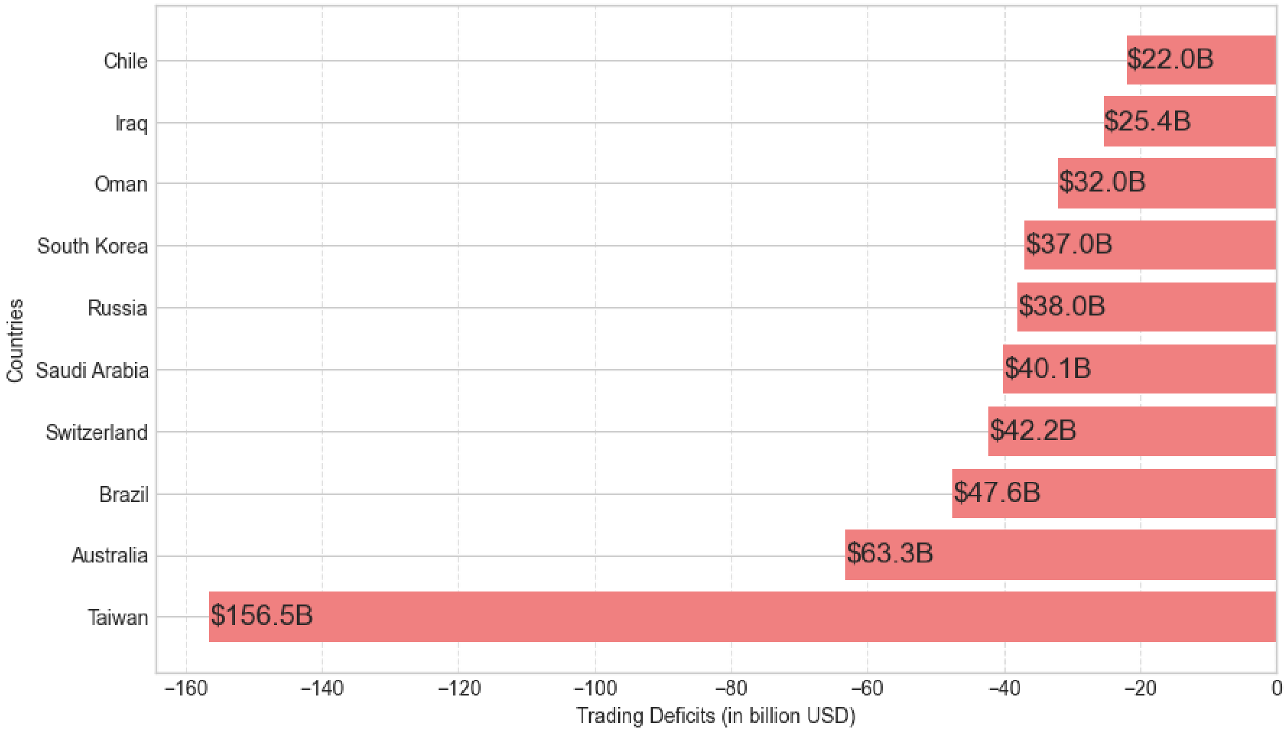

Figure 11: Countries Causing China's Greatest Trading Deficits (2022)

2.6 Analysis of China's Foreign Direct Investment

The last key metric that will be analysed in this paper is China's Foreign Direct Investment (FDI) inflows. This timeline provides information about key events that have influenced China's FDI inflows over the years and are analysed through this section.

Foreign Direct investment

Date	*Key event*
1978	Chinese economic reforms
1979	Law on Joint Ventures
1992	nationwide implementation of open policies for FDI
1997-1998	East Asian Financial Crisis
2001	China's entry into the World Trade Organization (WTO)
2006	Regulations for Merger and Acquisition (M&A) of Domestic Enterprises by Foreign Investors
2007	Anti-Monopoly Law
2007	Enterprise Income Tax Law
2007-2008	Financial crisis
2022	Ukraine-Russia war

China's open economy could not have succeded without the help of Foreign Direct Investments (Broadman & Sun, 1997). Mainly Asian developing countries have continuously been hosts of FDI throughouth the 20th century, with China and Vietnam receiving the largest amount of FDI. Figure 12 presents the flow of FDI into China from 1979 to 2022 (Macrotrends, 2023).

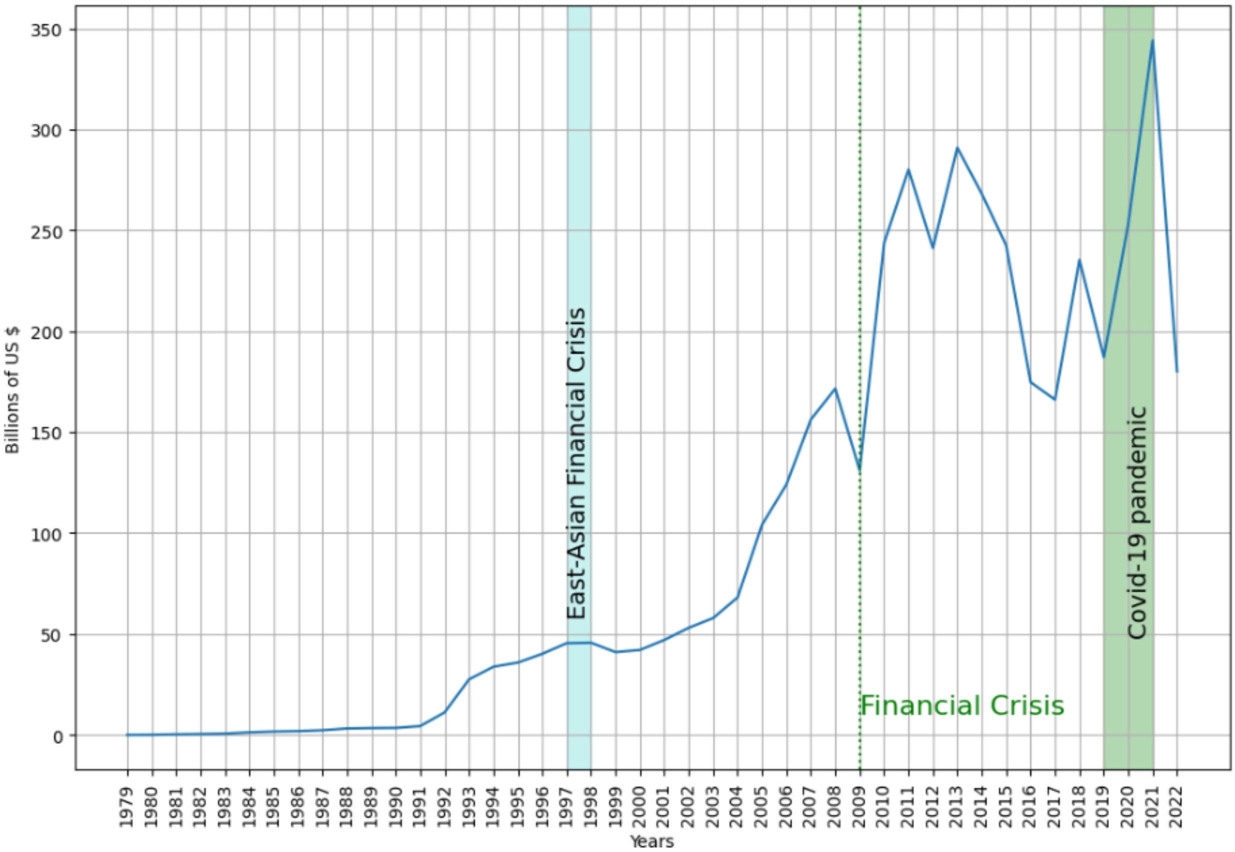

Figure 12: Foreign Direct Investment in China (1979-2022)

The 1979 Law on Joint Ventures was the beginning of FDI in China, giving investment incentives to other countries, Broadman and Sun (1997) recount. By 1992, China had received approximately one fourth of FDI flows towards developing countries and by 1995 that percent was at 40. They were and still are the largest source of external capital to China.

During the 20th centuty, there was a noticeable imbalance in the geographic spread of foreign capital across the country, with a distinct concentration of FDI in the eastern and southeastern regions (Chen, 2018). During the period from 1985 to 1992, the twelve coastal provinces, including Guangdong, Fujian, Beijing and Shanghai, gathered between 87% and 93% of the overall FDI inflows into China. Since 1989, their portion of FDI has consistently stayed above 90%. In 1980, China founded four Special Economic Zones in the provinces of Guangdong and Fujian. These zones include Shenzhen, Zhuhai, Xiamen and Shantou. During this initial phase of China's globalization efforts, the Chinese government was cautious in allowing FDI into its domestic economy. Consequently, foreign investors approached investing in China with a level of caution as can be seen in Figure 12, where FDI inflows are kept at a low lever for the first twelve years after the reforms. From 1979 to 1991 the annual average inflow in China was approximately $1.63 billion.

The second phase of FDI in China began in 1992, following Deng Xiaoping's influential tour of the southern coastal areas and Special Economic Zones (Chen, 2018). This phase marked a departure from regional to nationwide implementation of open policies for FDI, with new regulations and policies encouraging FDI inflows. The preferential policies initially applied to 14 coastal cities were extended to 52 cities and selective service industries were opened for limited FDI participation, including aviation, telecommunications, banking and retail trade. To spur economic growth and address regional disparities, the Western Development Strategy was launched in 1998, covering 12 provinces and regions, Chen continues. In the 1990s, the Chinese government further liberalized its FDI regime, enacting various laws and regulations.

This period saw the establishment of a more consistent and systematic FDI regulatory framework, leading to remarkable growth in FDI inflows, reaching $11.15 billion in 1992 and doubling to $27.51 billion in 1993. However, FDI faced challenges in 1999 and 2000, experiencing a decline, mainly attributed to the impact of the East Asian Financial Crisis (1997-1998) (International Monetary Fund, 1998). This crisis substantially weakened the outward investment abilities of East and South-East Asian economies, which had previously been significant investors in China. This decline can also be seen in Figure 12 where a curve is formed in 1999.

The third phase of China's Foreign Direct Investment evolution started in 2002 after its entry into the World Trade Organization (WTO) the previous year (Chen, 2018). In anticipation, the Chinese Government amended key laws in 2000 and 2001 and also after 2002 in order to meet its WTO commitments. This led China to experience a resurgence in FDI inflows, which kept until 2008. FDI inflows surged from $47.05 billion in 2001 to $171.53 billion in 2008. In 2005 China issued a new Company Law that simplified company establishment requirements and expanded shareholder rights. In 2006 the Regulations for Merger and Acquisition (M&A) of Domestic Enterprises by Foreign Investors established new rules for foreign investors acquiring interests in China's domestic companies. In 2007 China enacted its first Anti-Monopoly Law, treating foreign and domestic businesses equally. The same year they passed the Enterprise Income Tax Law, which unified the tax rates for both foreign and domestic enterprises at 25%.

The influence of the Global Financial Crisis on FDI inflows in China was impactful. It led to a decrease in inflows to $131.05 billion in 2009. Subsequently, there was a recovery, with inflows reaching $280.07 billion in 2011. The following 10-year period can be characterised by some fluctuations. During the covid-19 pandemic, FDI rose with $187.16 billion in 2019, $253.09 billion in 2020 and $344.07 billion in 2021. This resurgence was driven by robust M&A markets and substantial growth in international project finance, facilitated by loose financing conditions and major infrastructure stimulus packages. However, the global economic landscape underwent a dramatic shift in 2022. The war in Ukraine, combined with lingering pandemic effects, triggered a triple crisis involving food, fuel, and finance in several countries (Santander, 2023). In 2022, Figure 12 shows a notable decline with $180.16 billion, almost half than the previous year.

In summary, FDI inflows suggest a dynamic economic evolution. From cautious beginnings in 1979 to substantial reforms spurred by WTO entry in the early 2000s, each phase reflects China's adaptability. Despite challenges like the East Asian Financial Crisis and the Global Financial Crisis, strategic adjustments have consistently attracted significant FDI, showcasing China's appeal for international investments.

2.6.1 FDI Sources

In 2022 as Huld (2023) mentions, Foreign Direct Investment inflows into China continue to experience consistent growth, with notable increases from key source countries. Remarkably, South Korea, Germany, and the UK witnessed substantial year-on-year growth in their investments in China, registering increases of 64.2%, 52.9% and 40.7%, respectively. Furthermore, there was a remarkable surge in investments from the European Union (EU), contrary to the previous year. Investments from the EU recorded a substantial year-on-year growth of 92.2%, indicating a noteworthy shift from the 10.4% decrease observed in 2021. Simultaneously, investments from countries associated with the Belt and Road Initiative, spanning sixteen in western Asia, nine in South Asia and five in central Asia, grew by 17.2%.

Additionally, investments from the ASEAN[5] countries experienced an 8.2% increase, contributing to the overall positive trajectory of FDI inflows into China.

According to the Ministry of Commerce of the Republic's People of China (2023), the significant role of large-scale projects in driving foreign investment has been strengthened. The realized foreign investments in projects with contracted foreign investments exceeding $100 million in 2022, reached 653.47 billion yuan (US$90,116 billion), marking a 15.3% increase than the previous year. This accounted for 53% of the total actual foreign investment in the country, playing a crucial role in stabilizing and supporting foreign investments.

2.6.2 Distribution of Inward FDI Stock in China in 2021 by Country/Region of Origin

The distribution of inward Foreign Direct Investment stock in China in 2021, illustrated in Figure 13, reveals a concentration of investments from various countries and regions (National Bureau of Statistics, 2023). Hong Kong stands out as the dominant contributor, accounting for 54.70% of the total FDI stock. The Virgin Islands, Japan and Singapore follow with shares of 6.90%, 4.70% and 4.60%, respectively. Evidently, the United States, South Korea and Taiwan also contribute, each representing around 3-4% of the FDI stock. The data in Figure 13 underscores a diverse array of countries contributing to China's FDI, showing a diverse international investment profile in China, with key players from both neighboring and distant regions participating in its economic landscape (Macrotrends, 2023).

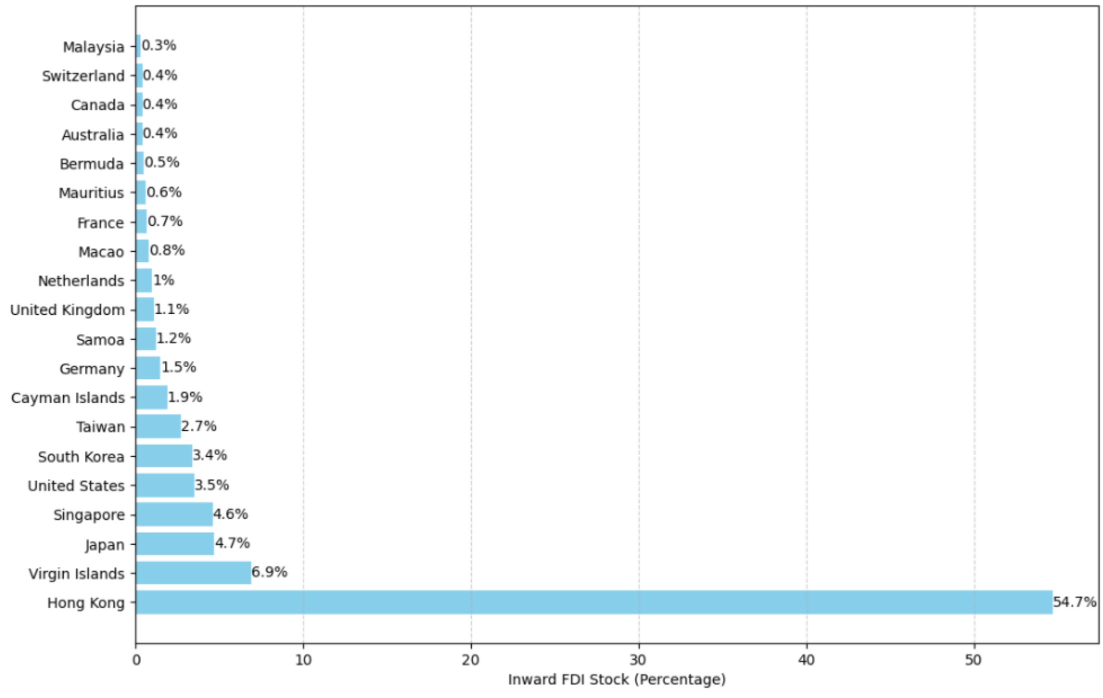

Figure 13: Distribution of Inward FDI Stock in China in 2021 by Country/Region of Origin

Foreign Direct Investment has notable positive spill-over effects[6] on the productivity of Chinese domestic firms operating within the same industry in the manufacturing sector (Chen, 2011). However, Chen's research indicates that FDI does not generate significant spill-over effects on the productivity of domestic firms through backward and forward industrial connections. This

[5] The Association of Southeast Asian Nations (ASEAN) consists of ten member countries: Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand and Vietnam (ASEAN Secretariat, 2020).
[6] According to Kenton (2019): "Spillover effect refers to the impact that seemingly unrelated events in one nation can have on the economies of other nations."

entails raw material suppliers and product manufacturers. The first contribute raw or processed materials (blocks) for the creation of products and the second transforms these blocks into products. This phenomenon is primarily attributed to the extensive involvement of FDI firms in processing trade, which disrupts the industry linkages between FDI and domestic firms. These findings suggest that host countries can enhance the positive spill-over effects of FDI and consequently boost domestic firm productivity by reinforcing industrial connections between domestic and FDI firms.

FDI has also played a positive role in Chinese exports (Chen, 2018). They advance China's global trade in its substantial involvement in processing trade, particularly in the coastal areas. In the span of 2002 to 2012, according to data from the Chinese Ministry of Commerce in 2015 (Chen, 2018), processing trade by FDI firms constituted over 80% of China's overall processing trade and more than 60% of the total trade was conducted by FDI firms. Specifically, the processing exports by FDI firms represented more than 70% of their total exports.

2.7 Comparison of Economic metrics

Concluding Section 2, a comparison including all key economic indicators covered in this report over time will be conducted. Given that all these indicators seem to be of different numerical scales, in order to compare them, they need to be transformed into a common index. To achieve this standardization, each indicator, namely GDP, Inflation, Imports, Exports and FDI underwent scaling using the MinMaxScaler() function from the sklearn library on a scale from 0 to 1. The results were then plotted together, as illustrated in Figure 14.

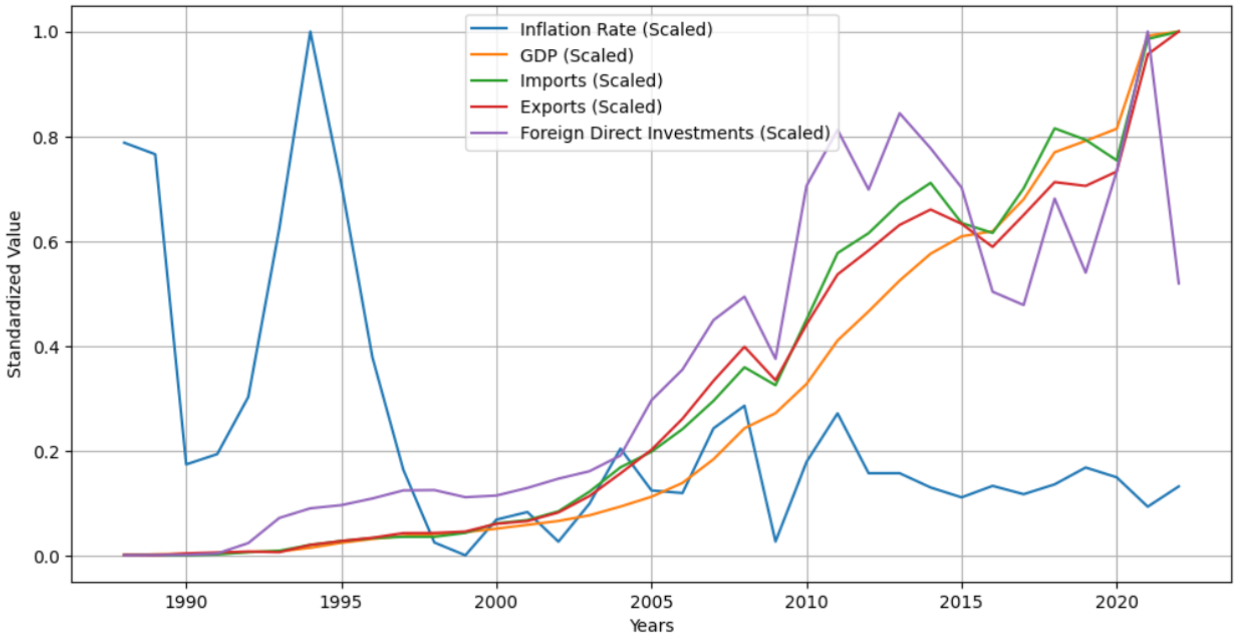

Figure 14: Comparison of Economic Metrics

The comprehensive comparison of economic indicators of Figure 14 reveals a consistent upward trajectory in GDP, Imports, Exports, and FDI over the analysed period. In particular, all these metrics, except for inflation, have followed a similar trend since the early 21st century, indicating a steady evolution in China's fiscal policy. The period before 2000 saw almost all metrics around nearly-zero values, reflecting the early stages of adjusting to the new economic reforms.

Conversely, the Inflation Rate demonstrates an opposing pattern. Over the years, there is a noticeable decline in inflation, signifying a trend toward lower and more stable inflation levels, an important trend for maintaining purchasing power and fostering a conducive economic environment.

Furthermore, GDP, imports and exports seem to be following the same trend, reaching their highest recorded values by the year 2022. In 2007, in the beginning of the financial crisis, all metrics experienced a downturn but subsequently recovered in the following year. A similar pattern is observed around the years 2016 and 2020 for all indicators. Regarding the decline in 2016, as suggested by (Hong et al., 2016), there was a deceleration in global trade growth initiated in 2015, with China playing a significant role due to weakened investments and the ongoing process of rebalancing its economic structure. This rebalancing affected 60 other nations, as China serves as a key transmitter of economic trends. As mentioned in section 2.5.1, China imports intermediate goods and exports finished products, implying that a decrease in demand from its export destinations could lead to a corresponding decline in its exports. This, in turn has cascading effects on numerous other nations that rely on China for goods. Regarding the fall in 2020, the impact of the COVID-19 pandemic on the Chinese economy, as detailed in Section 2.5.2, was substantial.

3 Forecasting

The first part of this research report dealt with analysing key metrics of the Chinese Economy. The second part will cover the annual forecasting of economic values. The aim of this part is to empirically construct statistical models for predicting GDP, inflation, imports and exports values and FDI inflows of China, utilizing the Univariate Autoregressive Integrated Moving Average (ARIMA) model proposed by Box and Jenkins in 1976 (Eissa, 2020).

3.1 Literature Review

The use of ARIMA models on time series data has been proven a valuable experimental tool for both developed and emerging nations, whether the data is presented annually or quarterly (Eissa, 2020). There has been a broad range of experiments forecasting economic values and especially regarding GDP and inflation. Researchers had been searching for the best model that captures the data's evolution in order to accurately predict future values. Noteworthy examples include the period from 1959 to 2011 for India, where GDP growth rates were predicted, with the results suggesting that an ARIMA (1, 2, 2) model was the most suitable fit (Maity & Chatterjee, 2012) and economist Dritsaki, who leveraged data from 1980 to 2013, in order to forecast Greece's real GDP rate in 2015 using an ARIMA (1, 1, 1) model. The statistical outcomes indicated a consistently improving forecasted Greek GDP rate (Dritsaki, 2015). Moreover, Qu et al. (2021) conducted a study on the GDP data of Hubei Province, China spanning from 1978 to 2019, employing the ARIMA model to predict GDP Figures for the years 2020 and 2021. His findings show that the ARIMA model is adept at capturing the developmental trend and accurate short-term forecasts for GDP. Lastly, Sharma et al. (2020) forecasted India's FDI values with an accuracy of 96.4%, as evaluated the next years. Their ARIMA (1, 1, 4) model showed a rapid growth in FDI for 2023 and 2024. In conclusion, it is noteworthy to observe that all the above research projects have used ARIMA models with relatively low numbers of parameters but not the same ones. This observation suggests that even when dealing with similar types of data, a one-size-fits-all model may not be the optimal choice.

Seeing the success of the previous research, this report will attempt to forecast China's metrics for a 10-year time period. Seeing how the majority of the forecasts are 2-years we will consider the 2-year forecast to be more trustworthy and the next 8-years will be considered from a upwards/downwards trend classification side.

3.2 Methodology

The Autoregressive Integrated Moving Average Model, abbreviated as ARIMA, stands as a conventional statistical model employed for the forecasting and analysis of time series data. It is a regression type equation including lags of the dependent variables and/or lags of the forecast errors as independent variables (Shweta, 2021). In addition to outlining the model itself, its creators, Box and Jenkins, proposed a systematic approach for recognizing, estimating and validating models tailored to a particular time series dataset, called the Box-Jenkins Method (Brownlee, 2017). In general, an ARIMA model is a simpler AutoRegressive Moving Average combined with integration. The important elements are described below.

AR: stands for Autoregression, meaning that the model uses the dependent relationship between a current observation and a certain number of lagged observations,

I: stands for Integrated, which means subtracting an observation from its value at the previous time step (differenced raw observations) in order to achieve stationarity in the time series data and

MA: stands for Moving Average which is a model of dependency between a current observation and the residual errors derived from a moving average model applied to lagged observations.

These three abbreviations are integrated into the model in the form of parameters. There are three parameters (p, d, q), which indicate the specific ARIMA model. They can be defined as follows:

p: This is the lag order (the number of lag observations included in the model).
In order to identify the value of p, the Autocorrelation Function (ACF) plot is used. This diagnostic plot summarises the correlation of one observation with its lag values.

d: This is the degree of differencing (the number of times that the raw data are differenced in the model). The number of times that the raw data are differenced is the number of the parameter (if the data are stationary then d=0).

q: This is the order of moving average (the size of the moving average window). In order to identify the value of q, the Partial Autocorrelation Function (PACF) is used. This diagnostic plot presents the correlation of one observation with its lag values that is not accounted for by prior observations.

Both plots (ACF and PACF) are represented as bar charts, illustrating the 95% and 99% confidence intervals through horizontal lines. Bars intersecting these confidence intervals are more significant and worth noting. In this project, these plots will be used in order to initially construct an ARIMA model manually. Subsequently, the manually developed model will be compared with the automatically obtained best model using the statsmodel library with the use of Python. The code for the development of all models can be found in Appendix C. This comparative analysis aims to assess the consistency and effectiveness of both approaches in modeling the data.

The ARIMA models will be used in order to forecast future values for all metrics. The experiment will be conducted in Python in Jupyter Notebook.

The section of the report is structured as follows for each dataset:

1. **Exploratory Data Analysis (EDA):** The initial phase involves conducting an exploratory analysis of the dataset.

2. **ARIMA Model Parameterization**:
   a. Examination of Stationarity: The dataset will be assessed for stationarity using the Dickey-Fuller test. The Dickey-Fuller (DF) test, introduced by Dickey and Fuller in 1979, examines the null hypothesis (H_0) of the presence of a unit root in an autoregressive (AR) model. When a time series is characterised by stationarity, it means that the data are not dependent on time and can be modelled easier (Brownlee, 2017). Non-stationary time series can show seasonality and trends that can influence the mean, variance or other summary statistics of the model not making it accurate. The H_0 hypothesis suggests that the investigated data are not stationary (Jalil & Rao, 2019). The alternative hypothesis (H_1) indicates failure to reject the null hypothesis, meaning that there is stationarity in the data.

b. p Parameter Exploration: The p parameter will be decided through the Autocorrelation Function (ACF) plot.

c. q Parameter Examination: The q parameter will be investigated using the Partial Autocorrelation Function (PACF) plot.

3. **Manual ARIMA Model Development**: The insights gained from the above steps will be employed to manually construct an ARIMA model. The datasests will be split into 0.80 training sets in order to better train the model and evaluate it. The 0.88 split will give the models enough data to train in order to accurately forecast new values based on the majority of the dataset. A prediction will be plotted along with the test and train sets.

4. **Auto_model Function Implementation**: Subsequently, the auto_model function from the statsmodel library will be utilized to automatically identify the most suitable ARIMA model for the dataset.

5. **Model Comparison**: The generated ARIMA models will be applied to the datasets. The outcomes of all five models will be compared to asses their performance and consistency. The Mean absolute percentage error (MAPE) will be used to evaluate their peformance. The measure calculates a model's accuracy of forecast system by measuring the absolute percent of error for each dataset.

3.3 Gross Domestic Product Forecasting

In this section of the report, Chinese GDO values will be forecasted through the application of an Autoregressive Integrated Moving Average (ARIMA) model.

3.3.1 Dataset Description

The dataset that will be used for forecasting GDP values is sourced from Macrotrends (2023) and contains annual time series data of China between 1960 and 2022. It contains 63 observations and the dependent variable is GDP Per Capita (US $).

3.3.2 Exploratory Data Analysis (EDA)

In order to better understand the data, the time series will be analysed further.

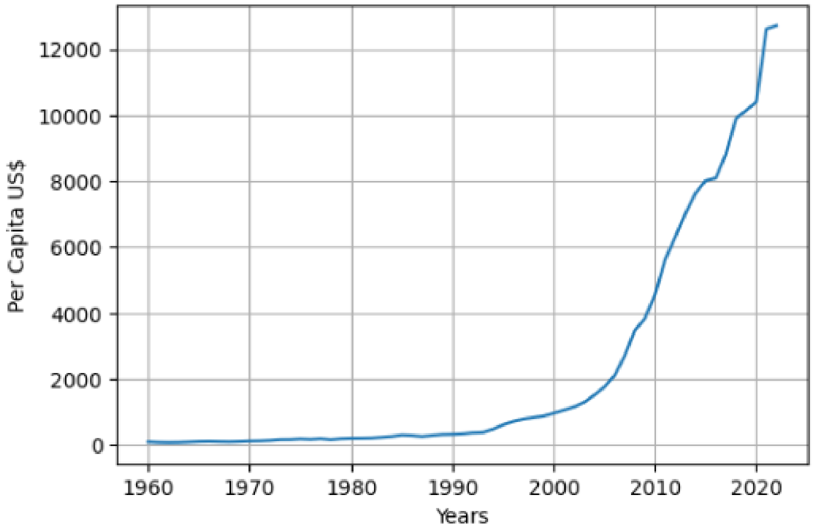

Figure 15: GDP Per Capita of China Line plot (1960 -2022)

Figure 15 shows the line plot for China's GDP Per Capita values. The line ascends over the years, indicating a positive and continuous growth trajectory. After 2000 the growth seems to be more rapid.

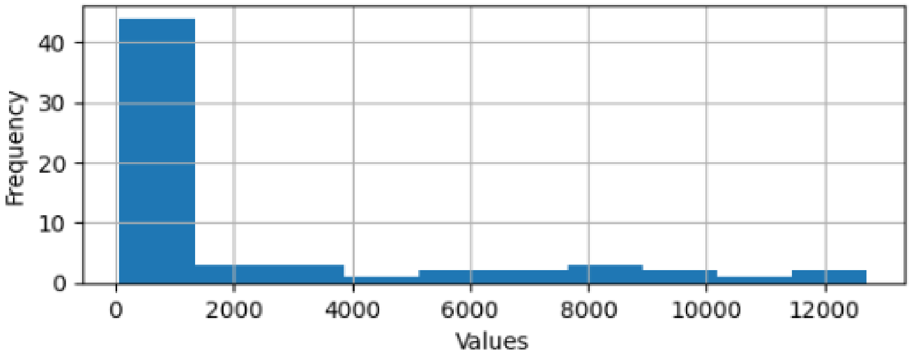

Figure 16: GDP of China Histogram of frequency distribution of values

The histogram in Figure 16 displays the frequency distribution of values in the dataset. The x-axis represents the values, and the y-axis represents their frequency. Notably, the majority of values cluster between 0 and 1000. However, the presence of values reaching up to 12,000 indicates higher occurrences of specific values outside the main concentration.

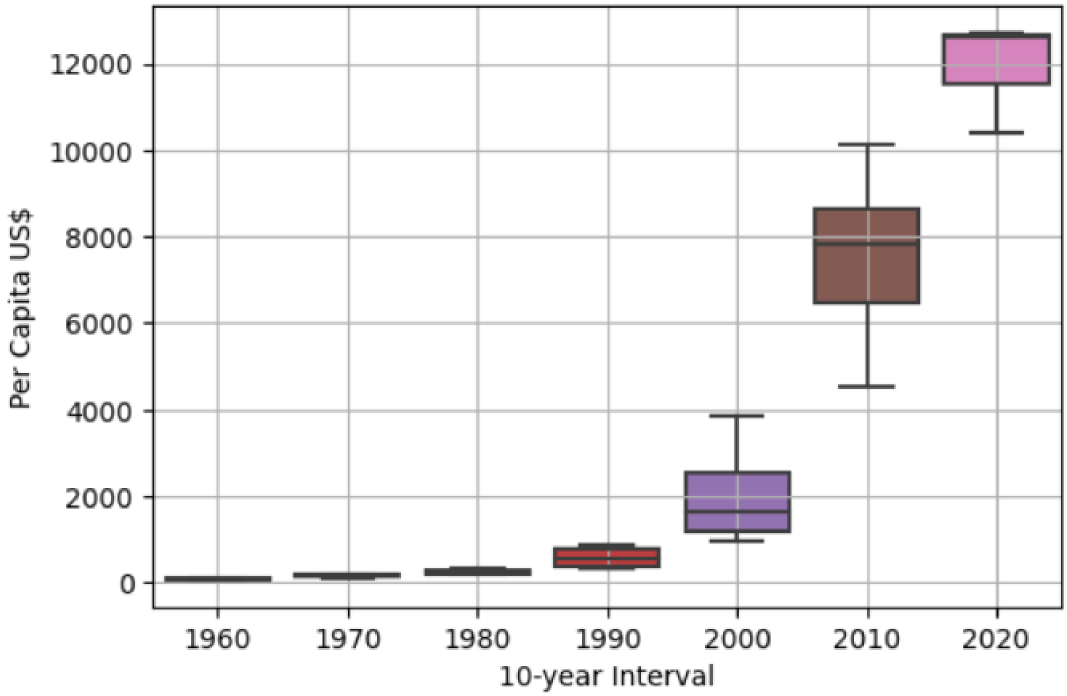

Figure 17: Box and Whiskers plot for Chinese GDP (1960-2022)

The box-and-whisker plot, depicted in Figure 17 illustrates the annual GDP per capita of China. Evidently, a visible trend is observed in both the median and the upper quartile of the boxes, which consistently ascend across the examined years. During the initial four 10-year intervals, the interquartile range (IQR), represented by the span between the first quartile (Q1) and the third quartile (Q3), as well as the length of the whiskers, remains relatively modest. In the subsequent three 10-year intervals, there is a noticeable expansion of the IQR, indicating an increased dispersion of the data. The most pronounced interval is observed in the sixth box, where the data range extends from nearly 4,200 to almost 10,000 US dollars. This widening of the IQR reflects a considerable variance in annual GDP per capita from 1960 to 2022.

3.3.3 Manual model construction

In order to understand the key compartments of an ARIMA model better (autoregressive (AR) terms, differencing (I) and moving average (MA)), as a first step, the model will be constructed manually. This means that the values for its parameters will be selected without help from any functions. As mentioned in section 3.2, the ARIMA model requires three parameters (p, d, q).

3.3.3.1 Stationarity

The d parameter of the manually constructed ARIMA model will be provided from the Dickey Fuller test. The adfuller() function was imported from the stats.models library in Python as shown in code snippet 1. X is the dependent variable (GDP Per Capita) used for this analysis.

Input:

```
from statsmodels.tsa.stattools import adfuller
X = series[' Per Capita (US $)']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Output:

```
ADF Statistic: -0.059821
p-value: 0.953296
Critical Values:
	1%: -3.563
	5%: -2.919
	10%: -2.597
```

Code snippet 1

In order to accept or fail to accept the null hypothesis, the p-value is examined. If the p-value is less than 0.05 then the H_0 is accepted, otherwise we fail to accept the H_0. In this case, as 0.953296 > 0.05 we fail to accept the H_0, meaning that at least one level of differencing is required. The data need to be differenced for the model to give an accurate prediction. Moreover, the critical values need to be smaller than the ADF Statistic in order to accept the H_0 and as can be seen in code snippet 1, all three are greater than -0.059821. All in all, d equals at least 1.

3.3.3.2 ACF/PACF

The next step in manually constructing an ARIMA model is to select the lag values for the Autoregression (AR) and Moving Average (MA) parameters. By reviewing the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) the p and q parameters, respectively, will be selected. With the use of the statsmodel library the plots in Figure 18 were created. The x axis shows the lag number and the y axis shows the correlation coefficient value between -1 and 1. In these plots the number of lag values have been limited to 30 for readability.

In order to interpret an AFC or PACF plots, we look at the confidence intervals (blue area on the plot). If an autocorrelation value goes beyond the confidence interval region, then it can be assumed that it is statistically significant. The number of the first statistically significant values in each plot give the p and q values, respectively.

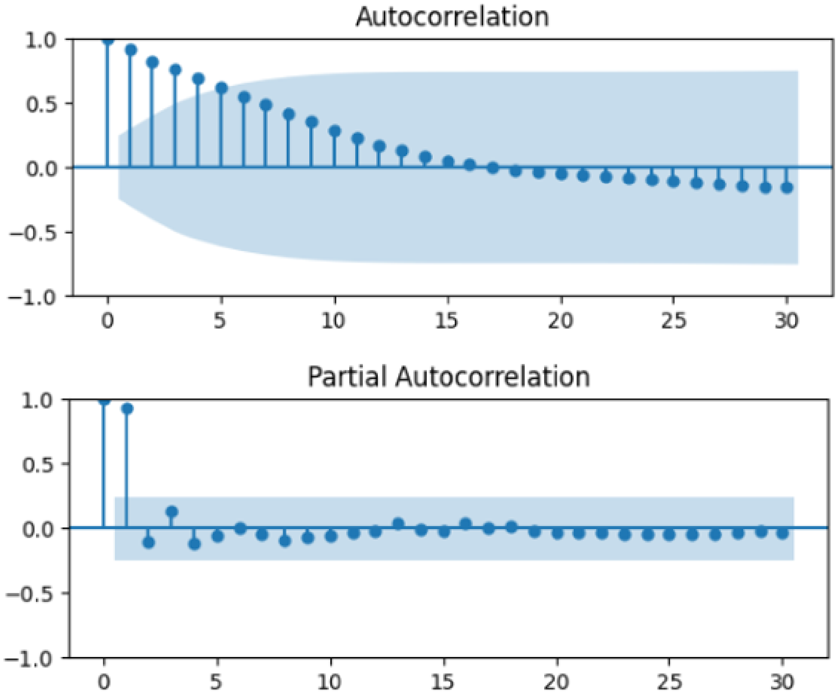

Figure 18: GDP ACF and PACF plot

The ACF plot in Figure 18 shows a significant lag for 4-5 years, while the PACF shows a significant lag for 2 years. The final model could be (4, 1, 2). Even though the 5th year shows a significant relationship as well, it is recommended to always choose the simplest model (Brownlee, 2017).

3.3.3.3 ARIMA model (4, 1, 2)

The first step into building the (4, 1, 2) model is to split the dataset into a 0.88 training set. From 63 observations, the first 55, from 1960 until 2014, will be used for training and the rest 8, from 2015 to 2022 for testing. As code snippet 2 shows, the ARIMA function with an order of (4, 1, 2) is used from the statsmodel library in Python. The data is fitted and then the get_forecast() function gives 8 predictions (length of the testing set). Afterwards all 3 datasets (train set, test set, ARIMA predictions) are plotted together in Figure 19.

Input:

```
train_size = int(len(X) * 0.88)
train, test = X[0:train_size], X[train_size:len(X)]
order = (4, 1, 2)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel.predict(start=y_pred_df.index[0], end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
```

Output:

```
RMSE: 595.2453704713009
```

Code snippet 2

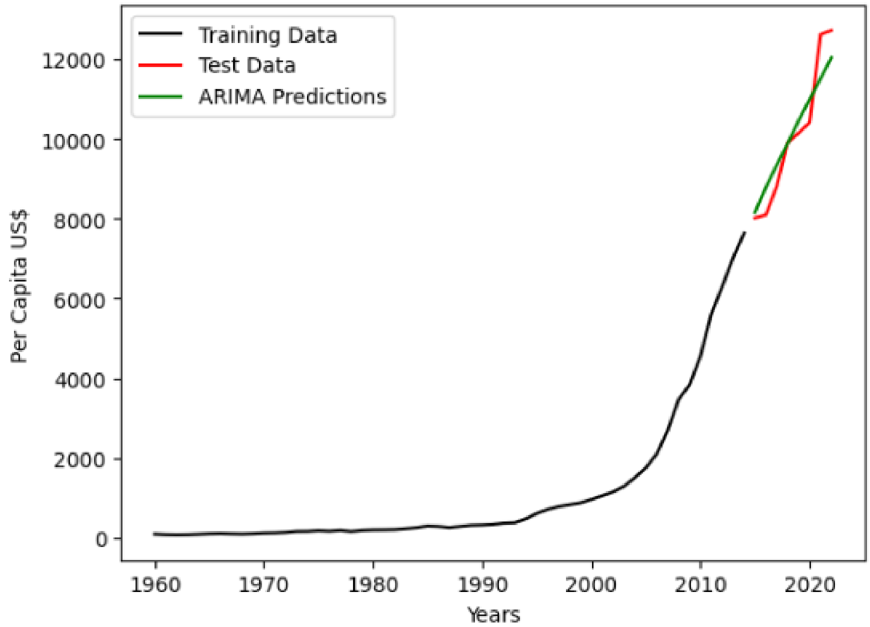

Figure 19: GDP prediction on train data with ARIMA model (4, 1, 2)

The manually constructed ARIMA (4, 1, 2) model seems to be close to the real values, according to Figure 19. The Root Mean Squared Error (RMSE) was employed as a quantitative measure to assess the accuracy of the ARIMA model predictions in comparison to the test set. The RMSE value was found to be 595.24 US$, indicating a reasonable level of prediction error. This suggests that the ARIMA (4, 1, 2) model captures the underlying patterns in the data and provides satisfactory predictions for GDP Per Capita. However, before any prediction gets made, the auto_model function will be utilized to indicate the optimal model for the most accurate predictions.

3.3.4 Automatic best model for 0.88 split

In order to select the best ARIMA model, the auto_arima() function is deployed from pmdarima library in Python. This function combines model parameters (p in range 0-6, q in range 0-6 and d in range 0-2) and determines the best model according to information criteria, which include the Akaike Information Criterion (AIC) among others. AIC is a mathematical method that evaluates how well the data are fitted to a model (Bevans, 2020). It is calculated from the number of independent variables integrated into the model and the maximum likelihood of the model, meaning how well it can reproduce the data. Based on AIC, the best-fit model is the one that can explain the largest amount of variation and at the same time use as few possible independent variables as possible. Code snippet 3 show auto_arima() function and the results being fitted. Afterwards the model is trained its statistical summary is printed.

Input:

```
model = auto_arima(train, start_p=0, max_p=6, start_q=0, max_q=6,
max_d=2, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(order=model.order)
model_fit.fit(train)
print("ARIMA Model Parameters:")
print(model_fit.summary())
```

Output:

```
Best ARIMA Order: (2, 2, 1)
ARIMA Model Parameters:
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   55
Model:               SARIMAX(2, 2, 1)   Log Likelihood                -317.886
Date:                Fri, 24 Nov 2023   AIC                            645.771
Time:                        13:15:26   BIC                            655.623
Sample:                             0   HQIC                           649.560
                                 - 55
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     28.3122     23.636      1.198      0.231     -18.014      74.639
ar.L1         -0.8673      0.127     -6.838      0.000      -1.116      -0.619
ar.L2         -0.5372      0.101     -5.303      0.000      -0.736      -0.339
ma.L1          0.6981      0.112      6.218      0.000       0.478       0.918
sigma2      9333.0125   1184.075      7.882      0.000    7012.268    1.17e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                32.55
Prob(Q):                              0.82   Prob(JB):                         0.00
Heteroskedasticity (H):              25.18   Skew:                             0.31
Prob(H) (two-sided):                  0.00   Kurtosis:                         6.79
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Code snippet 3

Based on AIC, the optimal model is SARIMAX(2, 2, 1). This seems quite different from the ARIMA (4, 1, 2) model that was constructed manually. A SARIMAX model can be defined as an ARIMA model with seasonality (Brownlee, 2017). The first two values from the ACF plot seem to be significant to the model, and only the first value of the PACF plot. The p-value of the coefficients is really low, except from the intercept, indicating that it is indeed a good model. Next the model will be fitted to the train dataset in code snippet 4 and plotted in figure 20.

Input:

```
order = (2, 2, 1)
SARIMAmodel = SARIMAX(train, order=order)
SARIMAmodel = SARIMAmodel.fit()
y_pred = SARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
SARIMAmodel.predict(start=y_pred_df.index[0], end=y_pred_df.index[-1])
sarima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", sarima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"])
/ test) * 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Output:

```
RMSE: 564.7141913057193
MAPE: 5.00%
```

Code snippet 4

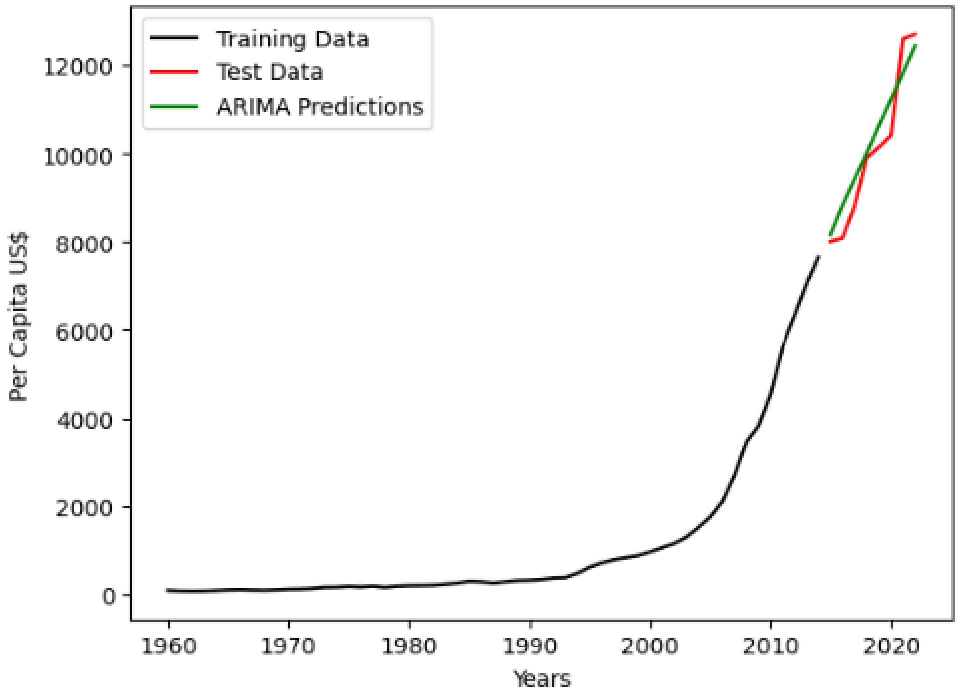

Figure 20: GDP predictions on train set with SARIMAX (2, 2, 2)

Figure 20 illustrates GDP predictions based on the training dataset. The predictions of the SARIMAX (2, 2, 1) model seem to be very close to the actual values, showing that the model is accurate. The RMSE value is 564.71 and MAPE is 5.00%, indicating that the model was off by 5%. Compared with the manual model of (4, 1, 2), this model is more accurate as it has a lower RMSE value Next, 10 step predictions will be made using the SARIMAX (2, 2, 1) model and the entire dataset as seen in code snippet 5 and then plotted, as seen in Figure 21.

Input:

```
forecast_steps = 10
arima_model = SARIMAX(X, order=(2, 2, 1))
arima_fit = arima_model.fit()
forecast = arima_fit.forecast(steps=forecast_steps)
```

Code snippet 5

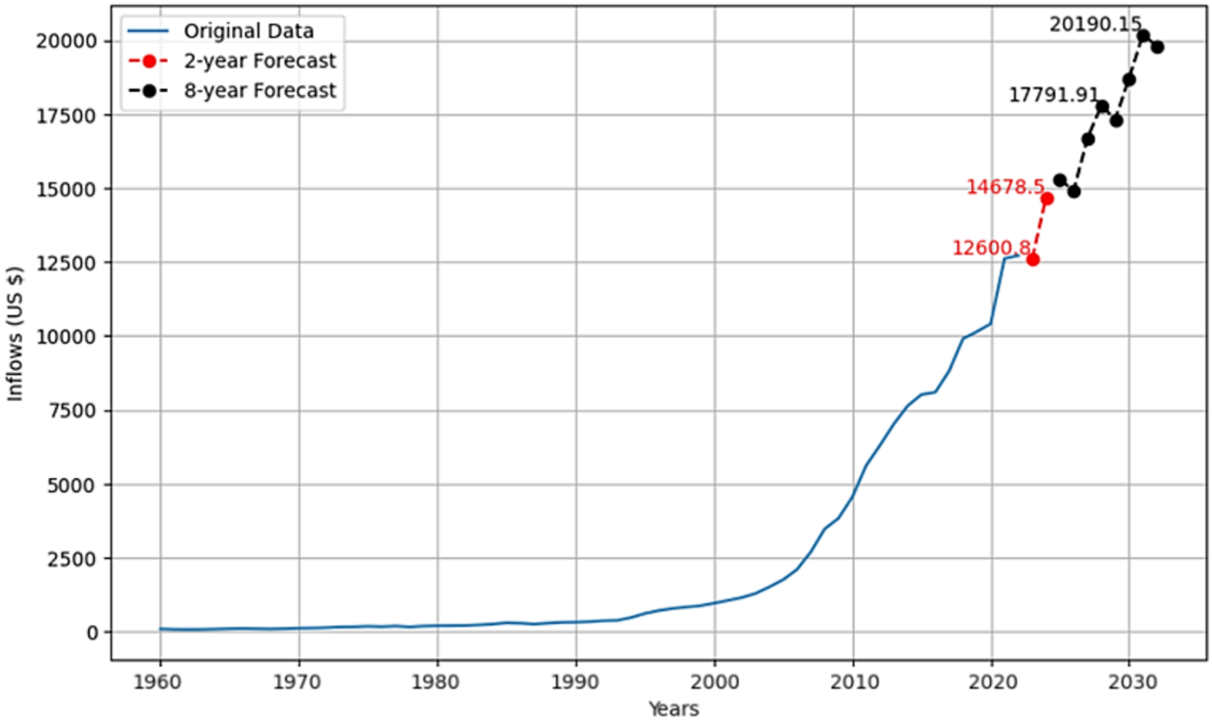

Figure 21: GDP prediction with ARIMA (2, 2, 1) model

Figure 21 shows the predictions made by the SARIMAX (2, 2, 1) model. Based on the model's predictions, GDP Per Capita of China will keep rising reaching an estimated 12,600.8 US$ in 2023 and 14,678.5 US$ in 2024. Subsequently, an upward trajectory is anticipated over the next 8 years, as the GDP Per Capita will keep rising reaching 20,190.15 US$ in 2031 before a projected decline.

3.4 Inflation Forecasting

In this section of the report, Chinese Inflation rate values will be forecasted through the application of an Autoregressive Integrated Moving Average (ARIMA) model.

3.4.1 Dataset Description

The dataset selected for analysis is obtained from Macrotrends (2023) and contains annual time series data of China between 1987 and 2022. It contains 36 observations and the dependent variable is Inflation Rate (%).

3.4.2 Exploratory Data Analysis (EDA)

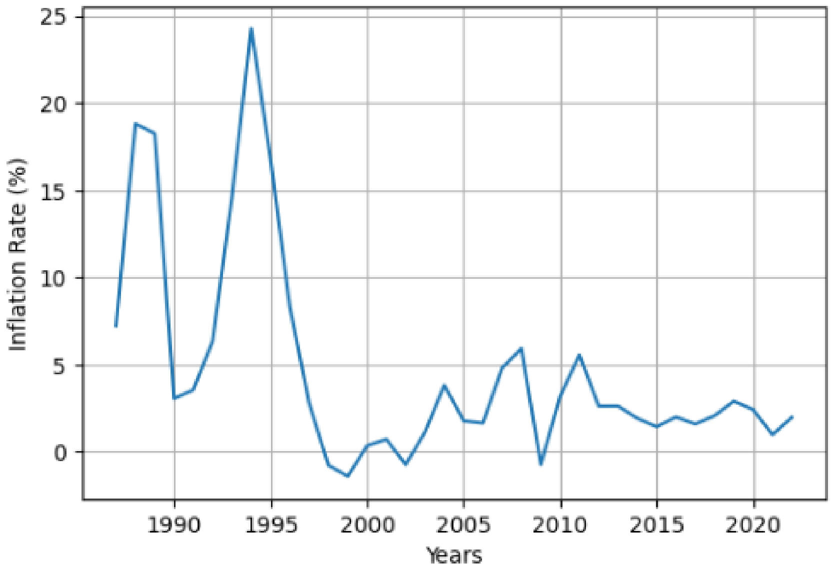

Figure 22: Inflation Rate of China Line plot (1987-2022)

The line plot representing China's Inflation Rate is illustrated in Figure 22. The data exhibit noticeable fluctuations, particularly within the first 20 observations. Notable peaks can be seen in the first half of the dataset.

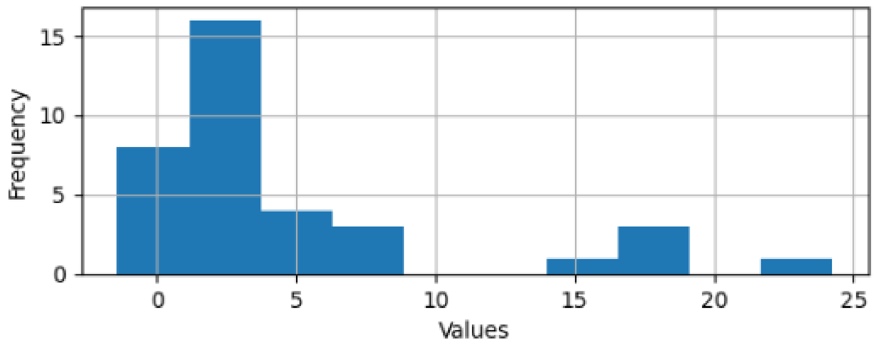

Figure 23: Inflation of China Histogram of frequency distribution of values

The histogram in Figure 23 displays the frequency distribution of values in the dataset. The x-axis represents the values, and the y-axis represents their frequency. Evidently, the majority of values cluster between 0 and 10. However, there are concentrations of values between 14 and 19 and between 12 and 24 outside of the main concentration.

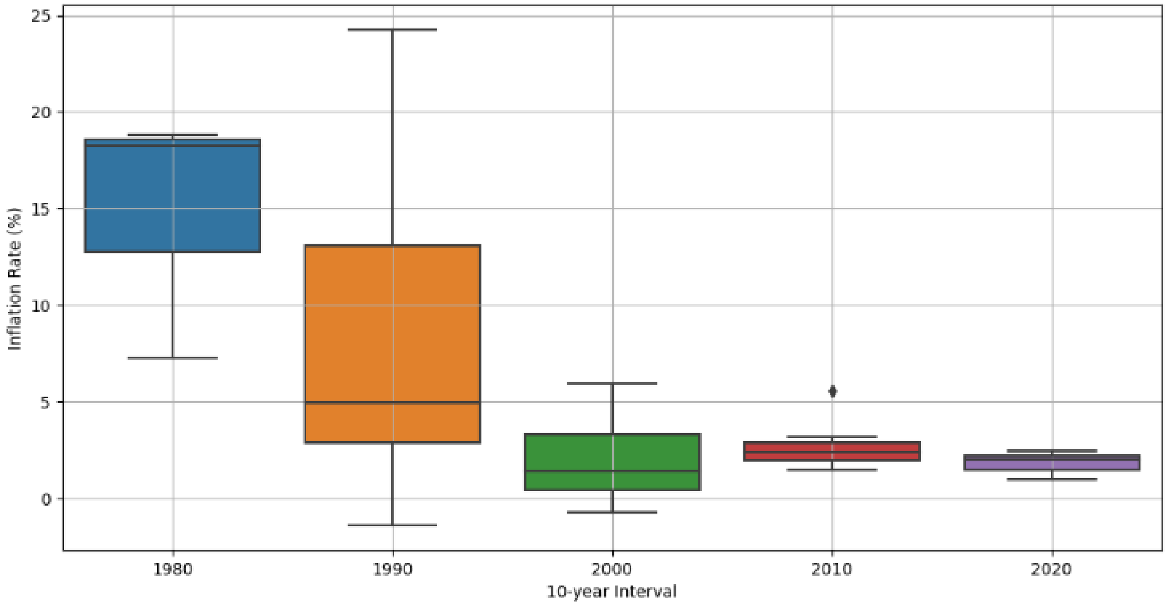

Figure 24: Inflation Rates of China Box and Whisker plots

The box-and-whisker plot, depicted in Figure 24 illustrates the Inflation Rates of China. Notably, a visible trend is observed in both the median and the upper quartile, which consistently descend across the examined years. During the initial two 10-year intervals, the IQR and the length of the whiskers, are quite large. In the subsequent three 10-year intervals, there is a noticeable shrinkage of the IQR, indicating a decreased dispersion of the data. The most pronounced interval is observed in the second box, where the data range extends from nearly 0 to almost 25 percent. This contraction of the IQR reflects a considerable variance in annual Inflation Rates from 1987 to 2022.

3.4.3 Manual model construction

A mentioned in section 3.2, firstly an ARIMA model will be constructed manually with its parameters selected from the Dickey Fuller test (d parameter) and the ACF and PACF plots (p and q parameters respectfully).

3.4.3.1 Stationarity

The d parameter for the ARIMA model will be provided using the adfuller() function from the statsmodel library in Python. The null hypothesis indicates stationarity in the data, while the alternative hypothesis suggests failure to reject the null hypothesis, meaning non stationarity. Code snippet 6 shows the results from the adfuller() function. X is the dependent variable (Inflation Rate (%)) used for this analysis.

Input:

```
series = pd.read_csv('china_inflation_new.csv', skiprows=16)
X = series[' Inflation Rate (%)']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Output:

```
ADF Statistic: -6.496480
p-value: 0.000000
Critical Values:
	1%: -3.689
	5%: -2.972
	10%: -2.625
```

Code snippet 6

The p-value and the ADF statistic are examined in order to accept either hypothesis. Since the p-value is 0.00 (less than 0.05) and the critical values are indeed larger than the ADF value, we can accept the H_0. The data are stationary and do not need to be differenced for the development of the model (d=0).

3.4.3.2 ACF/PACF

Next, the p and q parameters will be selected from the ACF and PACF plots. Figure 25 shows the plots that were created with the use of the statsmodel library in Python. The x axis shows the number of lags and the y axis shows the correlation coefficient value. The number of lags has been limited to 17 for readability reasons.

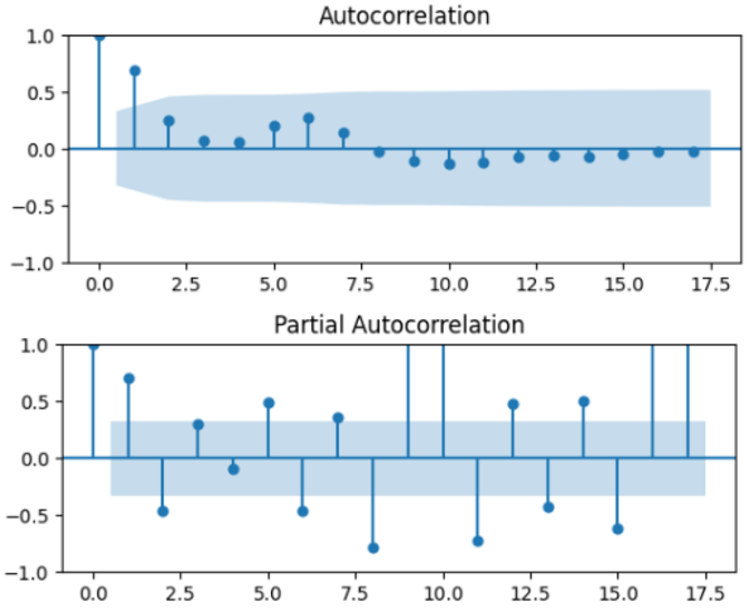

Figure 25: Inflation ACF and PACF

Both ACF and PACF plots reveal notable lags for the first 2 years. Based on these observations, a potential final model could be specified as (2, 1, 2).

3.4.3.3 ARIMA model (2, 1, 2)

Since the first step for building an ARIMA model is the split of the dataset, the 36 observations will be split into 31 training (from 1987 to 2017) and 5 testing (from 2018 to 2022) observations, following a 0.88 split. Next, the ARIMA() function with an order of (2, 1, 2) will be used from the statsmodel library in Python as shown in code snippet 7. The data will then be fitted and produce 5 forecasts from the get_forecast() function. Lastly, all three datasets will be plotted together in Figure 26.

Input:

```
train_size = int(len(X) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]
order = (2, 1, 2)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel.predict(start=y_pred_df.index[0], end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
```

Output:

```
RMSE: 0.7990846713627607
```

Code snippet 7

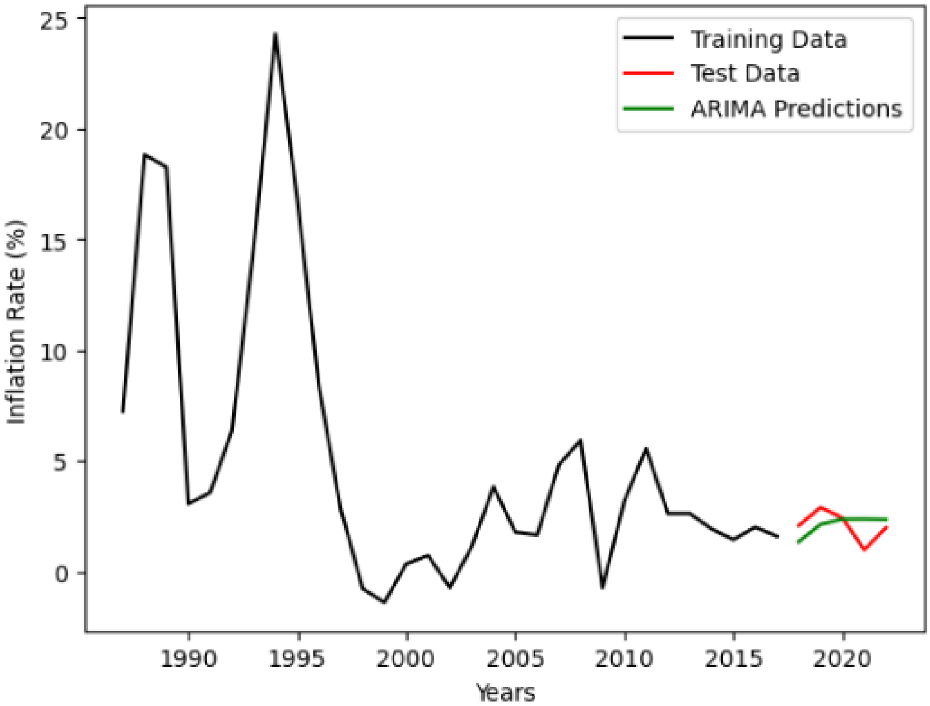

Figure 26: Inflation prediction on train data with ARIMA (2, 1, 2) model

The manually constructed ARIMA (2, 1, 2) model as illustrated in Figure 25, seems to be close to the real values with a RMSE of 0.79. This suggests that the ARIMA Predictions were wrong by 0.79%. By examining Figure 25, the ARIMA predictions seems to have captured the general mean of the test set. Subsequently, the auto_model() function will be used on the training data to identify and recommend the best model for the dataset.

3.4.4 Automatic best model for 0.88 split

In order to select the best ARIMA model for the dataset, the auto_arima function is deployed from the pmdarima library in Python. This function combines model parameters (p in range 0-6, q in range 0-6 and d in range 0-2) and determined the best model according to information criteria, which include the Akaike Information Criterion (AIC) among others. Code snippet 8 show auto_arima() function and the results being fitted. Afterwards the model is trained and its statistical summary is printed.

Input:

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit = model_fit.fit()
print("ARIMA Model Parameters:")
print(model_fit.summary())
```

Output:

```
Best ARIMA Order: (0, 1, 2)
ARIMA Model Parameters:
                               SARIMAX Results
==============================================================================
Dep. Variable:     Inflation Rate (%)   No. Observations:                31
Model:                 ARIMA(0, 1, 2)   Log Likelihood              -85.982
Date:                Mon, 27 Nov 2023   AIC                         177.965
Time:                        23:43:13   BIC                         182.168
Sample:                             0   HQIC                        179.310
                                 - 31
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.0849      0.215      0.394      0.693      -0.337       0.507
ma.L2         -0.6824      0.309     -2.205      0.027      -1.289      -0.076
sigma2        17.2885      6.318      2.736      0.006       4.906      29.671
===================================================================================
Ljung-Box (L1) (Q):                   0.38   Jarque-Bera (JB):                 2.16
Prob(Q):                              0.54   Prob(JB):                         0.34
Heteroskedasticity (H):               0.20   Skew:                             0.60
Prob(H) (two-sided):                  0.02   Kurtosis:                         3.54
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Code snippet 8

Based on code snippet 8, AIC suggests that the optimal model for this dataset is ARIMA(0, 1, 2), which is close with the manually constructed model with an order of (2, 1, 2). It seems that the first 2 lag values were not that significant. The p-value of the model's coefficients are below 0.05 except L1, but the model could still be considered good. The model is then fitted, as shown in code snippet 9 in the train dataset and plotted, as shown in figure 27.

Input:

```
order = (0, 1, 2)
arima_model = ARIMA(train, order=order)
arima_fit = arima_model.fit()
y_pred = arima_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = arima_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"])
/ test) * 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Output:

```
RMSE: 0.8045633361765978
MAPE: 40.86%
```

Code snippet 9

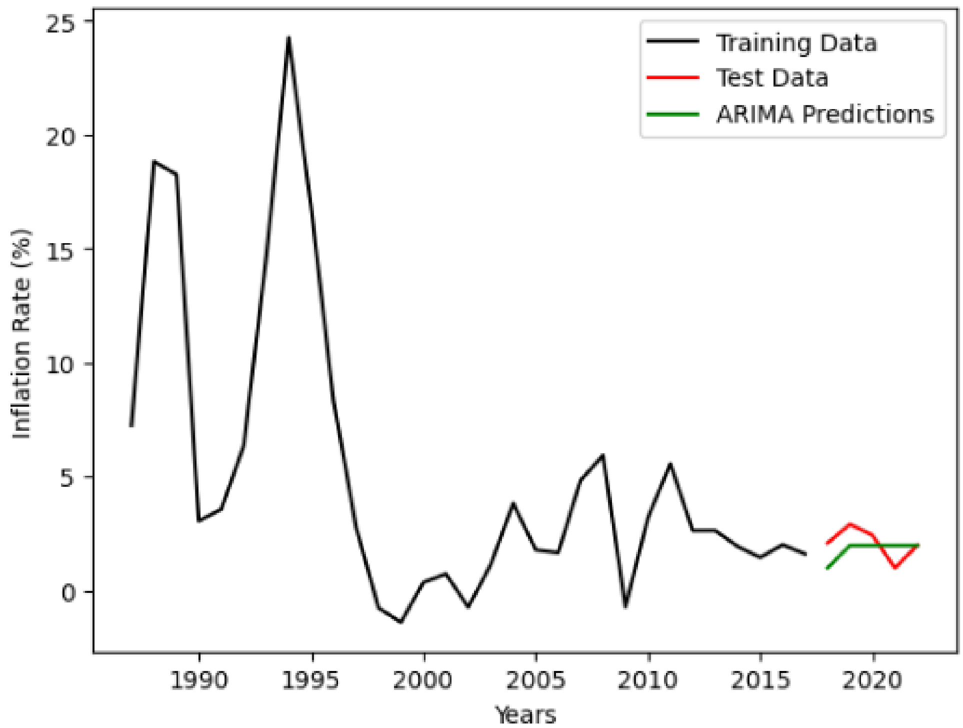

Figure 27: Inflation rate predictions on training set with ARIMA(0, 1, 2)

The ARIMA (0, 1, 2) model predicts different values from the test data with 0.80 RMSE and 40.86% MAPE. Even though 0.80 prediction error is very small, the MAPE depicts a different picture, indicating that the model is off by 40.86%. Compared with the manual model of (2, 1, 2), this model is slightly less as it has a higher RMSE value. Subsequently, a 10-year forecast will be made with the ARIMA (0, 1, 2) model with the entire dataset and then plotted as shown in code snippet 10 and Figure 28 respectively.

Input:

```
forecast_steps = 10
arima_model = ARIMA(X, order=(0, 1, 2))
arima_fit = arima_model.fit()
forecast = arima_fit.forecast(steps=forecast_steps)
```

Code snippet 10

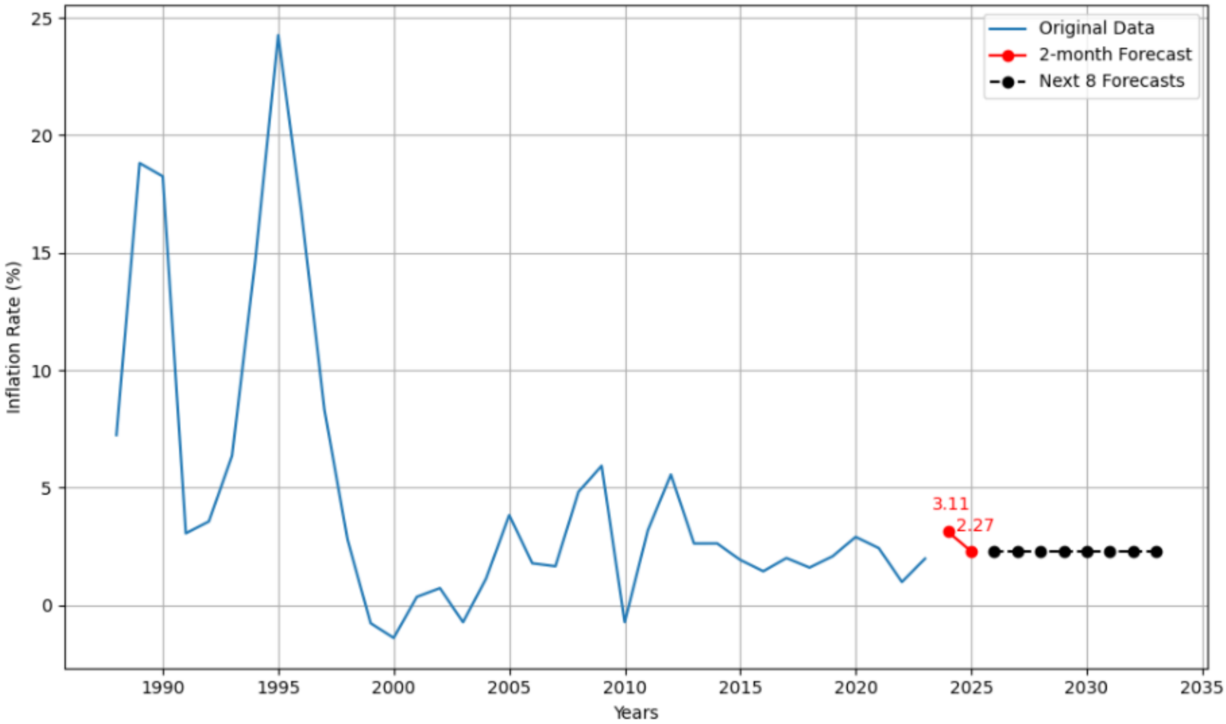

Figure 28: Inflation prediction on ARIMA (0, 1, 2) model

The predictions made by the ARIMA (0, 1, 2) model are depicted in Figure 28. Based on it, China’s inflation rate will rise to 3.11% in 2023 and fall to 2.27% in 2024. The next 8 years are forecasted to stay steady at 2.27% inflation rate.

3.5 FDI Forecasting

In this section of the report, Foreign Direct Investment values in China will be forecasted through the application of an Autoregressive Integrated Moving Average (ARIMA) model.

3.5.1 Dataset Description

The dataset that will be used for forecasting FDI values for China is sourced from Macrotrends (2023) and contains annual time series data of FDI values in China from 1979 to 2022. It contains 44 observations and the dependent variable is Inflows US $.

3.5.2 Exploratory Data Analysis (EDA)

The dataset will undergo an Exploratory Data Analysis to comprehensively examine its structure and characteristics.

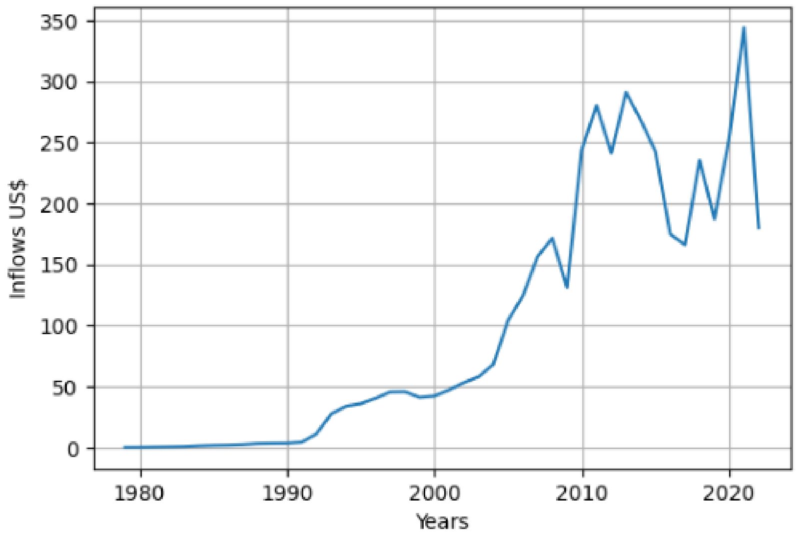

Figure 29: FDI inflows of China Line plot (1978-2022)

The line plot for China's FDI is shown in Figure 29. The values seem to follow an upward trend with some fluctuations towards the last 20 values.

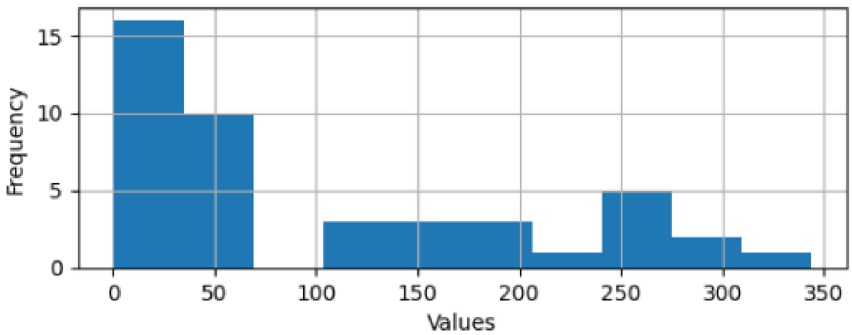

Figure 30: FDI inflows of China Histogram of frequency distribution of values

Figure 30 shows the histogram of the dataset, which shows the frequency distribution of values in the dataset. The x-axis represents the values, and the y-axis represents their frequency. The majority of the values are concentrated between approximately 0 and 60 and a second smaller concentration can be found between 100 and 350.

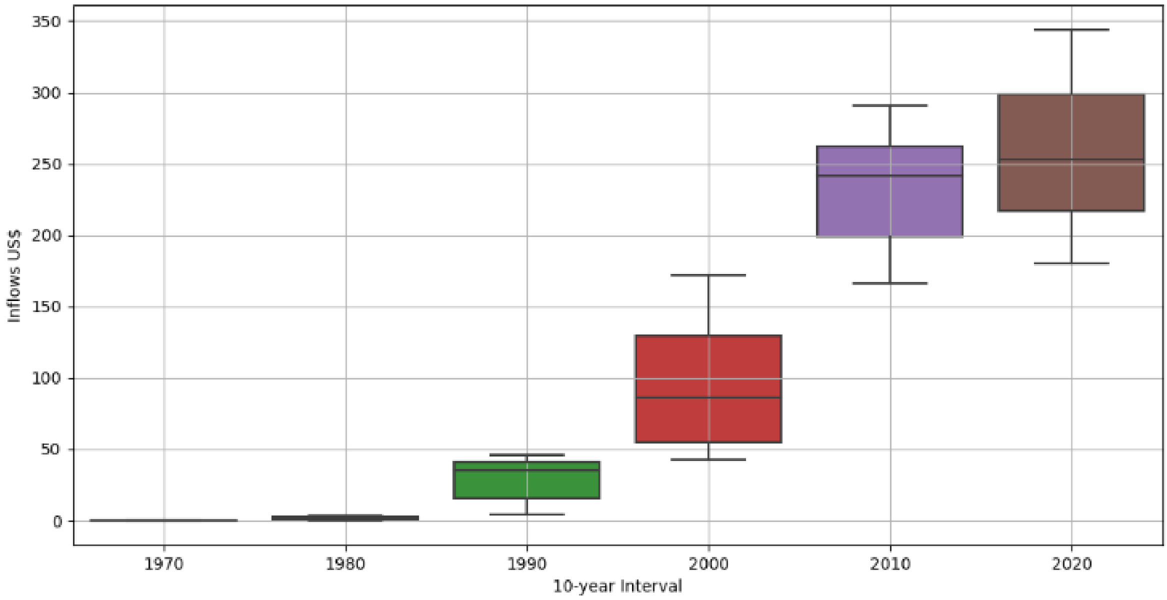

Figure 31: FDI inflows of China Box and Whisker plot (1978-2022)

The box-and-whisker plot, depicted in Figure 31 illustrates the FDI inflows of China. Notably, a visible trend is observed in both the median and the upper quartile, which consistently ascend across the examined years. During the initial four 10-year intervals, the IQR and the length of the whiskers, are expanding gradually. The last two 10-year intervals, seem to be having a similar median close to 250. The most pronounced interval is observed in the last box, where the data range extends from nearly 170 to almost 350 US$. This contraction of the IQR reflects a considerable variance in annual FDI inflows from 1978 to 2022.

3.5.3 Manual model construction

A manual ARIMA model regarding FDI values of China will be developed. Its three parameters (p, d, q) will be selected from the Dickey Fuller test and the ACF and PACF plots.

3.5.3.1 Stationarity

In order to determine the d parameter, the adfuller() function from the stats.model library in Python will be used in order to check for stationarity in the data. The H_0 indicates stationarity in the dataset, while the H_1 shows failure to reject the H_0, meaning the data are not stationary.
Code snippet 11 shows the Dickey Fuller test with X being the dependent variable (Inflows US$).

Input:

```
X= series[' Inflows  US $']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Output:

```
ADF Statistic: -1.393104
p-value: 0.585578
Critical Values:
        1%: -3.627
        5%: -2.946
        10%: -2.612
```

Code snippet 11

By examining the p-value and the ADF Statistic it can be seen that the data is non-stationary. Specifically, the p-value of 0.585578 is greater than 0.05 and the ADF Statistic of -1.393104 is greater than the critical values indicating a failure to reject the H_0 and non-stationarity. This suggests that d=1.

3.5.3.2 ACF/PACF

The next step is to select the p and q parameters. Figure 30 shows the ACF and PACF plots that were created from the statsmodel library in Python. The x axis shows the lag number and the y axis shows the correlation coefficient value. The plots show only the first 20 lags for readability.

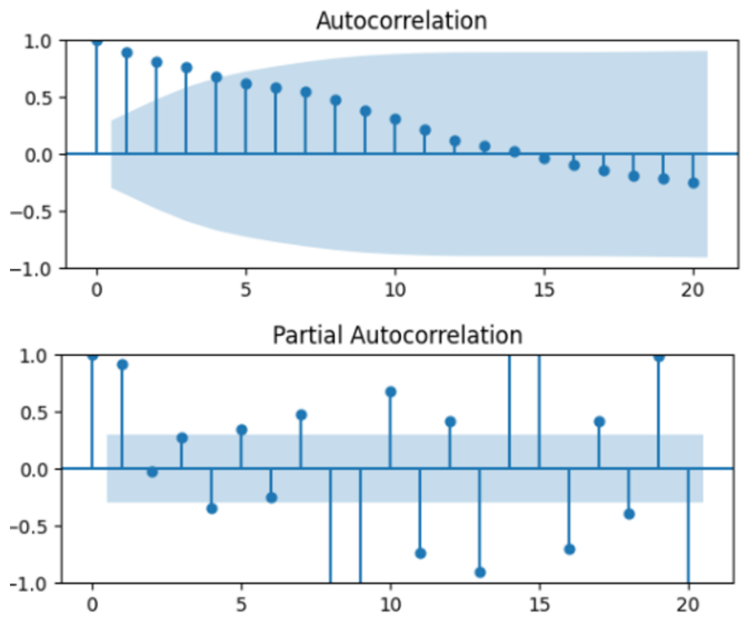

Figure 32: FDI ACF and PACF plots

The ACF in Figure 32 shows that the first 4 lags are significant, while the PACF shows that the first lag is significant. The final model could be indicated as (4, 1, 1).

3.5.3.3 ARIMA model (4, 1, 1)

To begin with, the dataset will be split into 0.88 training set and 0.12 testing set in order to train and evaluate the model. 32 observations (from 1978 to 2016) will be used to train the model and 6 (from 2017 to 2022) to test it from the total of 38 observations. The train set will be used for the ARIMA() function with an order of (4, 1, 1) from the statsmodel library, as seen in code snippet 12. Afterwards it will be fitted and the get_forecast() function will produce 6 observations. Lastly, the 3 datasets will be plotted together as illustrated in Figure 33.

Input:

```
train_size = int(len(y) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]
order = (4, 1, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel.predict(start=y_pred_df.index[0], end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
```

Output:

```
RMSE: 115.35301636066622
```

Code snippet 12

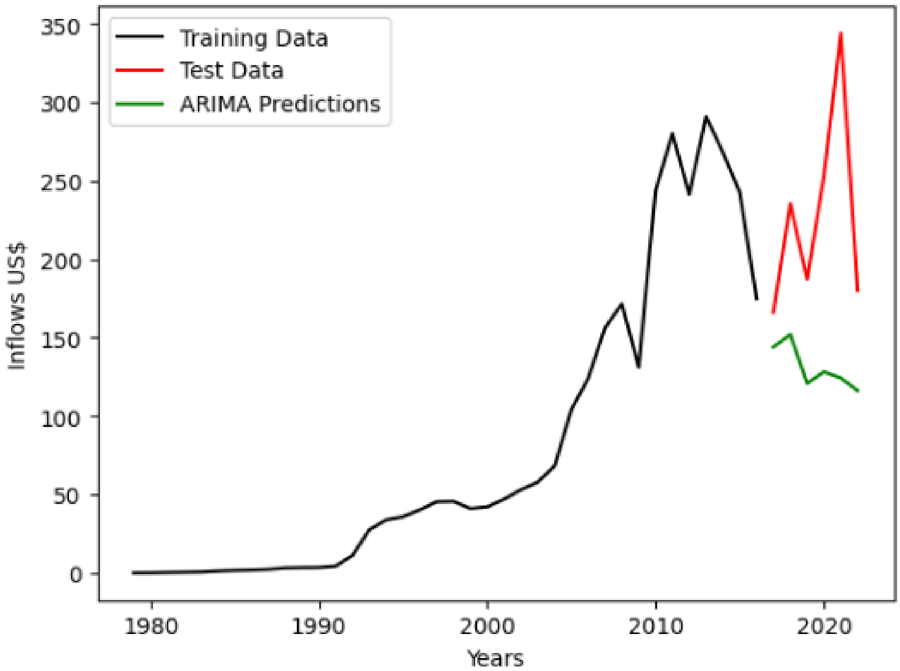

Figure 33: FDI predictions on train data with ARIMA (4, 1, 1) model

The manually constructed ARIMA (4, 1, 1) model exhibits obvious deviations from the observed test values, as seen in Figure 33. The RMSE of the model is 115.35 indicating a considerable level of predictive imprecision. The model seems inadequate in capturing the two big fluctuations and predicted a decline in inflows. Subsequently, the auto_model() function will be deployed in the entire dataset to show the optimal model the most accurate predictions of this dataset.

3.5.4 Automatic best model for 0.88 split

The identification of the best ARIMA model for the dataset will be conducted from the auto_arima function from the pmdarima library in Python. This function systematically explores a range of model parameters (p in range 0-6, q in range 0-6 and d in range 0-2) and selects the most suitable model based on information criteria, which include the Akaike Information Criterion (AIC) among others. Code snippet 13 shows the implementation of the auto_arima()

function and the results being fitted. Afterwards the model undergoes training and a statistical summary is printed.

Input:

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit_result = model_fit.fit()
print("\nARIMA Model Parameters:")
print(model_fit_result.summary())
```

Output:

```
Best ARIMA Order: (0, 1, 0)

ARIMA Model Parameters:
                     SARIMAX Results
==============================================================================
Dep. Variable:        Inflows  US $   No. Observations:           38
Model:             ARIMA(0, 1, 0)  Log Likelihood          -175.773
Date:            Fri, 24 Nov 2023   AIC                     353.546
Time:                   21:45:56   BIC                     355.157
Sample:                        0   HQIC                    354.114
                            - 38
Covariance Type:              opg
==============================================================================
            coef    std err        z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2    783.1880    92.483     8.468     0.000     601.925     964.452
===================================================================================
Ljung-Box (L1) (Q):                0.17   Jarque-Bera (JB):             51.42
Prob(Q):                           0.68   Prob(JB):                      0.00
Heteroskedasticity (H):        11349.97   Skew:                          1.07
Prob(H) (two-sided):               0.00   Kurtosis:                      8.36
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Code snippet 13

According to AIC, the optimal model is ARIMA (0, 1, 0), an indeed different model from the manually constructed ARIMA (4, 1, 1). According to it, no value from the ACF and the PACF plots were significant. The p-value of the coefficients] is 0.000, which suggests a good model. Next, the model will be fit with the train set and plotted. The results are shown in code snippet 14 and Figure 34.

Input:

```
order = (0, 1, 0)
arima_model = ARIMA(train, order=order)
arima_fit = arima_model.fit()
y_pred = arima_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = arima_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"])
/ test) * 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Output:

```
RMSE: 80.3555949553868
MAPE: 20.13%
```

Code snippet 14

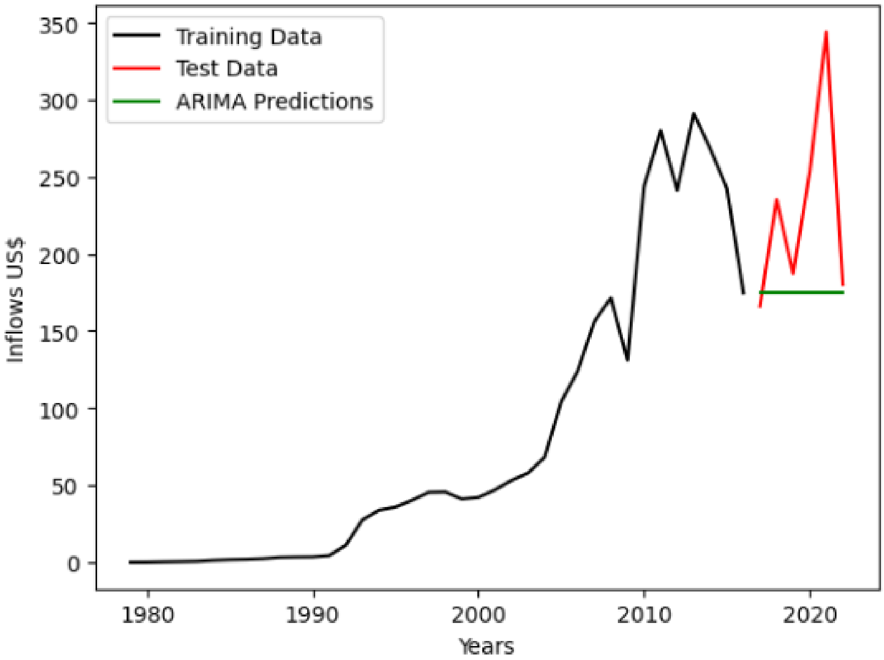

Figure 34: FDI predictions on train set with ARIMA (0, 1, 0)

As figure 34 depicts, the model gave a levelled prediction which does not show the fluctuations of the dataset. The model's RMSE value is 80.35, which means that the prediction has 80.35 US$ margin of error from the test set. Moreover, the MAPE value shows that the model is wrong by 20.13%. Compared with the manual model of (4, 1, 1), this model is more accurate as it has a lower RMSE value. Next, 10 step predictions will be made with the use of this model for the totality of the dataset. Code snippet 15 shows the code for the development of the model and Figure 35 shows the results in a plot.

Input:

```
arima_model = ARIMA(X, order=(0, 1, 0))
arima_fit = arima_model.fit()
forecast_steps = 10
forecast = arima_fit.forecast(steps=forecast_steps)
```

Code snippet 15

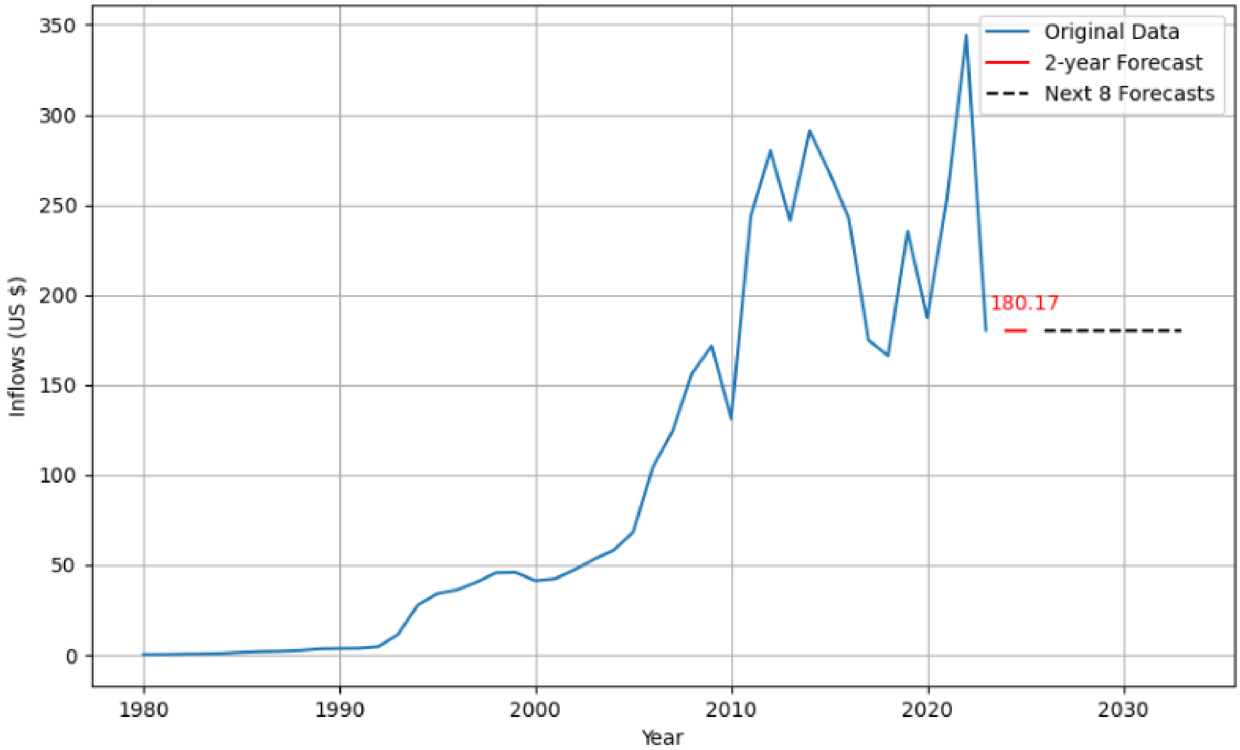

Figure 35: FDI forecast with ARIMA (0, 1, 0)

As can be seen in Figure 35, the newly ARIMA (0, 1, 0) model gives a consistent level forecast of 180.17 US $ for the next 10 years.

3.6

3.7 Imports Forecasting

In this section of the report, Chinese import values will be forecasted through the application of an Autoregressive Integrated Moving Average (ARIMA) model.

3.7.1 Dataset Description

The dataset that will be used for forecasting Import values for China is sourced from Macrotrends (2023) and contains annual time series data values of Chinese imports from 1960 to 2022. It contains 63 observations and the dependent variable is Billions of US $.

3.7.2 Exploratory Data Analysis (EDA)

The dataset will undergo an Exploratory Data Analysis to comprehensively examine its structure and characteristics.

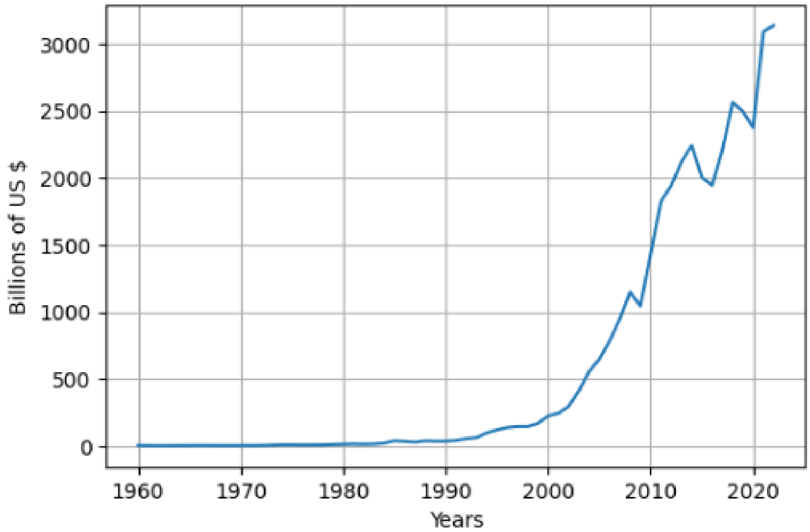

Figure 36: Imports of China Line plot (1960-2022)

China's imports are illustrated in Figure 36. There seems to be a positive growth, with a notable acceleration after 2000.

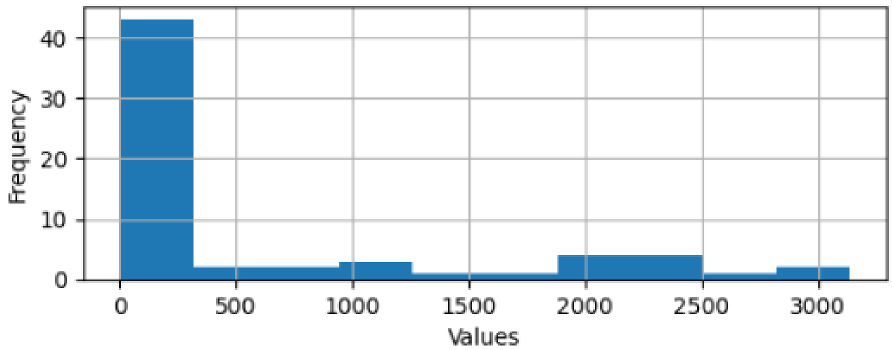

Figure 37: Imports of China Histogram of frequency distribution of values

China's imports' histogram is depicted in Figure 37. The x-axis represents the values, and the y-axis represents their frequency. The majority of the values are concentrated between 0 and 300 while the rest of the values span until 3100.

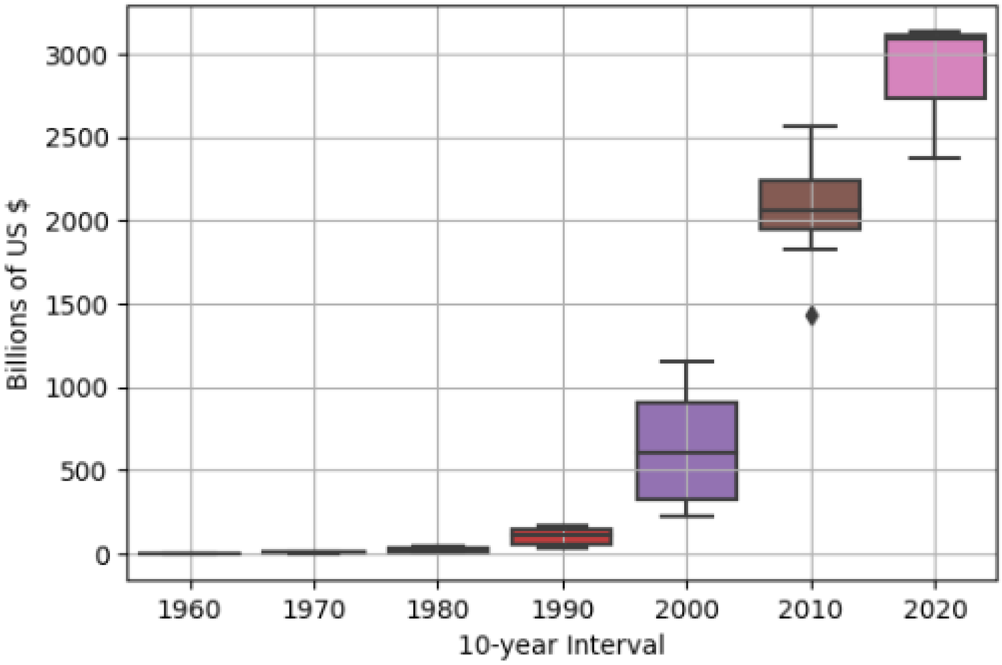

Figure 38: Imports of China Box and Whisker plot (1960-2022)

The box-and-whisker plot, depicted in Figure 38 illustrates the import values of China. Notably, a visible trend is observed in both the median and the upper quartile, which consistently ascend across the examined years. During the initial four 10-year intervals, the IQR of the boxes remain relatively modest. In the subsequent three 10-year intervals, there is a noticeable expansion of the IQR, indicating an increased dispersion of the data. The most pronounced interval is observed in the fifth box, where the data range extends from nearly 400 to almost 1,250 billion US $. This widening of the IQR reflects a considerable variance in annual Chinese imports from 1960 to 2022.

3.7.3 Manual model construction

A manual ARIMA model regarding import values of China will be developed. Its three parameters (p, d, q) will be selected from the Dickey Fuller test and the ACF and PACF plots.

3.7.3.1 Stationarity

The adfuller() function from the stats.model library will be deployed in order to check for the stationarity of the data. The H_0 indicates stationarity in the import's dataset, while the H_1 shows failure to reject the H_0, meaning the data are not stationary. Code snippet 16 shows the Dickey Fuller test with X being the dependent variable (Billions of US $).

Input:

```
X= series[' Billions of US $).']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Output:

```
ADF Statistic: -0.420202
p-value: 0.906671
Critical Values:
        1%: -3.566
        5%: -2.920
        10%: -2.598
```

Code snippet 16

The p-value of the statistic is 0.906671 which is greater than 0.05 and the critical values are greater than the ADF statistic, indicating that a failure to reject the H_0 and that the data are non-stationary. In the manually constructed ARIMA model d equals at least 1.

3.7.3.2 ACF/PACF

The p and q parameters will be selected next. The ACF and PACF generated from the statsmodel library are displayed in Figure 39. The correlation coefficient value is displayed on the y axis, and the lag number is displayed on the x axis. For readability, the plots only display the first 17 lags.

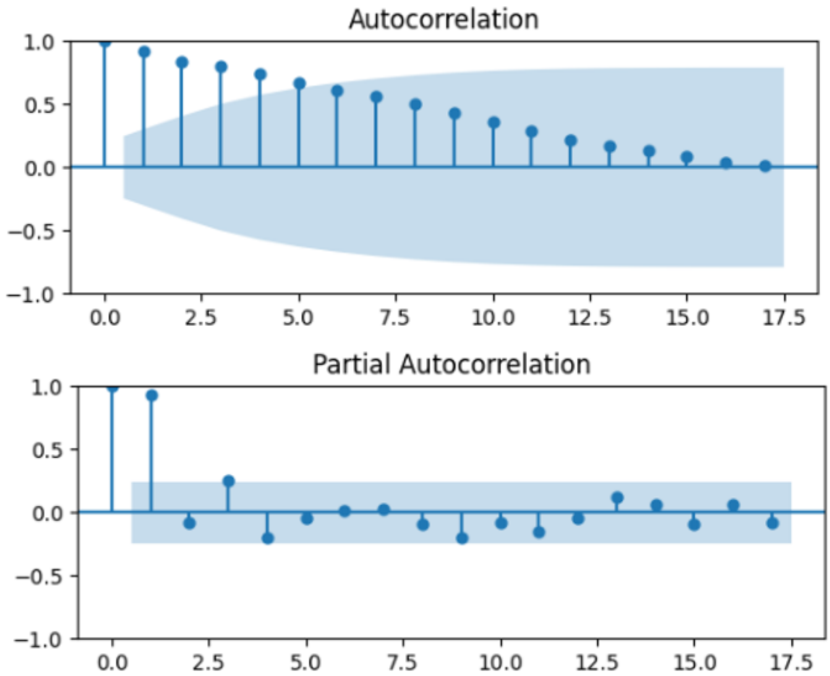

Figure 39: Imports ACF and PACF

The ACF in Figure 39 shows that the first 4-5 lags are significant, while the PACF shows that the first lag is significant. The final model could be indicated as (4, 1, 1).

3.7.3.3 ARIMA model (4, 1, 1)

The first step is to split the dataset into 0.80 train set and 0.12 test set. Of the 63 observations 55 (from 1960 to 2014) will be used to train the model and 8 (from 2015 to 2022) to test it. The ARIMA() function with an order of (4, 1, 1) will be used from the statsmodel library and then the dataset will be fitted with put through the get_forecast() function. This process will produce 8 new forecasted observations, as shown in code snippet 17. Figure 40 shows the 3 datasets plotted together.

Input:

```
train_size = int(len(y) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]
order = (4, 1, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel.predict(start=y_pred_df.index[0], end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
```

Output:

```
RMSE: 564.2058729575895
```

Code snippet 17

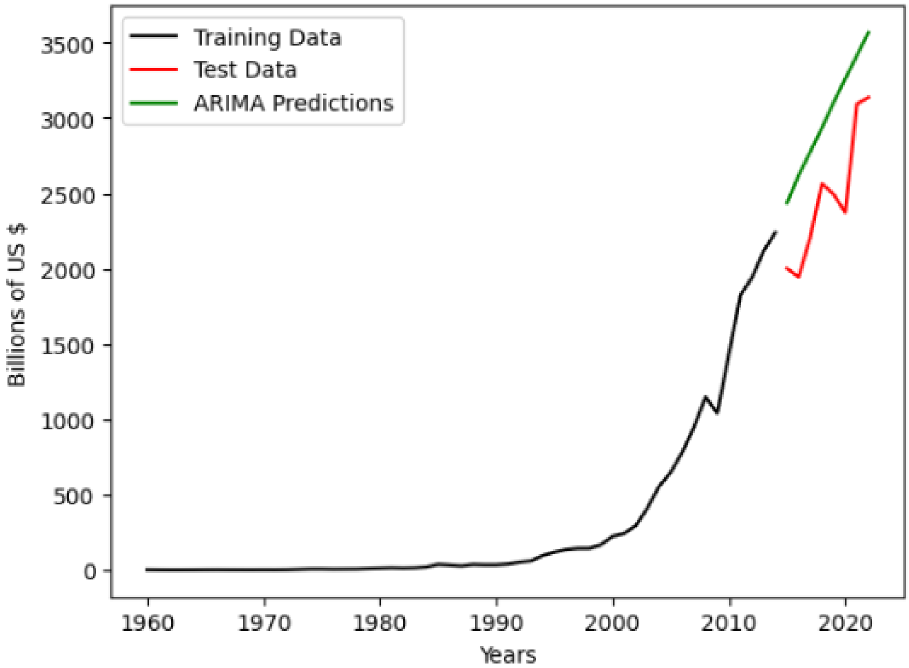

Figure 40: Imports predictions on train data with ARIMA (4, 1, 1) model

According to Figure 40, the ARIMA (4, 1, 1) model shows obvious deviations from the observed test values. As it was trained on data that were rising, it failed to capture the later fluctuations of the test data. Subsequently, the auto_model() function will be deployed to show the optimal model the most accurate predictions of this dataset.

3.7.4 Automatic best model for 0.88 split

The identification of the best ARIMA model for the dataset will be done conducted from the auto_arima function from the pmdarima library in Python. This function systematically explores a range of model parameters (p in range 0-6, q in range 0-6 and d in range 0-2) and selects the most suitable model based on information criteria, which include the Akaike Information Criterion (AIC) among others. The implementation of the auto_arima() function and the results being fitted are shown in code snippet 18. Afterwards the model undergoes training and a statistical summary is printed.

Input:

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit_result = model_fit.fit()
print("\nARIMA Model Parameters:")
print(model_fit_result.summary())
```

Output:

```
Best ARIMA Order: (0, 2, 1)
ARIMA Model Parameters:
                    SARIMAX Results
==============================================================================
Dep. Variable:      Billions of US $   No. Observations:            55
Model:              ARIMA(0, 2, 1)   Log Likelihood            -301.415
Date:               Fri, 24 Nov 2023   AIC                        606.830
Time:                    23:53:12   BIC                        610.770
Sample:                         0   HQIC                       608.345
                             - 55
Covariance Type:              opg
==============================================================================
             coef    std err        z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1     -0.7482      0.049   -15.164      0.000      -0.845      -0.651
sigma2    5017.8671    405.234    12.383      0.000    4223.623    5812.111
===================================================================================
Ljung-Box (L1) (Q):                 0.65   Jarque-Bera (JB):             212.95
Prob(Q):                            0.42   Prob(JB):                       0.00
Heteroskedasticity (H):         11560.71   Skew:                           1.30
Prob(H) (two-sided):                0.00   Kurtosis:                      12.47
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Code snippet 18

According to the AIC value, the optimal model is ARIMA (0, 2, 1), an indeed different model from the manually constructed ARIMA (4, 1, 1) model. The first values of the ACF plot are not significant according to the model, and only the first one is in the PACF plot. However, the p-value of the coefficients is 0.000, which suggests a good model. The model's AIC value is 606.830. Next the model will be used to make predictions using the training set, as seen in code snippet 19. The predictions will be plotted and illustrated in figure 41.

Input:

```
order = (0, 2, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel_fit = ARIMAmodel.fit()
y_pred = ARIMAmodel_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel_fit.predict(start=y_pred_df.index[0], end=y_pred_df.index[-
1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"])
/ test) * 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Output:

```
RMSE: 554.4576583118105
MAPE: 22.29%
```

Code snippet 19

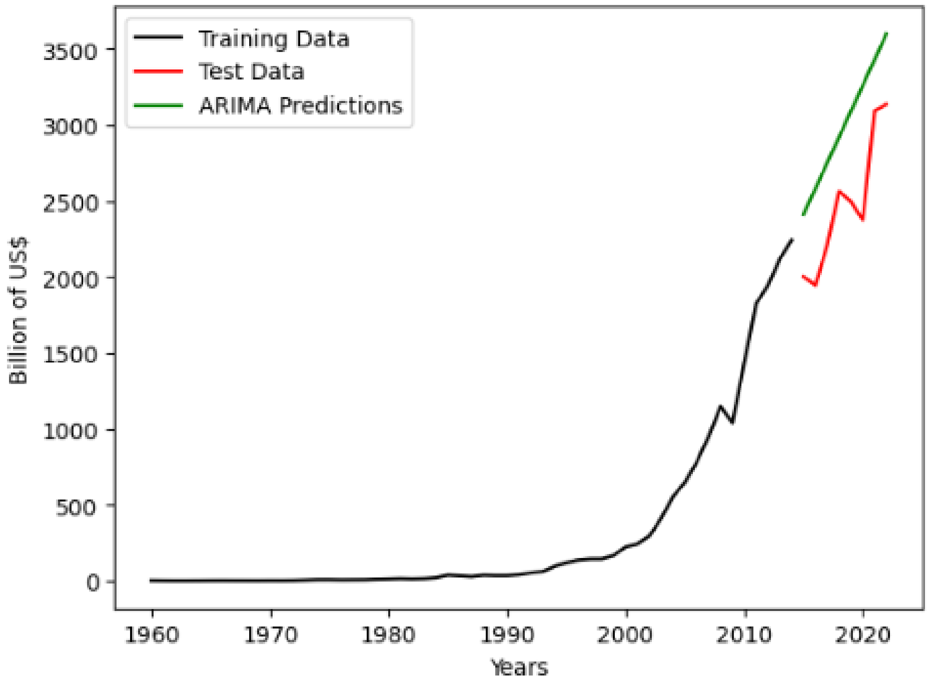

Figure 41: Imports predictions with training set with ARIMA (0, 2, 1)

ARIMA (0, 2, 1) model's predictions are shown in figure 41. The model does seem to have predicted kind of similar values with the test data. Their RMSE value is 554.45, which means that the predictions were wrong by 554.45 billion of US$. Moreover, the MAPE value indicates that the model is 22.29% wrong. However this model is better than the manually constructed ARIMA (4, 1, 1) as it has a lower RMSE value. Subsequently, this model will be used to make 10 step predictions using the whole dataset. Code snippet 20 shows the code for the development of the model and Figure 42 shows the results in a plot.

Input:

```
model = ARIMA(X, order=(0, 2, 1))
results = model.fit()
forecast_steps = 10
forecast_index = np.arange(len(X), len(X) + forecast_steps)
forecast_values =
results.get_forecast(steps=forecast_steps).predicted_mean
```

Code snippet 20

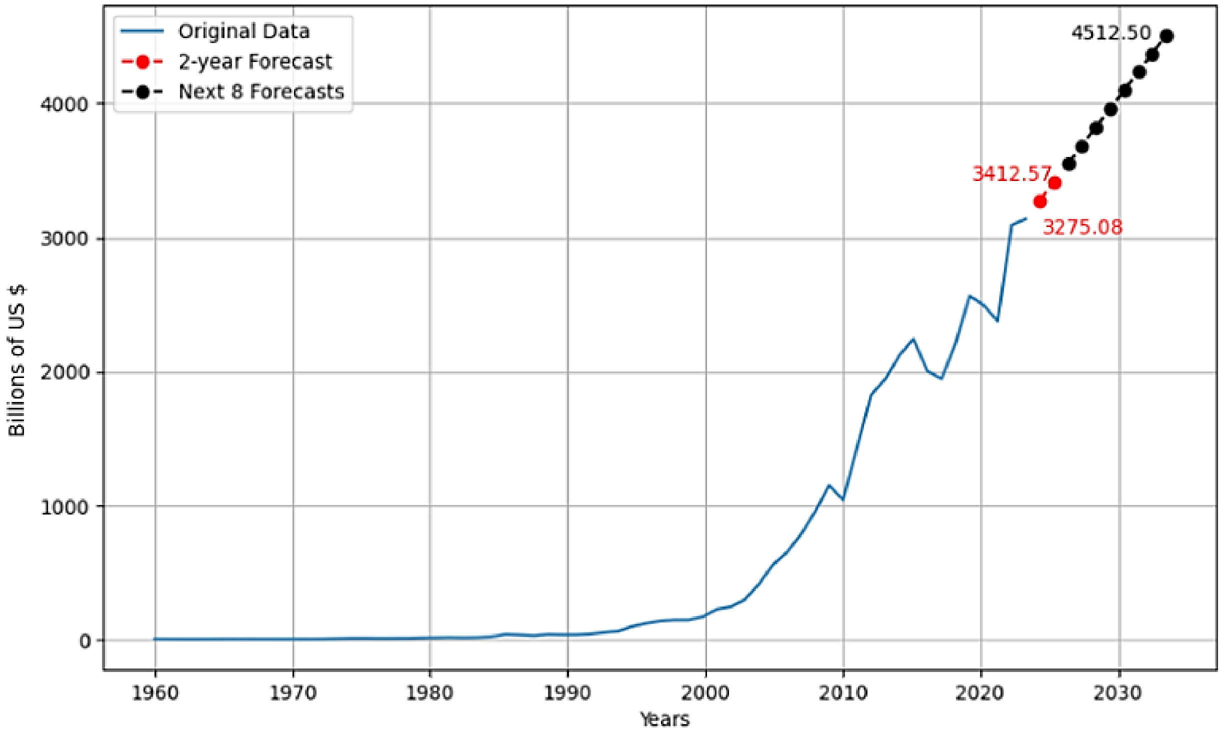

Figure 42: Imports Forecast with ARIMA (0, 2, 1) model

The forecasts derived from the ARIMA (0, 2, 1) model project an upward trend of imports in China. Predictions suggest an increase to 3,275.08 billion of US$ in 2023 and a further incline to 3,412.57 billion of US$ in the next year. This trajectory is expected to persist over the next 8 years, reaching 4512.50 billion of US$ in 2032, as visually represented in Figure 42.

3.8 Exports Forecasting

In this section of the report, Chinese export values will be forecasted through the application of an Autoregressive Integrated Moving Average (ARIMA) model.

3.8.1 Dataset Description

The dataset that will be used in order to forecast Chinese export values is sourced from Macrotrends (2023) and contains annual time series data of Chinese exports between 1960 and 2022. It contains 63 observations and the dependent variable is Billions of US $.

3.8.2 Exploratory Data Analysis (EDA)

In order to better understand the data, the time series will be analysed further

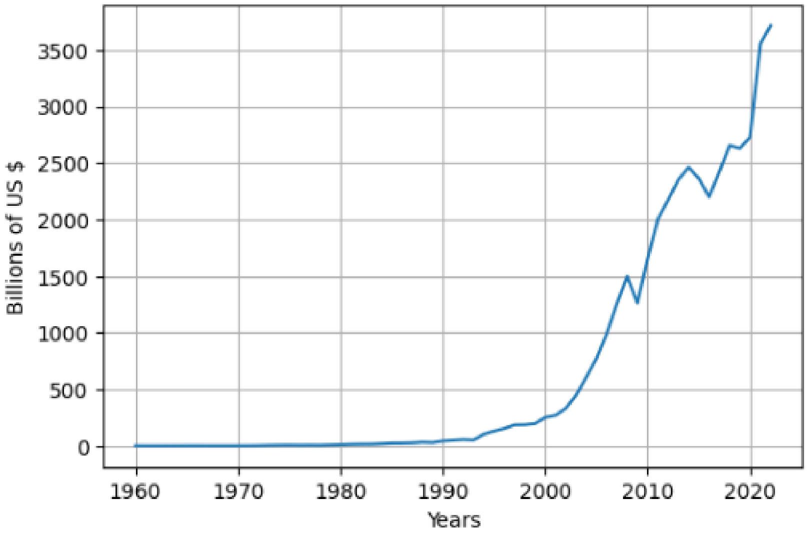

Figure 43: Exports of China Line plot (1960-2022)

In Figure 43, China's exports' line plot can be seen. There is a positive upward trend in the data that gets more intense after 2000.

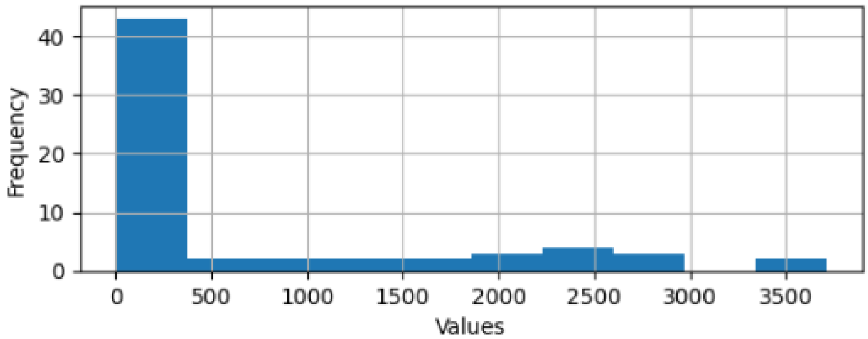

Figure 44: Exports Density plot

The dataset's histogram is illustrated in Figure 44. The x-axis represents the values, and the y-axis represents their frequency. The majority of the values are concentrated between approximately 0 and 40, spanning up until 3000. Some outliners are evident close to 3500.

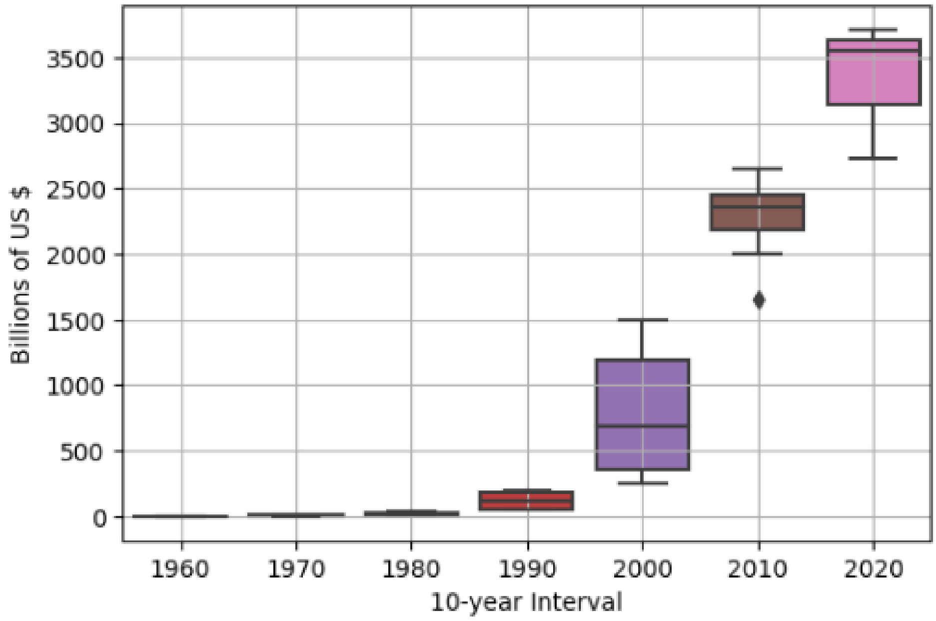

Figure 45: Exports of China Box and Whisker plot (1960-2022)

The box-and-whisker plot, depicted in Figure 45 illustrates the exports of China. Notably, a visible trend is observed in both the median and the upper quartile, which consistently ascend across the examined years. During the initial four 10-year intervals, the IQR and the length of the whiskers, are increasing at a very gradual pace. The next three 10-year intervals, seem to be ascending swiftly. The most pronounced interval is observed in the fifth box, where the data range extends from nearly 350 to 1500 billion US$. This contraction of the IQR reflects a considerable variance in annual exports from 1960 to 2022.

3.8.3 Manual model construction

In this part, an ARIMA model will be constructed manually with its parameters selected from the Dickey Fuller test (d parameter) and the ACF and PACF plots (p and q parameters respectfully).

3.8.3.1 Stationarity

The d parameter for the ARIMA model will be provided using the adfuller() function from the statsmodel library in Python. The null hypothesis indicates stationarity in the data, while the alternative hypothesis suggests failure to reject the null hypothesis, meaning non stationarity. Code snippet 21 shows the results from the adfuller() function. X is the dependent variable (Billions of US $) used for this analysis.

Input:

```
series = pd.read_csv('china-exports.csv', skiprows=16)
X = series[' Billions of US $']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Output:

```
ADF Statistic: 3.209569
p-value: 1.000000
Critical Values:
    1%: -3.560
    5%: -2.918
    10%: -2.597
```

Code snippet 21

The p-value and the ADF statistic are examined in order to accept either hypothesis. Since the p-value is greater than 0.05 and the critical values are smaller than the ADF, we fail to accept the H_0, meaning that at least one level of differencing is required. Because the p-value is 1, the d parameter will be set to 2.

3.8.3.2 ACF/PACF

From the statsmodel libraty, plots ACF and PACF were created in order to accurately select the p and q parameters in Figure 46. The x axis shows the lag number and the y axis shows the correlation coefficient value. The plots show only the first 20 lags for readability.

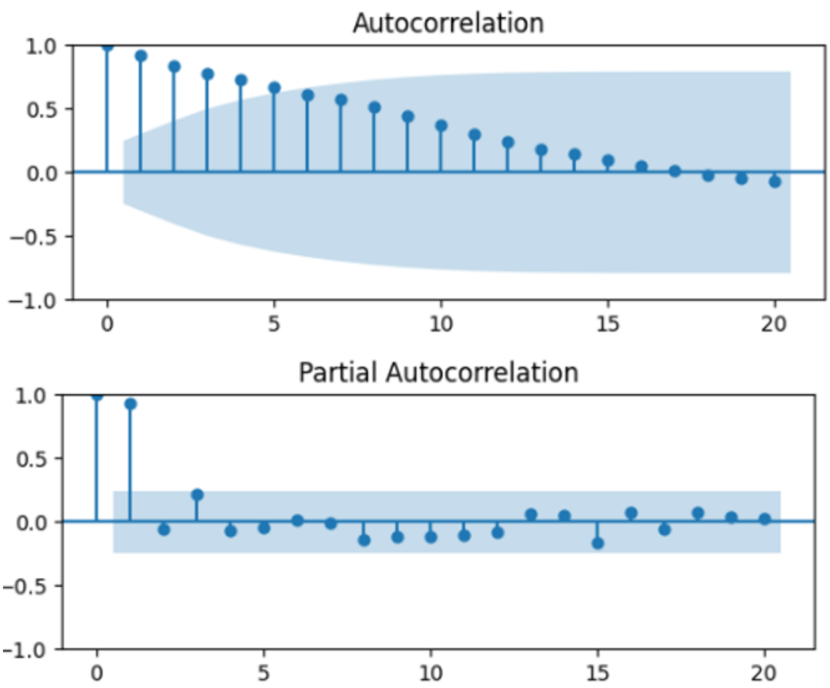

Figure 46: Exports ACF and PACF

The ACF plot in Figure 46 shows that the first 5 lags are significant, while the PACF plot shows that the first lag is significant. The final model could be indicated as (5, 2, 1).

3.8.3.3 ARIMA model (5, 2, 1)

In order to create the ARIMA model, the dataset the dataset will be split into 0.88 training set and 0.12 testing set. From 63 observations, the 55 (from 1960 to 2014) will be used to train the model and 8 (from 2015 to 2022) to test it. Code snippet 23 depicts the ARIMA () function, the fitting of the model, the get_forecast() function for the 8 forecasts that will be projected while Figure 47 illustrates a plot of the 3 datasets.

Input:

```
train_size = int(len(y) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]
order = (5, 2, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel.predict(start=y_pred_df.index[0], end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
```

Output:

```
RMSE: 553.164372650585
```

Code snippet 23

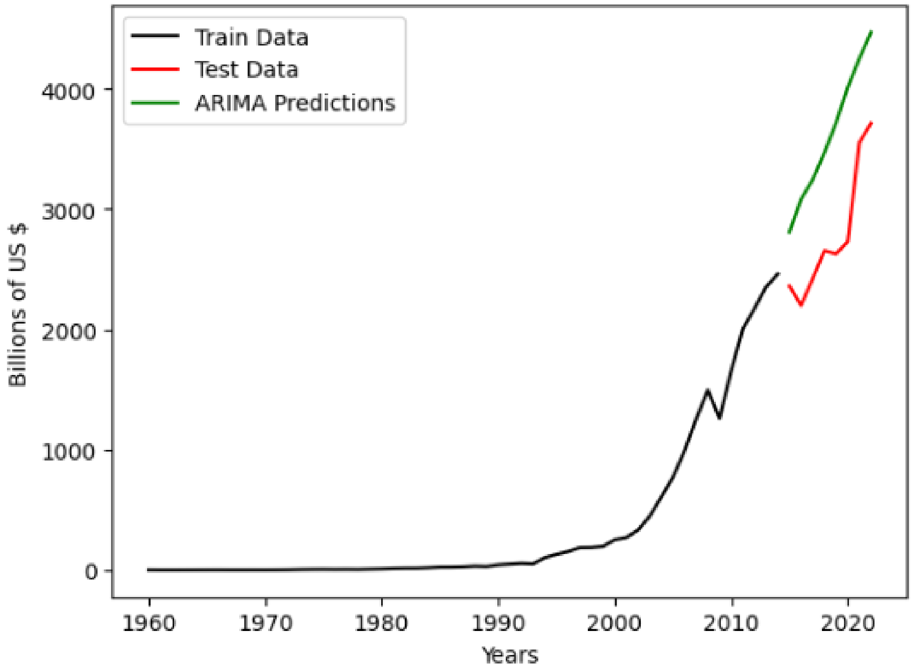

Figure 47: Exports predictions on train data with ARIMA (5, 2, 1) model

The manually constructed ARIMA (5, 2, 1) model exhibits obvious deviations from the observed test values, as seen in Figure 47. The RMSE of the model is 553.16 indicating a small level of predictive imprecision. However, the model seems to continue the rising trend of the data and does not capture the decline that appears in the test data. Subsequently, the auto_model() function will be deployed on the entire dataset to show the optimal model with the most accurate predictions of this dataset.

3.8.4 Automatic best model for 0.88 split

In order to determine the best ARIMA model for this dataset, the auto_arima function from the pmdarima library in Python will be used. This function systematically explores a range of model parameters (p in range 0-6, q in range 0-6 and d in range 0-2) and selects the most suitable model based on information criteria, which include the Akaike Information Criterion (AIC) among others. Code snippet 24 shows the implementation of the auto_arima() function and the results being fitted. Afterwards the model undergoes training and a statistical summary is printed.

Input:

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit_result = model_fit.fit()
print("\nARIMA Model Parameters:")
print(model_fit_result.summary())
```

Output:

```
Best ARIMA Order: (0, 2, 3)
ARIMA Model Parameters:
                               SARIMAX Results
==============================================================================
Dep. Variable:      Billions of US $   No. Observations:                   55
Model:                ARIMA(0, 2, 3)   Log Likelihood                -307.571
Date:               Fri, 24 Nov 2023   AIC                            623.142
Time:                       03:08:14   BIC                            631.023
Sample:                            0   HQIC                           626.172
                                - 55
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.9674      0.085    -11.356      0.000      -1.134      -0.800
ma.L2         -0.0542      0.190     -0.285      0.775      -0.427       0.318
ma.L3          0.3430      0.170      2.022      0.043       0.011       0.675
sigma2      6230.5853    750.258      8.305      0.000    4760.106    7701.064
===================================================================================
Ljung-Box (L1) (Q):                   0.11   Jarque-Bera (JB):               746.57
Prob(Q):                              0.74   Prob(JB):                         0.00
Heteroskedasticity (H):           17114.88   Skew:                            -2.94
Prob(H) (two-sided):                  0.00   Kurtosis:                        20.42
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Code snippet 24

According to the function, which is based on the AIC value, the best model for the dataset is of order (0, 2, 3), quite different from the manually selected ARIMA (5, 2, 1) model. It seems that none of the values in the ACF plot were significant for the model, but the first three values of the PACF plot were significant. The p-values of the coefficients are all below 0.05 except for L2, but is can still be considered a good model. The model's AIC value is 623.142. Next the model will be fit with the training set as shown in code snippet 25 and plotted, as shown in Figure 48.

Input:

```
order = (0, 2, 3)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel_fit = ARIMAmodel.fit()
y_pred = ARIMAmodel_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] =
ARIMAmodel_fit.predict(start=y_pred_df.index[0], end=y_pred_df.index[-
1])
arima_rmse = np.sqrt(mean_squared_error(test,
y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"])
/ test) * 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Output:

```
RMSE: 615.8541419320143
MAPE: 21.84%
```

Code snippet 25

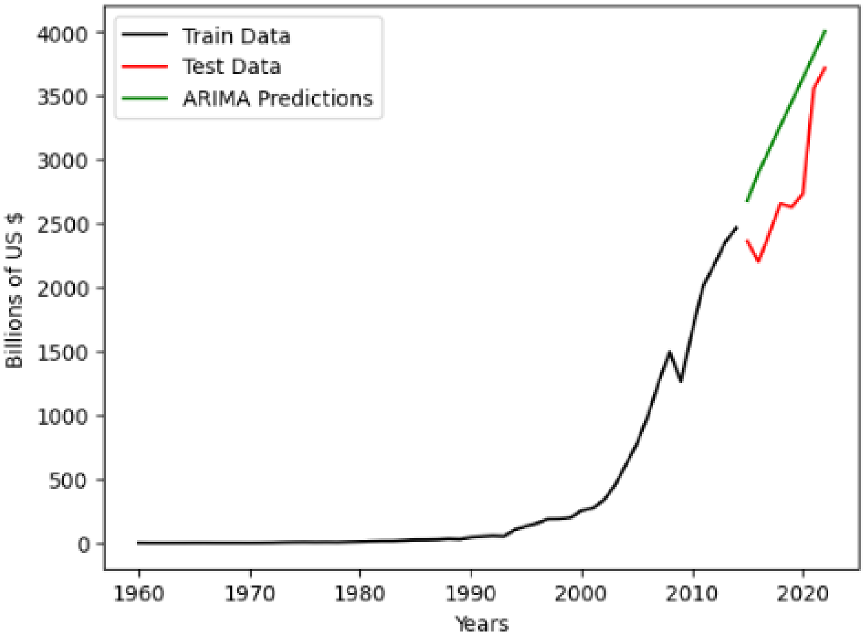

Figure 48: Exports predictions made with the training set with ARIMA (0, 2, 3)

As seen in figure 48, ARIMA (0, 2, 3) model's predictions are similar but a little different that the actual values. The model's RMSE value suggests that the model is off by 615.85 billion of US$, which is 21.84%. Compared to the manual ARIMA (5, 2, 1) model, this one seems to be slightly less accurate as it has a larger RMSE value. Next, 10 step predictions will be made with the use of this model for the totality of the dataset. Code snippet 25 shows the code for the development of the model and Figure 49 shows the results in a plot.

Input:

```
model = ARIMA(X, order=(0, 2, 3))
results = model.fit()
forecast_steps = 10
forecast_values =
results.get_forecast(steps=forecast_steps).predicted_mean
```

Code snippet 25

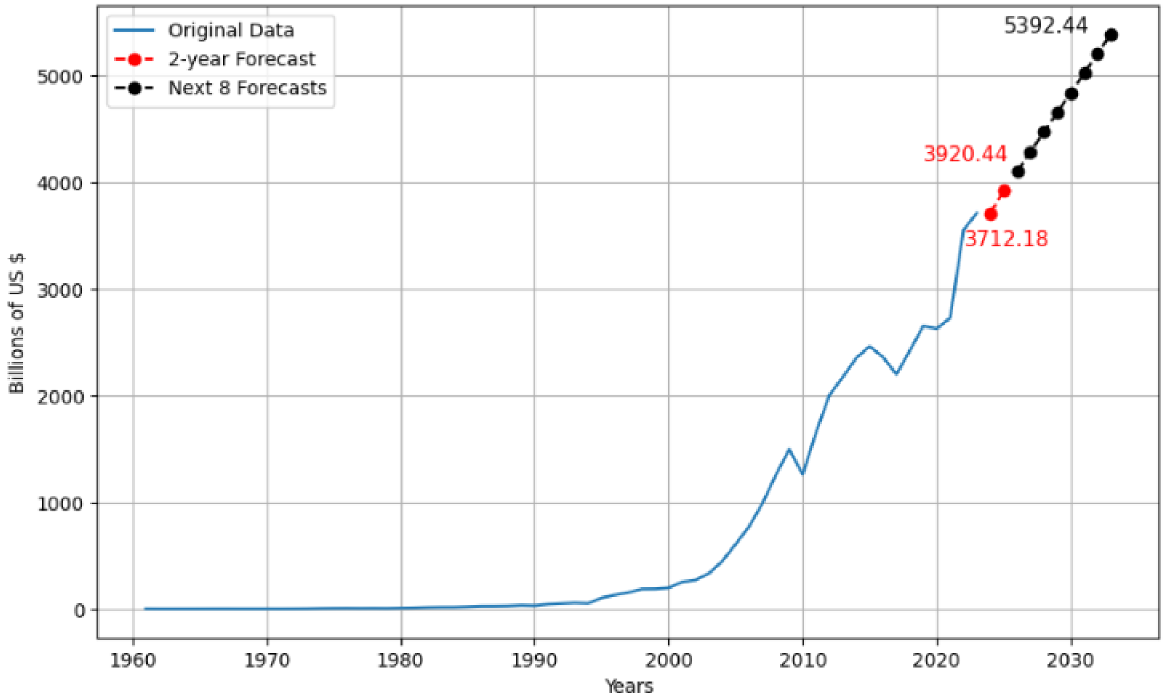

Figure 49: Exports Forecast with ARIMA (0, 2, 3) model

The forecasts derived from the ARIMA (0, 2, 3) model project an upward trend in Chinese exports. Predictions suggest an increase to 3712.18 US$ in 2023 and a further rise to 3920.44 US$ in the next year. This trajectory is expected to persist over the next 8 years, reaching 5392.44 US$ in 2032, as visually represented in Figure 49.

4 Comparative Results and Discussion

In the experiments that took part in sections 3.3 to 3.7, ten ARIMA model were developed. Five of them were created manually from the ACF and PACF plots and the adfuller() function. Subsequently, five new models were constructed, which were optimal for the specific dataset.

The models and their forecasts are presented in Table 4.

Table 4: Forecasts of ARIMA models for all metrics (2023-2032)

Year	Forecast for GDP Per Capita	Forecast for Inflation Rate	Forecast for Imports	Forecast for Exports	Forecast for FDI
Manually constructed models	ARIMA (4, 1, 2)	ARIMA (2, 1, 2)	ARIMA (4, 1, 1)	ARIMA (5, 2, 1)	ARIMA (4, 1, 1)
Optimal models	SARIMAX (2, 2, 1)	ARIMA (0, 1, 2)	ARIMA (0, 2, 1)	ARIMA (0, 2, 3)	ARIMA (0, 2, 1)
RMSE	564.71	0.80	554.45	615.185	80.35
MAPE	5.00%	40.86%	22.29%	21.84%	20.13%
2023	12600.801514	3.115434	3275.085283	3712.180218	180.166881
2024	14678.500482	2.279638	3412.576458	3920.447852	180.166881
2025	15295.853520	2.279638	3550.067632	4104.448050	180.166881
2026	14910.495259	2.279638	3687.558806	4288.448247	180.166881
2027	16695.318436	2.279638	3825.049981	4472.448445	180.166881
2028	17791.912086	2.279638	3962.541155	4656.448642	180.166881
2029	17323.365909	2.279638	4100.032329	4840.448839	180.166881
2030	18710.138270	2.279638	4237.523504	5024.449037	180.166881
2031	20190.148504	2.279638	4375.014678	5208.449234	180.166881
2032	19812.701847	2.279638	4512.505852	5392.449432	180.166881

For the GDP dataset, a SARIMAX (2, 2, 1) model was chosen indicating that the first 2 lags of the ACF plot were significant (hence q=2), the data was differenced 2 times (hence d=2) and the only the first lag of PACF plot was significant (hence p=1). This model differs a lot from the manually selected model of (4, 1, 2). It's MAPE value is 5.00%. As seen in Table 4, the SARIMAX (4, 1, 2) model shows continous growth, which seems to be immitating the trend of the previous 10 years of the dataset.

For the inflation dataset, an ARIMA(0, 1, 2) model was developed indicating that there are no significant lags in the ACF plot (hence p=0), the data was differenced once (hence d=1) and the first 2 lags of the PACF were significant for the model (hence q=2). The model shows some similarity with the manually selected model of (2, 1, 2) and its MAPE value is 40.86%. Lastly, the predictions project a small rise and then a stable forecast for the last 9 years. This prediction may be coming from the trend that the dataset is showing. While the data start with big fuctuations (as seen in the line plot in Figure 28), gradually it starts leveling.

For the imports dataset, an ARIMA (0, 2, 1) model was chosen indicating that no lags from the ACF plot were significant (hence q=0), the data was differenced 2 times (hence d=2) and only the first lag of the PACF plot was significant (hence q=1). The model shows very little similarity with the manually selected model of (4, 1, 1) and its MAPE value is 22.29%. The data in Table 4 indicate a slow future growth for China's imports.

For the exports dataset, an ARIMA (0, 2, 3) model was chosen indicating that no significant plots in the ACF plot (hence p=0), the data was differenced twice (hence d=2) and the first 3 lags of the PACF were significant for the model (hence q=3). The model shows little similarity with the manually selected model of (5, 2, 1) and its MAPE value is 21.84%. Lastly, the model notably indicates a rapid rise in Chinese exports.

For the FDI dataset, an ARIMA (0, 1, 0) model was chosen indicating that no significant lags, neither in the ACF plot nor in the PACF plot (hence p=0, q=0) and that the data was differenced once (hence d=1). The model shows little similarity with the manually selected model of (4, 1, 1) and its MAPE value is 20.13%. Lastly, the model predicts a stable value for FDI inflows for the next 10 years.

In order to evaluate the models fairly the MAPE measure will be used. Based on it, the SARIMAX (2, 2, 1) model fo GDP seems to be showing the most accurate predictions with 5.00% error. Next is ARIMA (0, 2, 1) for FDI with 20.13%, ARIMA (0, 2, 3) for exports with 21.84% error and ARIMA (0, 2, 1) for imports with 22.29%. The model with the least accurate predictions is ARIMA (0, 1, 2) for Inflation rate with 40.86% error. The depictions of their predictions and be seen together in Figure 50.

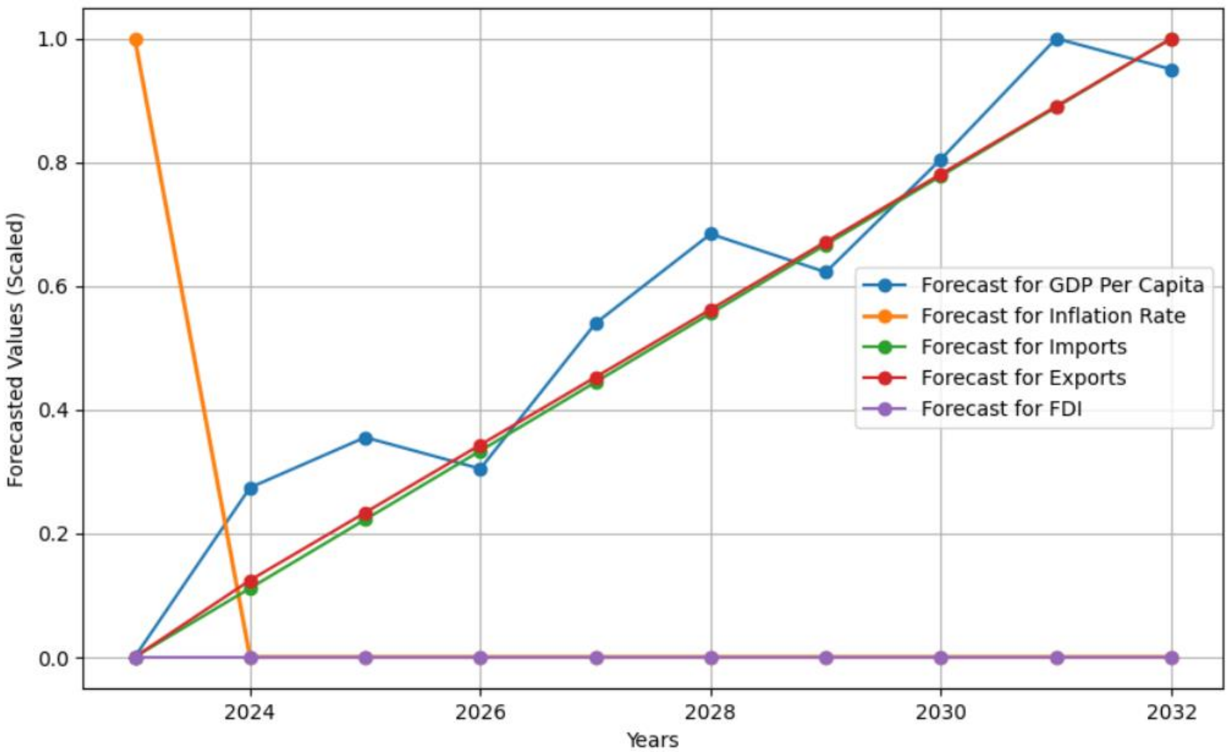

Figure 50: Forecasted Values of China (2023-2032)

In conclusion, Figure 50 shows the forecasted values of Table 4 scaled, with the use of the MinMaxScaler() function. All models seem to be forecasting a positive trajectory for China's economy with reduced inflation (orange line) and increased GDP and trade values. FDI inflows also remain steady.

5 Project Management

The project was executed according to the following timeline:

Project Start Date: Wednesday, 19 July 2023
Project End Date: Thursday, 7 December 2023

The idea of this project dates back on 19/07/2023 when I approached Dr. Mark Johnston and discussed potential project topics. Following the finalisation of the topic in September, a project plan was created in order to monitor and evaluate the project's progress.

The topic selection and initial research was successfully completed before the deadline for the Ethics submission deadline and on October 9, 2023, the project received approval from both Dr. Mark Johnston and Dr. Alireza Daneshkhah.

Originally, the plan detailed the execution of specific project components on a monthly basis. The task was to write the analysis part in October and the forecast part in November. The final conclusions would be written in December. However, by the third week, it became apparent that certain aspects required more time than initially anticipated. In response, I adopted a more intensive approach, committing to providing weekly drafts to my supervisor. Every Thursday at 4:00 pm, a scheduled meeting took place between Dr. Mark Johnston and myself. Dr. Mark Johnston provided feedback and discussed the next steps of the project. This level of accountability, which served as both pressure and motivation enabled me to push myself and adhere to the revised schedule. A new timeline, in the format of a Gantt chart, which can be found in Appendix B, was created in early October and shows the duration each part required. A visible pattern is evident after review, as a significant portion of the project components experienced delays, compared to the initial plan. Nonetheless, the final draft was successfully completed on Thursday, 07 December 2023.

Throughout the project, various challenges were encountered. The initial data selection process proved to be time-consuming due to the abundance of datasets related to the Chinese economy. After careful evaluation, annual data up until 2022 were selected. Additionally, difficulties arose during the model development phase, as ARIMA models were more challenging than expected. For this reason, two weeks were invested in order to gain proficiency in ARIMA modeling, allowing for the successful continuation of the project.

In conclusion, there were many ups and down in the writing of the project, largely stemming from unforeseen time constraints. Despite these obstacles, the project was ultimately conducted successfully and adhered to the designated timeline.

6 Social, Legal, Ethical and Professional Considerations

Since economic data regarding China is the most crucial part of this report, the critical examination of their validity becomes a paramount concern. The majority of this data originates from public sources and particularly with respect to China, there exists a possibility that these values were provided by official Chinese organizations.

An ethical dilemma arises concerning the credibility of this data, as numerous researchers have raised concerns about its accuracy and potential exaggeration. Some researchers, such as Tyrrell (2020), assert that Chinese data cannot be unequivocally trusted. China has been criticized for presenting data that appears exceptionally smooth compared to more authentic datasets, leading to questions regarding their reliability. A good example is China ‘s nominal gross domestic product growth, which has been overestimated by 2 percentage points for nine years from 2008 to 2016 (Leng, 2019).

Further evidence supporting scepticism towards the validity of Chinese economic data is found in an admission by Chinese officials in 2015, as reported by Global Times (2015). The smooth out of data was, reportedly, attributed to repetitive accounting, lack of unified sources and incomplete information. A deeper reason for this phenomenon can be found in the prevailing central system in China, where local officials may receive promotions based on favourable statistics related to GDP.

Furthermore, it is important to note that this report included GDP data for China dating back to the year 1500. It is vital to acknowledge that these historical data points are speculative assumptions, considering that economic metrics such as GDP were not conceived during that period. The values presented have been retroactively calculated based on backdated economic models and present-day data, introducing an element of estimation and reconstruction into the historical economic narrative.

However, as these datasets are posted in official databases, such as World Bank, Statista, macrotrends and IMF, known for their commitment to accuracy and reliability, it is reasonable to consider them as credible and assumed to be true unless proven otherwise.

7 Conclusions

7.1 Achievements

Considering the main questions and aims of this report, it can be concluded that the project was completed successfully. The initial research questions, which can be found in the Introduction, are the following:

Evolution of the Chinese Economy:

1) How has the Chinese economy evolved over time? How have the macroeconomic trends evolved?

Long-Term Economic Trends and Correlations:

2) Are there any long-term economic trends in China and do they correlate with significant global events? Are there any trends that cannot be explained? (In case of unexplained by known events economic disruptions, this project will delve deeper into the political and social context of those periods).

Recurring Patterns and Cycles:

3) Are there any recurring patterns or cycles in the Chinese economy, and if so, what insights do they offer into economic dynamics?

Predictive Capabilities of Economic Analysis:

4) Can time series forecasting through the ARIMA model, trade analysis and advanced modeling techniques contribute to predicting future economic trends, trade relationships and financial market disruptions in China? What are their predictions and can they be trusted?

The report analysed the trajectory of the Chinese economy at length revealing a consistent positive trend in its key economic metrics. Moreover, significant global events have consistently showed that they always had consequences on the Chinese economy with exports emerging as a particularly influenced aspect, reliable on the demands of other nations. Additionally, the analysis showed a number of recurring patterns and cycles, as shifts in world events or disruptions showed to influence China’s metrics. Nonetheless, the events that seemed to influence China more, are those regarding its fiscal policy and domestic occurrences (e.g. economic reforms). Lastly, the ARIMA model appeared to be a sufficient tool in predicting future values for all five key metrics but inflation. GDP was successfully predicted with 5% error, Inflation showed an unexpected 40.86% error, imports were predicted with 22.29% error, exports with 21.84% error and FDI inflows with 20.13% error. Following an 0.88 data split, the ARIMA models underwent training on the available data and generated 10-year forecasts that demonstrate a credible and plausible outlook. The developed models show similarities with the observed in the literature review in section 3.1, as the also contain low value parameters. The utilization of the ARIMA model enhances the predictive capabilities of the analysis, providing valuable insights into the anticipated trajectory of China's economic metrics in the coming years.

7.2 Future Work

Despite the fact that the report was completed successfully, certain ideas that were considered in the beginning could not be pursued due to limited time. In order to produce a complete economic analysis of the Chinese economy, more metrics could have been taken into account. In the future, I would like to explore the interconnections between government expenditure within specific sectors, such as manufacturing and trade, and its potential impact on the overall economy. Additionally, I believe that a comprehensive analysis of the Chinese economy spanning distinct historical periods, including the Mao era, Republic of China era, and the

People’s Republic of China era would be beneficial. The investigation would aim to identify any economic reforms, analogous to those implemented in 1978, that were introduced earlier, potentially experiencing interruptions and subsequent resumptions. That data could be beneficial for a comparative analysis, facilitating the identification of any similar trends between the two eras.

In addition, I believe that more comparisons between China’s economy with other similar sized economies/countries, such as the United States and India, would be interesting. By analysing import and export volumes, the comparative strengths and weaknesses of each economy in the global market could be identified.

Furthermore, China is a country with a big balance of payments surplus, meaning that China has loaned money to many nations worldwide. It would be interesting to analyse these data, look for any patterns and see what impact this had in China’s economic evolution.

Moreover, after gaining an understanding of the ARIMA models and how their parameters work, I would enjoy implementing the same models into alternative dataset splits (e.g. a 0.50 split) and investigating the projected forecasts. Lastly, as the models gave 5 10-year forecasts, I aspire to conduct an evaluation of their projections a decade hence, inspecting the accuracy of their predictions.

8 Bibliography and References

AJMC Staff. (2021, January 2). *A Timeline of COVID-19 Developments in 2020*. The American Journal of Managed Care. https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020

ASEAN Secretariat. (2020). *ASEAN Member States*. Asean.org. https://asean.org/member-states/

Bettelheim C. (1988) Economic reform in China, The Journal of Development Studies, 24:4, 15-49, DOI: 10.1080/00220388808422081

Bevans, R. (2020). *Akaike information criterion | when & how to use it*. Scribbr. https://www.scribbr.com/statistics/akaike-information-criterion/#:~:text=The%20Akaike%20information%20criterion%20(AIC

Britannica Dictionary. (2023a). *Trade deficit Definition & Meaning | Britannica Dictionary*. Www.britannica.com. https://www.britannica.com/dictionary/trade-deficit

Britannica Dictionary. (2023b). *Trade surplus Definition & Meaning | Britannica Dictionary*. Www.britannica.com. https://www.britannica.com/dictionary/trade-surplus

Britannica, T. Editors of Encyclopaedia. (2022). Self-Strengthening Movement | Chinese history. In *Encyclopædia Britannica*. https://www.britannica.com/event/Self-Strengthening-Movement

Britannica, T. Editors of Encyclopaedia (2023c). Cold War. Encyclopedia Britannica. https://www.britannica.com/event/Cold-War

Broadman, H. G., & Sun, X. (1997). The Distribution of Foreign Direct Investment in China. *The World Economy*, *20*(3), 339–361. https://doi.org/10.1111/1467-9701.00073

Brownlee J. (2017). *How to Create an ARIMA Model for Time Series Forecasting in Python*. Machine Learning Mastery. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/

Bruce-Lockhart, A. (2017, June 26). *China's $900 billion New Silk Road. What you need to know*. World Economic Forum. https://www.weforum.org/agenda/2017/06/china-new-silk-road-explainer/

Burdekin, R. (2000). *A Service of Leibniz Information Centre for Economics*. https://www.econstor.eu/bitstream/10419/94624/1/2000-08.pdf

Cambridge Dictionary. (2023). *Hyperinflation*. @CambridgeWords. https://dictionary.cambridge.org/dictionary/english/hyperinflation

Chen, C. (2011), Foreign Direct Investment in China: Location determinants, investor differences and economic impacts, Cheltenham, UK: Edward Elgar. doi.org/ 10.4337/9781781001141.

Chen, C. (2018). *The liberalisation of FDI policies and the impacts of FDI on China's economic development*. https://press-files.anu.edu.au/downloads/press/n4267/pdf/ch29.pdf

Chen, C.-S., & Lieberthal, K. G. (2019). China - Trade. In The Editors of Encyclopaedia Britannica (Ed.), *Encyclopædia Britannica*. https://www.britannica.com/place/China/Trade

Chien, K. (2008). China okays $586 billion stimulus. *Reuters*. https://www.reuters.com/article/uk-china-economy-idUKTRE4A817E20081109

China Power Team. (2020). "Unpacking China's GDP". *Unpacking China's GDP*. ChinaPower Project. https://chinapower.csis.org/tracker/china-gdp/

Cordesman, A. H. (2023). China's Emergence as a Superpower. *Www.csis.org*. https://www.csis.org/analysis/chinas-emergence-superpower

Deepanshu Sharma and Kritika Phulli. Box-Jenkins ARIMA Modelling: Forecasting FDI in India. *ScienceOpen Posters*. 2020. DOI: 10.14293/S2199-1006.1.SOR-.PPZCKJY.v1

Deng, K., & Shengmin, S. (2019). China's Extraordinary Population Expansion and Its Determinants during the Qing Period, 1644-1911. *Population Review*, *58*(1). https://doi.org/10.1353/prv.2019.0001

Dritsaki, C. (2015). Forecasting Real GDP Rate Through Econometric Models: An Empirical Study from Greece. Journal of International Business and Economics, 3, 13-19. https://doi.org/10.15640/jibe.v3n1a2

Eissa, N. (2020). Forecasting the GDP per Capita for Egypt and Saudi Arabia Using ARIMA Models. *Research in World Economy*, *11*(1), 247. https://www.sciedu.ca/journal/index.php/rwe/article/view/17182/10875

Franke, H., Suzuki, C., & Knightley, D. N. (2023). *China | Culture, History, Maps, & People*. Encyclopedia Britannica. https://www.britannica.com/place/China#ref70957

Global Times . (2015). *Inflated statistics wreak havoc on economies of Northeast China - Global Times*. Www.globaltimes.cn. https://www.globaltimes.cn/content/958306.shtml

Harrison, V., & Palumbo, D. (2019). China anniversary: How the country became the world's "economic miracle." *BBC News*. https://www.bbc.co.uk/news/business-49806247

Hong, G. H., Lee, J., Liao, W., & Seneviratne, D. (2016). China and Asia in Global Trade Slowdown. *IMF Working Papers*, *16*(105), 1. https://doi.org/10.5089/9781484368565.001

Hook, G. D. (2022). *China's Silk Roads* (L. R. Kurtz, Ed.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/pii/B9780128201954001175

Hsiao, K. H. (1971). *Money and Monetary Policy in Communist China*. Columbia University Press: New York. https://www.cambridge.org/core/journals/journal-of-asian-studies/article/abs/money-and-monetary-policy-in-communist-china-by-katharine-huang-hsiao-new-york-columbia-university-press-1971-308-pp-1250/7C7E500F1EEC9AA09A59EDDFE3FACAE7

Hu, X. (2016). Legacy of Tiananmen Square Incident in Sino-US Relations (post-2000). *East Asia*, *33*(3), 213–232. https://doi.org/10.1007/s12140-016-9259-5

Huld, A. (2023, January 20). *FDI in China Sees Slow but Steady Growth in 2022*. China Briefing News. https://www.china-briefing.com/news/china-records-steady-fdi-growth-in-2022

Information Office of the Ministry of Commerce (2023). *The country's use of foreign capital will grow steadily in 2022*. Www.mofcom.gov.cn. http://www.mofcom.gov.cn/article/syxwfb/202301/20230103379768.shtml

International Monetary Fund. (1998,). *The Asian Crisis: Causes and Cures*. Finance and Development | F&D. https://www.imf.org/external/pubs/ft/fandd/1998/06/imfstaff.htm

Investopedia. (2022). *What Is Per Capita GDP?* Investopedia. https://www.investopedia.com/terms/p/per-capita-gdp

ISHIHARA, K. (1990). INFLATION AND ECONOMIC REFORM IN CHINA. *The Developing Economies*, *28*(2), 180–201. https://doi.org/10.1111/j.1746-1049.1990.tb00180.x

Jahn, M. (2021, July 21). *What country is the world's largest exporter of goods?* Investopedia. https://www.investopedia.com/ask/answers/011915/what-country-worlds-largest-exporter-goods.asp

Jahn, M. (2023, August 30). *What country is the world's largest exporter of goods?* Investopedia. https://www.investopedia.com/ask/answers/011915/what-country-worlds-largest-exporter-goods.asp

Jalil, A., & Rao, N. H. (2019). *Chapter 8 - Time Series Analysis (Stationarity, Cointegration, and Causality)* (B. Özcan & I. Öztürk, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/pii/B9780128167977000084

Kenton, W. (2019). *Understanding the Spillover Effect*. Investopedia. https://www.investopedia.com/terms/s/spillover-effect.asp

Kobayashi, S., Baobo, J., & Sano, J. (1999). *The "Three Reforms" in China: Progress and Outlook*. Japan Research Institute. https://www.jri.co.jp/english/periodical/rim/1999/RIMe199904threereforms/

Leng, S. (2019). *China "exaggerated" GDP data by 2 percentage points for at least nine years, new study says*. South China Morning Post. https://www.scmp.com/economy/china-economy/article/2189052/china-exaggerated-gdp-data-2-percentage-points-least-nine

Li, L., Willett, T. D., & Zhang, N. (2012). The Effects of the Global Financial Crisis on China's Financial Market and Macroeconomy. *Economics Research International*, *2012*(1), 1–6. https://doi.org/10.1155/2012/961694

Logubayom, I. A., Nasiru, S., & Luguterah, A. (2013). Modelling the Rate of Treasury Bills in Ghana. *Mathematical Theory and Modeling*, *3*(11), 102. https://www.iiste.org/Journals/index.php/MTM/article/view/7689

Macrotrends. (2023). *Macrotrends | the Long Term Perspective on Markets*. Macrotrends.net. https://www.macrotrends.net/

Maddison, A. (2006). *China in the world economy: 1300-2030*. Go.gale.com. https://go.gale.com/ps/i.do?id=GALE%7CA168399571&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=10834346&p=AONE&sw=w&userGroupName=anon%7Ef93554d6&aty=open-web-entry

Maity, B., & Chatterjee, B. (2012). FORECASTING GDP GROWTH RATES OF INDIA: An Empirical Study. *International Journal of Economics and Management Sciences, 1*(9), 52–58. https://www.hilarispublisher.com/open-access/forecasting-gdp-growth-rates-of-india-an-empirical-study-2162-6359-1-082.pdf

McKinsey Global Institute. (2019). *China and the world Inside the dynamics of a changing relationship*. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/China/China%20and%20the%20world%20Inside%20the%20dynamics%20of%20a%20changing%20relationship/MGI-China-and-the-world-Full-report-June-2019-vF.ashx

Miami University. (n.d.). *Silk Road ExplorAsian*. Miami University. Retrieved November 7, 2023, from https://miamioh.edu/cas/centers-institutes/havighurst-center/additional-resources/havighurst-special-programming/silk-road/index.html

Morrison, W. M. (2019, June 25). *China's Economic Rise: History, Trends, Challenges, and Implications for the United States*. Everycrsreport.com; Congressional Research Service. https://www.everycrsreport.com/reports/RL33534.html

National Bureau of Statistics of China. (October 31, 2022). Actually utilized foreign direct investment (FDI) in China in 2021, by leading country or region (in billion U.S. dollars) [Graph]. In Statista. Retrieved November 15, 2023, from https://www.statista.com/statistics/257099/value-of-foreign-direct-investment-in-china-by-country/

Ning, W., Kuan-jiang, B., & Zhi-fa, Y. (2010). Analysis and Forecast of Shaanxi GDP Based on the ARIMA Model. *Asian Agricultural Research*, *02*(01), 1–4. https://ideas.repec.org/a/ags/asagre/93238.html

OECD. (2023a). *Conversion rates - Purchasing power parities (PPP) - OECD Data*. TheOECD. https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm

OECD. (2023b). *GDP and spending - Nominal GDP forecast - OECD Data*. TheOECD. https://data.oecd.org/gdp/nominal-gdp-forecast.htm

Office of the Historian. (2019). *Milestones: 1899–1913 - Office of the Historian*. State.gov. https://history.state.gov/milestones/1899-1913/chinese-rev

Overholt, W. H. (2016). China and the evolution of the world economy. *China Economic Review*, *40,* 267–271. https://doi.org/10.1016/j.chieco.2016.07.005

Park, D., Ramayandi, A., & Shin, K. (n.d.). *Why Did Asian Countries Fare Better during the Global Financial Crisis than during the Asian Financial Crisis?* https://www.piie.com/publications/chapters_preview/6741/04iie6741.pdf

Perkins, D.H. (2013). Agricultural Development in China, 1368-1968 (1st ed.). Routledge. https://doi.org/10.4324/9781315082776

Perry, E. J., & Wong, C. (1985). The Political Economy of Reform in Post-Mao China. In *JSTOR* (1st ed., Vol. 2). Harvard University Asia Center. https://www.jstor.org/sTable/j.ctt1sq5tbp

Phillips, T. (2016, May 11). *The Cultural Revolution: all you need to know about China's political convulsion*. The Guardian; Guardian Media Group. https://www.theguardian.com/world/2016/may/11/the-cultural-revolution-50-years-on-all-you-need-to-know-about-chinas-political-convulsion

Qu Haiqing, He Xianping (2021). Research on Hubei GDP Forecast Model Based on Time Series Analysis. Journal of Hubei University of Economics (Humanities and Social Sciences Edition), 2021,18 (09): 37-39.

Santander. (2023, October). *Foreign investment in China - Santandertrade.com*. Santandertrade.com. https://santandertrade.com/en/portal/establish-overseas/china/foreign-investment

Shweta. (2021, July 30). *Introduction to Time Series Forecasting — Part 2 (ARIMA Models)*. Medium. https://towardsdatascience.com/introduction-to-time-series-forecasting-part-2-arima-models-9f47bf0f476b

The Editors of Encyclopedia Britannica. (2019). Tiananmen Square incident | Chinese history [1989]. In *Encyclopædia Britannica*. https://www.britannica.com/event/Tiananmen-Square-incident

The Language Nerds Team. (2019, August 23). *Tea if by sea, Cha if by land: Why the world has only two words for tea. | THE LANGUAGE NERDS*. The Language Nerds. https://thelanguagenerds.com/2019/tea-if-by-sea-cha-if-by-land-why-the-world-had-only-two-words-for-tea/

TUCKER, A. M. H. (2023). *The Global 2000 2023*. Forbes. https://www.forbes.com/lists/global2000/?sh=41b093fb5ac0

Tyrrell, P. (2020, November 8). *The Problem of False Chinese Economic Data*. The Heritage Foundation. https://www.heritage.org/international-economies/commentary/the-problem-false-chinese-economic-data

UNESCO, & Zhaowen, G. (n.d.). *About the Silk Roads | SILK ROADS*. En.unesco.org; UNESCO. https://en.unesco.org/silkroad/about-silk-roads

US bank. (2023, September 18). *Analysis: China's Economy and Its Influence on Global Markets | U.S. Bank*. Www.usbank.com. https://www.usbank.com/investing/financial-perspectives/market-news/chinas-economic-influence.html

Walter, C. (1985). Dual Leadership and the 1956 Credit Reforms of the People's Bank of China. *The China Quarterly*, 102, 277 - 290. https://doi.org/10.1017/S0305741000029945.

Weinberg, J. (2013). *The Great Recession and its Aftermath*. Federal Reserve History. https://www.federalreservehistory.org/essays/great-recession-and-its-aftermath

Workman, D. (2023a). *China's Economic Rise: History, Trends, Challenges, and Implications for the United States*. Www.everycrsreport.com. https://www.everycrsreport.com/reports/RL33534.html#_Toc12530873

Workman, D. (2023b). *China's Top Trading Partners 2022*. Www.worldstopexports.com. https://www.worldstopexports.com/chinas-top-import-partners/?expand_article=1

World Bank. (2023). *The World Bank In China*. World Bank. https://www.worldbank.org/en/country/china/overview

World Integrated Trade Solution. (2018). *TRADE SUMMARY FOR CHINA*. Worldbank.org. https://wits.worldbank.org/countrysnapshot/en/CHN/textview

Worldometers.info. (2023). *China Population (2020) - Worldometer*. Www.worldometers.info. https://www.worldometers.info/world-population/china-population

Yao, K., & Zhang, E. (2023, January 17). China's 2022 economic growth one of the worst on record, post-pandemic policy faces test. *Reuters*.

https://www.reuters.com/world/china/chinas-economy-slows-sharply-q4-2022-growth-one-worst-record-2023-01-17/

Zhang, C., & Clovis, J. (2010). China inflation dynamics: Persistence and policy regimes. *Journal of Policy Modeling*, *32*(3), 373–388. https://doi.org/10.1016/j.jpolmod.2010.02.007

Zhou Mingwei. (2014). *Promoting China-Japan relations through culture - China.org.cn*. China.org.cn. http://www.china.org.cn/opinion/2014-06/18/content_32690843.htm

Appendix A – Timeline of all key events

Gross Domestic Product

Date	*Key event*
1368- 1644	Ming Dynasty
1644-1912	Qing Dynasty
1700-1830	Population growth
1840-1860	Opium Wars
1851-1864	Taiping Rebellion
1861-1895	Self-Strengthening Movement
1912	Republic of China
1919	The May forth movement
1949	Declaration of People’s Republic of China
1958-1962	The Great Leap Forward
1966-1976	The Cultural Revolution
1978	Chinese economic reforms
1979	Reform on People’s Bank of China
1989	The Tiananmen Square incident
2007-2008	Financial Crisis

Inflation

Date	*Key event*
130 BC -1453	Silk roads
1958-1962	The Great Leap Forward
1978	Chinese economic reforms
1988	Announcement of Economic reforms being implemented by all means
1992	South China Tour Speech
2007-2008	Financial crisis

Trade

Date	*Key event*
1960	Sino-Soviet breach
1978	Chinese economic reforms
2007-2008	Financial crisis

Foreign Direct investment

Date	*Key event*
1978	Chinese economic reforms
1979	Law on Joint Ventures
1992	nationwide implementation of open policies for FDI
1997-1998	East Asian Financial Crisis
2001	China's entry into the World Trade Organization (WTO)
2006	Regulations for Merger and Acquisition (M&A) of Domestic Enterprises by Foreign Investors
2007	Anti-Monopoly Law
2007	Enterprise Income Tax Law
2007-2008	Financial crisis
2022	Ukraine-Russia war

Appendix B – Project Management Records

Project Management

Task Name	Start	End	Delay	Status	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11	W 12
					22	29	6	13	20	27	3	10	17	24	1	8
					Fri	Fri	Fri	Fri	Fri	Fri	Fri	Fri	Fri	Fri	Fri	Fri
Topic Finding			Yes	Completed												
Research Topic Area	19-Jul-23	22-Sep-23	No	Completed												
Topic Analysis & Selection	22-Sep-23	28-Sep-23	No	Completed												
Proposal			100%	Completed												
Research Questions & Aims	22-Sep-23	28-Sep-23	No	Completed												
Detailed Planning of Stages	28-Sep-23	02-Oct-23	Yes	Completed												
Ethics Application	01-Oct-23	03-Oct-23	No	Completed												
Resubmition and Approval	07-Oct-23	09-Oct-23	No	Completed												
Material Selection				Completed												
Material Selection	07-Oct-23	19-Oct-23	Yes	Completed												
Evaluate Sources	15-Oct-23	19-Oct-23	Yes	Completed												
Data Selection	07-Oct-23	19-Oct-23	Yes	Completed												
Literature Review	15-Oct-23	19-Oct-23	No	Completed												
Economic Analysis				Completed												
China's History	07-Oct-23	17-Oct-23	Yes	Completed												
GDP Analysis	07-Oct-23	01-Nov-23	Yes	Completed												
Inflation Analysis	27-Oct-23	01-Nov-23	Yes	Completed												
Imports Analysis	03-Nov-23	08-Nov-23	Yes	Completed												
Exports Analysis	03-Nov-23	08-Nov-23	Yes	Completed												
FDI inflows Analysis	09-Oct-23	16-Nov-23	Yes	Completed												
Final Comparison	30-Nov-23	05-Nov-23	No	Completed												
Forecast			100%	Completed												
Methodology	06-Nov-23	20-Nov-23	Yes	Completed												
Forecasts for all metrics	22-Nov-23	06-Dec-23	No	Completed												
Final Comparison	22-Nov-23	06-Dec-23	Yes	Completed												
Conclusions			100%	Completed												
Future works	02-Dec-23	06-Dec-23	No	Completed												
Project Management	02-Dec-23	06-Dec-23	No	Completed												
Student Reflection	02-Dec-23	06-Dec-23	No	Completed												
Social, Legal, Ethical and Profe	02-Dec-23	06-Dec-23	No	Completed												
Final Revision	06-Dec-23	07-Dec-23	No	Completed												

Appendix C – Programming Code

Imports

```
from pandas import read_csv
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.dates as mdates
from pandas import DataFrame
from pandas import Grouper
from matplotlib import pyplot
from pandas import datetime
from pandas import concat
from sklearn.metrics import mean_squared_error
from math import sqrt
from pandas import Series
import warnings
warnings.filterwarnings("ignore")
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
from statsmodels.tsa.statespace.sarimax import SARIMAX
from pmdarima import auto_arima
from pmdarima.arima import ARIMA
from statsmodels.tsa.arima.model import ARIMA
```

Figure 1: Nominal GDP and GDP at PPP for China (2022)

```
ppp= pd.read_csv('nominal-ppp.csv')
ppp_sorted_nominal = ppp.sort_values('Nominal (current US$)', ascending=False)
ppp_sorted_ppp = ppp.sort_values('PPP (current international $)',
ascending=False)
fig, axs = plt.subplots(1, 2, figsize=(12, 6))
sns.set(style="dark")
ax1 = sns.barplot(x='Nominal (current US$)', y='Country Name',
data=ppp_sorted_nominal, palette=['skyblue' if country != 'China' else 'red'
for country in ppp_sorted_nominal['Country Name']], ax=axs[0])
axs[0].set_xlabel('Current US$')
axs[0].set_ylabel('Country')
for index, value in enumerate(ppp_sorted_nominal['Nominal (current US$)']):
    ax1.text(value, index, str(round(value, 2)), va='center', ha='left')
ax2 = sns.barplot(x='PPP (current international $)', y='Country Name',
data=ppp_sorted_ppp, palette=['skyblue' if country != 'China' else 'red' for
country in ppp_sorted_ppp['Country Name']], ax=axs[1])
axs[1].set_xlabel('Current international $')
axs[1].set_ylabel('Country')
for index, value in enumerate(ppp_sorted_ppp['PPP (current international
$)']):
    ax2.text(value, index, str(round(value, 2)), va='center', ha='left')
plt.tight_layout()
plt.show()
```

Figure 51: Share of Global GDP at PPP (1500-2020)

```
data=pd.read_csv("gdp1500.csv")
columns =['China', 'USA', 'India', 'Japan', 'Germany', 'Russia/USSR', 'UK',
'France', 'Rest of World']
colours =['#DC143C', '#BFEFFF', '#FF7D40', '#03A89E', '#9467bd', '#8c564b',
'#e377c2', '#7f7f7f', '#87CEFF']
plt.figure(figsize=(10, 6))
plt.stackplot(data['X.1'], *[data[column] for column in columns],
labels=columns, colors=colours)
plt.xlabel('Year')
plt.ylabel('Stacked GDP  (%)')
plt.title('Share of Global GDP since 1500' )
plt.axvline(x=1644, color='black', linestyle='--',linewidth=4)
plt.axvline(x=1912, color='black', linestyle='--',linewidth=4)
plt.text(1500,20,  'The Ming Dynasty', rotation=0, fontsize=16)
plt.text(1650,20,  'The Qing Dynasty', rotation=0, fontsize=16)
plt.axvline(x=1840, color='blue', linestyle='--' )
plt.text(1840,80,  'Opium Wars', rotation=0, fontsize=14, color='blue')
plt.axvline(x=1860,color='blue', linestyle='--')
plt.axvline(x=1851, color='green', linestyle='--')
plt.text(1852,70,  'Taiping Rebelion', rotation=0, fontsize=14, color='green')
plt.axvline(x=1864, color='green', linestyle='--')
plt.axvline(x=1861, color='yellow', linestyle='--')
plt.text(1863,90,  'Self Strength. Movem.', rotation=0, fontsize=14,
color='yellow')
plt.axvline(x=1895,color='yellow', linestyle='--' )
plt.axvline(x=1949, color='brown', linestyle='--' )
plt.text(1950,30,  'Mao Zedong', rotation=90, fontsize=14, color='brown')
plt.text(1980,30,  'Reforms', rotation=90, fontsize=14, color='black')
plt.axvline(x=1976, color='black', linestyle='--')
plt.axvline(x=1939, color='black', linestyle='--')
plt.text(1940,10,  'WW2',rotation=90, fontsize=14, color='black')
plt.axvline(x=1912, color='black', linestyle='--')
plt.text(1915,20,  'Creation of PRC', rotation=90, fontsize=14, color='black')
plt.text(1915,4,  'WW1', rotation=90,fontsize=14, color='black')
plt.axvspan(1861, 1895, color='yellow', alpha=0.3)
plt.axvspan(1851, 1864, color='green', alpha=0.5)
plt.axvspan(1840,1860, color='blue', alpha=0.3)
plt.legend(loc='upper left')
plt.show()plt.axvline(x=1976, color='black', linestyle='--' )
plt.axvline(x=1939, color='black', linestyle='--')
plt.text(1940,10,  'WW2', rotation=90, fontsize=14, color='black')
plt.axvline(x=1912, color='black', linestyle='--')
plt.text(1915,20,  'Creation of PRC', rotation=90, fontsize=14, color='black')
plt.text(1915,4,'WW1', rotation=90, fontsize=14, color='black')
plt.axvspan(1861, 1895, color='yellow', alpha=0.3)
plt.axvspan(1851, 1864, color='green', alpha=0.5)
plt.axvspan(1840, 1860, color='blue', alpha=0.3)
plt.legend(loc='upper left')
plt.show()
```

Figure 3: Chinese GDP (Billions of US $) (1978-2022)

```
gdp= pd.read_csv('china_gdp_before.csv',skiprows=16)
gdp = gdp[(gdp['date'] >= 1978)]
plt.figure(figsize=(12, 8))
plt.plot(gdp['date'], gdp[' GDP ( Billions of US $)'])
plt.xlabel('date')
plt.ylabel('GDP (Billions of US $)')
plt.xticks(gdp['date'], rotation=90)
plt.axvline(x=2007, color='green', linestyle=':')
plt.text(2007, 7500, 'Financial Crisis', rotation=90, fontsize=14,
color='green')
plt.axvline(x=1989, color='green', linestyle=':')
plt.text(1989, 6000, 'Tiananmen Square', rotation=90, fontsize=14,
color='green')
plt.grid( linewidth=0.1)
plt.show()
```

Figure 4: Annual % Change of GDP of China (1978-2022)

```
gdp= pd.read_csv('china_gdp_before.csv',skiprows=16)
gdp = gdp[(gdp['date'] >= 1978)]
years = gdp['date']
annual_change = gdp[' Annual % Change']
plt.figure(figsize=(12, 8))
bars= plt.bar(years, annual_change, color='skyblue')
plt.xlabel('Years')
plt.ylabel('Annual % Change')
plt.xticks(gdp['date'], rotation=90)
plt.grid(axis='y')
plt.text(1988, 11.7226, '11.22%', ha='center', fontsize=12, color='black')
plt.text(1989, 5.2063, '4.20%', ha='center', fontsize=12, color='black')
plt.text(1990.5, 4.5203, '3.92%', ha='center', fontsize=12, color='black')
plt.text(1992, 14.2245, '14.22%', ha='center', fontsize=12, color='black')
plt.text(2007, 14.2309, '14.23%', ha='center', fontsize=12, color='black')
plt.text(2009, 9.3987, '9.39%', ha='center', fontsize=12, color='black')
plt.text(2010, 10.6359, '10.63%', ha='center', fontsize=12, color='black')
plt.axvline(x=2007, color='green', linestyle=':', linewidth=2)
plt.text(2007, 15, 'Financial Crisis', fontsize=14, color='green',
fontweight='bold')
plt.axvline(x=1989, color='green', linestyle=':', linewidth=2)
plt.text(1989, 15, 'Tiananmen Square', fontsize=14, color='green',
fontweight='bold')
plt.text(2019, 5.9505, '5.95%', ha='center', fontsize=12, color='black')
plt.text(2020, 2.2386, '2.23%', ha='center', fontsize=12, color='black')
plt.text(2022, 2.9908, '2.99%', ha='center', fontsize=12, color='black')
bars[10].set_color('red')
bars[11].set_color('red')
bars[12].set_color('red')
bars[14].set_color('red')
bars[29].set_color('red')
bars[32].set_color('red')
bars[31].set_color('red')
bars[31].set_color('red')
bars[41].set_color('red')
bars[42].set_color('red')
bars[44].set_color('red')
plt.show()
```

Figure 5: Inflation Rate (%) of China (1987-2022)

```
inflation= pd.read_csv('china_inflation_new.csv',skiprows=16)
sns.set(style="whitegrid")
custom_colors = sns.color_palette("Set2")
plt.figure(figsize=(12, 8))
plt.fill_between(inflation['year'], inflation[' Inflation Rate (%)'],
color=custom_colors[0], alpha=0.4)
plt.xlabel('Years', fontsize=12)
plt.xticks(inflation['year'], rotation=90)
plt.ylabel('Inflation Rate (%)', fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.text(1988, 18.8118, '18.81%', fontsize=12, color='black')
plt.text(1990, 3.0523, '3.05%', fontsize=12, color='black')
plt.text(1994, 24.257, '24.25%', fontsize=12, color='black')
plt.text(1999, -1.4015, '-1.40%', fontsize=12, color='black')
plt.text(2002, -0.732, '-0.73%', fontsize=12, color='black')
plt.text(2008, 5.9253, '5.92%', fontsize=12, color='black')
plt.text(2009, -0.7282, '-0.72%', fontsize=12, color='black')
plt.text(2011, 2.9908, '2.99%', fontsize=12, color='black')
plt.text(2019, 2.8992, '2.89%', fontsize=12, color='black')
plt.text(2021, 0.981, '0.98%', fontsize=12, color='black')
plt.show()
```

Figure 7: Trade values of China (1976-2022)

```
imports = pd.read_csv('china_imports.csv',skiprows=16)
exports = pd.read_csv('china_exports.csv',skiprows=16)
plt.figure(figsize=(12, 8))
plt.plot(imports['year'], imports[' Billions of US $'], label='Imports')
plt.plot(exports['year'], exports[' Billions of US $'], label='Exports')
plt.axvline(x=2009, color='green', linestyle=':')
plt.text(2009, 84, 'Financial crisis', rotation=0, fontsize=16, color='green')
plt.text(2020, 700, 'Covid-19 pandemic', rotation=90, fontsize=14,
color='black')
plt.axvspan(2019, 2021, color='green', alpha=0.3)
plt.xlabel('Years')
plt.ylabel('Amount (in Billions of US $)')
plt.xticks(imports['year'], rotation=90)
plt.grid(True)
plt.legend()
plt.show()
```

Figure 8: Change of % in China's trade values (1976-2022)

```
imports = pd.read_csv('china_imports.csv',skiprows=16)
exports = pd.read_csv('china_exports.csv',skiprows=16)
plt.figure(figsize=(12, 8))
plt.plot(imports['year'], imports[' Billions of US $'], label='Imports')
plt.plot(exports['year'], exports[' Billions of US $'], label='Exports')
plt.axvline(x=2009, color='green', linestyle=':')
plt.text(2009, 84, 'Financial crisis', rotation=0, fontsize=16, color='green')
plt.text(2020, 700, 'Covid-19 pandemic', rotation=90, fontsize=14,
color='black')
plt.axvspan(2019, 2021, color='green', alpha=0.3)
plt.xlabel('Years')
plt.ylabel('Amount (in Billions of US $)')
plt.xticks(imports['year'], rotation=90)
```

```
plt.grid()
plt.legend()
plt.show()
```

Figure 9: China's Exports by Country in 2022 (in billions of US$ and %)

```
countries = ['United States', 'Hong Kong', 'Japan', 'South Korea', 'Vietnam',
'India', 'Netherlands',
             'Germany', 'Malaysia', 'Taiwan', 'United Kingdom', 'Singapore',
'Australia', 'Thailand', 'Mexico']
exports = [582.8, 297.5, 172.9, 162.6, 147, 118.5, 117.7, 116.2, 93.7, 81.6,
81.5, 81.2, 78.8, 78.5, 77.5]
percentages = [16.2, 8.3, 4.8, 4.5, 4.1, 3.3, 3.3, 3.2, 2.6, 2.3, 2.3, 2.3,
2.2, 2.2, 2.2]
fig, ax1 = plt.subplots(figsize=(12, 8))
ax2 = ax1.twinx()
ax1.bar(countries, exports, color='b')
ax2.plot(countries, percentages, 'r', linewidth= 6, alpha= 0.7)
ax1.set_xlabel('Countries')
ax1.set_ylabel('Exports in billion USD', color='b')
ax2.set_ylabel('Percentage of China's total exports', color='r')
plt.setp(ax1.get_xticklabels(), rotation=45, ha="right")
for i, v in enumerate(exports_billion):
    ax1.text( i, v + 10, str(v), color='black', ha='center', fontsize= 14)
for i, v in enumerate(percentages):
    ax2.text(i, v - 0.5, str(v)+ '%', color='red', ha='center', fontsize= 16)
plt.grid(False)
plt.show()
```

Figure 10: Countries Generating China's Biggest Trading Surpluses (2022)

```
countries = ['United States', 'Hong Kong', 'Netherlands', 'India', 'Mexico',
'United Kingdom', 'Vietnam', 'Singapore', 'Philippines', 'Poland']
surpluses = [403.8, 289.7, 105.2, 101.0, 60.1, 59.7, 59.0, 47.2, 41.6, 33.1]
sorted = sorted(range(len(surpluses)), key=lambda k: surpluses[k],
reverse=True)
sorted_countries = [countries[i] for i in sorted]
sorted_surpluses = [surpluses[i] for i in sorted]
plt.figure(figsize=(12, 8))
plt.barh(sorted_countries, sorted_surpluses, color='skyblue')
plt.xlabel('Trading Surpluses (in billion USD)')
plt.ylabel('Countries')
plt.grid(axis='x', linestyle='--', alpha=0.6)
for i, value in enumerate(sorted_surpluses):
    plt.text(value, i, f'${value}B', va='center', fontsize=14)
plt.show()
```

Figure 11: Countries Causing China's Greatest Trading Deficits (2022)

```
deficit_countries = ['Taiwan', 'Australia', 'Brazil', 'Switzerland', 'Saudi
Arabia', 'Russia', 'South Korea', 'Oman', 'Iraq', 'Chile']
deficits = [-156.5, -63.3, -47.6, -42.2, -40.1, -38.0, -37.0, -32.0, -25.4, -
22.0]
plt.figure(figsize=(12, 8))
plt.barh(deficit_countries, deficits, color='lightcoral')
plt.xlabel('Trading Deficits (in billion USD)')
plt.ylabel('Countries')
plt.grid(axis='x', linestyle='--', alpha=0.6)
```

```
for i, value in enumerate(deficits):
    plt.text(value, i, f'${abs(value)}B', va='center', ha='left' if value < 0
else 'right', fontsize=14)
plt.show()
```

Figure 12: Foreign Direct Investment in China (1979-2022)

```
forinvestements = pd.read_csv('china-foreign-direct-
investment.csv',skiprows=16)
plt.figure(figsize=(12, 8))
plt.plot(forinvestements['year'], forinvestements[' Inflows  US $'])
plt.axvspan(1997, 1998, color='mediumturquoise', alpha=0.3)
plt.axvline(x=2009, color='green', linestyle=':')
plt.text(2009, 10.435, 'Financial Crisis', rotation=0, fontsize=16,
color='green')
plt.text(1997, 60.4801,  'Ease-Asian Financial Crisis', rotation=90,
fontsize=14, color='black')
plt.axvspan(2019, 2021, color='green', alpha=0.3)
plt.text(2020, 50.4801,  'Covid-19 pandemic', rotation=90, fontsize=14,
color='black')
plt.xlabel('Years')
plt.ylabel('Billions of US $')
plt.xticks(forinvestements['year'], rotation=90)
plt.grid()
plt.show()
```

Figure 13: Distribution of Inward FDI Stock in China in 2021 by Country/Region of Origin

```
countries = ['Hong Kong', 'Virgin Islands', 'Japan', 'Singapore', 'United
States', 'South Korea',
             'Taiwan', 'Cayman Islands', 'Germany', 'Samoa', 'United Kingdom',
'Netherlands',
             'Macao', 'France', 'Mauritius', 'Bermuda', 'Australia', 'Canada',
'Switzerland', 'Malaysia']
fdi = [54.70, 6.90, 4.70, 4.60, 3.50, 3.40, 2.70, 1.90, 1.50, 1.20, 1.10, 1,
0.80, 0.70, 0.60, 0.50, 0.40, 0.40, 0.40, 0.30]
plt.figure(figsize=(12, 8))
plt.barh(countries, fdi, color='skyblue')
plt.xlabel('Inward FDI Stock (Percentage)')
plt.grid(axis='x', linestyle='--', alpha=0.6)
for i, value in enumerate(fdi_stock):
    plt.text(value, i, f'{value}%', va='center')
plt.show()
```

Figure 14: Comparison of Economic Metrics

```
inflation=pd.read_csv('china_inflation_new.csv', skiprows=16)
inflation['date'] =pd.to_datetime(inflation['date'])
inflation['date'] =inflation['date'].dt.year
gdp = pd.read_csv('china_gdp_before.csv', skiprows=16)
imports = pd.read_csv('china_imports.csv', skiprows=16)
imports['date'] = pd.to_datetime(imports['date'])
imports['date'] =imports['date'].dt.year
exports = pd.read_csv('china_exports.csv', skiprows=16)
exports['date'] =pd.to_datetime(exports['date'])
exports['date'] =exports['date'].dt.year
forinvestments = pd.read_csv('china-foreign-direct-investment.csv',
skiprows=16)
forinvestments['date'] =pd.to_datetime(forinvestments['date'])
```

```
forinvestments['date'] = forinvestments['date'].dt.year
inf=inflation[inflation['date'] > 1987]
gd=gdp[gdp['date'] > 1987]
imp=imports[imports['date'] > 1987]
exp=exports[exports['date'] > 1987]
fori=forinvestments[forinvestments['date'] > 1987]
inf_scaled=inf.iloc[:, 1]
gd_scaled = gd.iloc[:, 1]
imp_scaled=imp.iloc[:, 1]
exp_scaled= exp.iloc[:, 1]
fori_scaled =fori.iloc[:, 1]
scaler_minmax= MinMaxScaler()
inf_minmax =scaler_minmax.fit_transform(inf_scaled.values.reshape(-1, 1))
gd_minmax =scaler_minmax.fit_transform(gd_scaled.values.reshape(-1, 1))
imp_minmax=scaler_minmax.fit_transform(imp_scaled.values.reshape(-1, 1))
exp_minmax =scaler_minmax.fit_transform(exp_scaled.values.reshape(-1, 1))
fori_minmax =scaler_minmax.fit_transform(fori_scaled.values.reshape(-1, 1))
plt.figure(figsize=(12, 8))
plt.plot(inf_minmax, label='Inflation Rate (Scaled)')
plt.plot(gd_minmax, label='GDP (Scaled)')
plt.plot(imp_minmax, label='Imports (Scaled)')
plt.plot(exp_minmax, label='Exports (Scaled) ')
plt.plot(fori_minmax, label='Foreign Direct Investments (Scaled) ')
plt.xlabel('Years')
plt.ylabel('Standardized Value')
plt.legend(loc=9)
plt.grid()
plt.show()
```

Figure 15: GDP Per Capita of China Line plot (1960 -2022)

```
series= pd.read_csv('china_gdp_before.csv',skiprows=16)
X=series[' Per Capita (US $)']
plt.figure(figsize=(6, 4))
plt.plot(series['date'], X)
plt.ylabel('Per Capita US$')
plt.xlabel('Date')
plt.grid()
plt.show()
```

Figure 16: GDP of China Histogram of frequency distribution of values

```
series= pd.read_csv('china_gdp_before.csv',skiprows=16)
X=series[' Per Capita (US $)']
pyplot.figure(1)
pyplot.subplot(211)
X.hist()
plt.xlabel('Values')
plt.ylabel('Frequency')
pyplot.show()
```

Figure 17: Box and Whiskers plot for Chinese GDP (1960-2022)

```
series = pd.read_csv('china_gdp_before.csv', skiprows=16)
X=series[' Per Capita (US $)']
series['10-Year Interval']=series['date'] // 10 * 10
plt.figure(figsize=(6, 4))
sns.boxplot(x='10-Year Interval', y=X, data=series)
plt.xlabel('10-year Interval')
```

```
plt.ylabel('Per Capita US$')
plt.grid()
plt.show()
```

GDP Manual model construction

Dickey Fuller

```
series = pd.read_csv('china_gdp_before.csv', skiprows=16)
X = series[' Per Capita (US $)']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Figure 18: GDP ACF and PACF plot

```
series = pd.read_csv('china_gdp_before.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Per Capita (US $)']
pyplot.figure()
pyplot.subplot(211)
plot_acf(X, lags=30, ax=pyplot.gca())
pyplot.subplot(212)
plot_pacf(X, lags=30, ax=pyplot.gca())
pyplot.show()
```

Figure 19: GDP prediction on train data with ARIMA model (4, 1, 2)

```
dates = pd.to_datetime(series['date'])
train_size = int(len(X) * 0.88)
train, test = X[0:train_size], X[train_size:len(X)]
order = (4, 1, 2)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])

arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)

plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Per Capita US$')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()
arma_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
```

Automatic model for GDP

```
model = auto_arima(train, start_p=0, max_p=6, start_q=0, max_q=6, max_d=2,
suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(order=model.order)
model_fit.fit(train)
print("ARIMA Model Parameters:")
print(model_fit.summary())
```

Figure 20: GDP predictions on train set with SARIMAX (2, 2, 2)

```
SARIMAmodel = SARIMAX(train, order=order)
SARIMAmodel_fit = SARIMAmodel.fit()
y_pred = SARIMAmodel_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = SARIMAmodel_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Per Capita US$')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='SARIMA Predictions')
plt.legend()
plt.show()
sarima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", sarima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"]) / test)
* 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Figure 21: GDP prediction with ARIMA (2, 2, 1)

```
dates = pd.to_datetime(series['date'])
forecast_steps = 10
arima_model = SARIMAX(X, order=(2, 2, 1))
arima_fit = arima_model.fit()
forecast_index = pd.date_range(start=dates.iloc[-1], periods=forecast_steps +
1, freq='A')[1:]
forecast_values = arima_fit.get_forecast(steps=forecast_steps).predicted_mean

plt.figure(figsize=(12, 8))
plt.plot(dates, X, label='Original Data')
plt.plot(forecast_index[:2], forecast_values[:2], label='2-year
Forecast',linestyle='dashed', color='red', marker='o')
plt.plot(forecast_index[2:], forecast_values[2:],  label='8-year Forecast',
linestyle='dashed', color='black', marker='o')
for i, txt in enumerate(forecast_values[:2]):
    if i == 0 or i == 1:
        plt.text(forecast_index[i], txt, f'{round(txt, 2)}', ha='right',
va='bottom', color='red')
for i, txt in enumerate(forecast_values[5:9]):
    if i == 0 or i == 3:
        plt.text(forecast_index[i + 5], txt, f'{round(txt, 2)}', ha='right',
va='bottom', color='black')
plt.title('')
```

```
plt.xlabel('Index')
plt.ylabel('Inflows (US $)')
plt.legend()
plt.grid()
plt.show()
```

Figure 22: Inflation Rate of China Line plot (1987-2022)

```
series = pd.read_csv('china_inflation_new.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflation Rate (%)']
plt.figure(figsize=(6, 4))
plt.plot(series['date'], X)
plt.ylabel('Per Capita US$')
plt.xlabel('Date')
plt.grid()
plt.show()
```

Figure 23: Inflation of China Histogram of frequency distribution of values

```
series = pd.read_csv('china_inflation_new.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflation Rate (%)']
pyplot.figure(1)
pyplot.subplot(211)
X.hist()
plt.xlabel('Values')
plt.ylabel('Frequency')
pyplot.show()
```

Figure 24: Inflation Rates of China Box and Whisker plot

```
series = pd.read_csv('china_inflation_new.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflation Rate (%)']
series['10-Year Interval']=series['date'] // 10 * 10
plt.figure(figsize=(12, 6))
sns.boxplot(x='10-Year Interval', y=X, data=series)
plt.xlabel('10-year Interval')
plt.ylabel('Inflation Rate (%)')
plt.grid()
plt.show()
```

Inflation Manual model construction

Dickey Fuller

```
series = pd.read_csv('china_inflation_new.csv', skiprows=16)
X = series[' Inflation Rate (%)']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Figure 25: Inflation ACF and PACF

```
series = pd.read_csv('china_inflation_new.csv', skiprows=16)
X = series[' Inflation Rate (%)']
pyplot.figure()
pyplot.subplot(211)
plot_acf(X, lags=17, ax=pyplot.gca())
pyplot.subplot(212)
plot_pacf(X, lags=17, ax=pyplot.gca())
pyplot.show()
```

Figure 26: Inflation prediction on train data with ARIMA (2, 1, 2) model

```
train_size = int(len(X) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]
order = (2, 1, 2)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
dates = pd.to_datetime(series['date'])
plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Inflation Rate (%)')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()
arma_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arma_rmse)
```

Automatic model for Inflation

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit = model_fit.fit()
print("ARIMA Model Parameters:")
print(model_fit.summary())
```

Figure 27: Inflation rate predictions on training set with ARIMA(0, 1, 2

```
order = (0, 1, 2)
arima_model = ARIMA(train, order=order)
arima_fit = arima_model.fit()
y_pred = arima_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = arima_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Inflation Rate (%)')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
```

```
plt.legend()
plt.show()
arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"]) / test)
* 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Figure 28: Inflation prediction on ARIMA (0, 1, 2) model

```
dates = pd.to_datetime(series['date'])
forecast_steps = 10
arima_model = ARIMA(X, order=(0, 1, 2))
arima_fit = arima_model.fit()
forecast_index = pd.date_range(start=dates.iloc[-1], periods=forecast_steps +
1, freq='A')[1:]
forecast_values = arima_fit.get_forecast(steps=forecast_steps).predicted_mean

plt.figure(figsize=(12, 8))
plt.plot(dates, X, label='Original Data')
plt.plot(forecast_index[:2], forecast_values[:2], label='2-year
Forecast',linestyle='dashed', color='red', marker='o')
plt.plot(forecast_index[2:], forecast_values[2:],  label='8-year Forecast',
linestyle='dashed', color='black', marker='o')
plt.text(36, 3.9 , ' 3.11', ha='right', va='bottom', color='red')
plt.text(39, 3 , '2.27', ha='right', va='bottom', color='red')
plt.xlabel('Index')
plt.ylabel('Inflows (US $)')
plt.legend()
plt.grid()
plt.show()
```

Figure 29: FDI inflows of China Line plot (1978-2022)

```
series = pd.read_csv('china-foreign-direct-investment.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflows  US $']
plt.figure(figsize=(6, 4))
plt.plot(series['date'], X)
plt.ylabel('Inflows US$')
plt.xlabel('Years')
plt.grid()
plt.show()
```

Figure 30: FDI inflows of China Histogram of frequency distribution of values

```
series = pd.read_csv('china-foreign-direct-investment.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflows  US $']
pyplot.figure(1)
pyplot.subplot(211)
X.hist()
plt.xlabel('Values')
plt.ylabel('Frequency')
pyplot.show()
```

Figure 31: FDI inflows of China Box and Whisker plot (1978-2022)

```
series = pd.read_csv('china-foreign-direct-investment.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflows  US $']
series['10-Year Interval']=series['date'] // 10 * 10
plt.figure(figsize=(12, 6))
sns.boxplot(x='10-Year Interval', y=X, data=series)
plt.xlabel('10-year Interval')
plt.ylabel('Inflows US$')
plt.grid()
plt.show()
```

FDI Manual model construction

Dickey Fuller

```
X= series[' Inflows  US $']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Figure 32: FDI ACF and PACF plots

```
series = pd.read_csv('china-foreign-direct-investment.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'], format='%d/%m/%Y')
series['date'] = series['date'].dt.year
X = series[' Inflows  US $']
pyplot.figure()
pyplot.subplot(211)
plot_acf(X, lags=20, ax=pyplot.gca())
pyplot.subplot(212)
plot_pacf(X, lags=20, ax=pyplot.gca())
pyplot.show()
```

Figure 33: FDI predictions on train data with ARIMA (4, 1, 1) model

```
train_size = int(len(y) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]

order = (4, 1, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)

plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Inflows US$')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
```

```
plt.show()
```

Automatic model for GDP

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit_result = model_fit.fit()
print("\nARIMA Model Parameters:")
print(model_fit_result.summary())
```

Figure 34: FDI predictions on train set with ARIMA (0, 1, 0)

```
order = (0, 1, 0)
arima_model = ARIMA(train, order=order)
arima_fit = arima_model.fit()
y_pred = arima_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = arima_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Inflows US$')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()

arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"]) / test)
* 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Figure 35: FDI forecast with ARIMA (0, 1, 0)

```
arima_model = ARIMA(X, order=(0, 1, 2))
arima_fit = arima_model.fit()
forecast_steps = 10
forecast = arima_fit.forecast(steps=forecast_steps)
forecast_dates = pd.date_range(start=X.index[-1] + pd.DateOffset(years=1),
periods=forecast_steps, freq='A')

plt.figure(figsize=(10, 6))
plt.plot(X.index, X, label='Original Data')
plt.plot(forecast_dates[:2], forecast[:2], label='2-year Forecast',
linestyle='solid', color='red')
plt.plot(forecast_dates[2:], forecast[2:], label='Next 8 Forecasts',
linestyle='dashed', color='black')
plt.annotate(f'{forecast[0]:.2f}', (forecast_dates[1], forecast[1]),
textcoords="offset points", xytext=(0, 10), ha='center', color='red')
plt.text(2023, 3.9 , ' 3.11', ha='right', va='bottom', color='red')
plt.text(2024, 3 , '2.27', ha='right', va='bottom', color='red')
plt.xlabel('Year')
plt.ylabel('Inflows (US $)')
plt.legend()
plt.grid()
```

```
plt.show()
```

Figure 36: Imports of China Line plot (1960-2022)

```
series = pd.read_csv('china-imports.csv', skiprows= 16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
plt.figure(figsize=(6, 4))
plt.plot(series['date'], X)
plt.ylabel('Billions of US $')
plt.xlabel('Years')
plt.grid()
plt.show()
```

Figure 37: Imports of China Histogram of frequency distribution of values

```
series = pd.read_csv('china-imports.csv', skiprows= 16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
pyplot.figure(1)
pyplot.subplot(211)
X.hist()
plt.xlabel('Values')
plt.ylabel('Frequency')
pyplot.show()
```

Figure 38: Imports of China Box and Whisker plot (1960-2022)

```
series = pd.read_csv('china-imports.csv', skiprows= 16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
series['10-Year Interval']=series['date'] // 10 * 10
plt.figure(figsize=(6, 4))
sns.boxplot(x='10-Year Interval', y=X, data=series)
plt.xlabel('10-year Interval')
plt.ylabel('Billions of US $')
plt.grid()
plt.show()
```

Imports Manual model construction

Dickey Fuller

```
series = pd.read_csv('china-imports.csv', skiprows= 16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Figure 39: Imports ACF and PACF

```
series = pd.read_csv('china-imports.csv', skiprows= 16)
series['date'] = pd.to_datetime(series['date'])
```

```
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
pyplot.figure()
pyplot.subplot(211)
plot_acf(X, lags=17, ax=pyplot.gca())
pyplot.subplot(212)
plot_pacf(X, lags=17, ax=pyplot.gca())
pyplot.show()
```

Figure 40: Imports predictions on train data with ARIMA (4, 1, 1) model

```
train_size = int(len(y) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]

order = (4, 1, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)

plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()
```

Automatic model for Imports

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit_result = model_fit.fit()
print("\nARIMA Model Parameters:")
print(model_fit_result.summary())
```

Figure 41: Imports predictions with training set with ARIMA (0, 2, 1)

```
order = (0, 2, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel_fit = ARIMAmodel.fit()
y_pred = ARIMAmodel_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
plt.plot(dates[0:train_size], train, color="black", label="Training Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Billion of US$')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()
arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
```

```
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"]) / test)
* 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Figure 42: Imports Forecast with ARIMA (0, 2, 1) model

```
model = ARIMA(X, order=(0, 2, 1))
results = model.fit()
forecast_steps = 10
forecast_index = np.arange(len(X), len(X) + forecast_steps)
forecast_values = results.get_forecast(steps=forecast_steps).predicted_mean

dates = pd.to_datetime(series['date'])
plt.figure(figsize=(12, 8))
plt.plot(dates, X, label='Original Data')
plt.plot(forecast_index[:2], forecast_values[:2], label='2-year
Forecast',linestyle='dashed', color='red', marker='o')
plt.plot(forecast_index[2:], forecast_values[2:],  label='Next 8 Forecasts',
linestyle='dashed', color='black', marker='o')
plt.text(2023, 3000 , '3275.08', ha='right', va='bottom', color='red')
plt.text(2024, 3400 , '3412.57', ha='right', va='bottom', color='red')
plt.text(2031, 4450 , '4512.50', ha='right', va='bottom', color='black')
plt.xlabel('Billion of US $')
plt.ylabel('Values')
plt.legend()
plt.grid()
plt.show()
```

Figure 43: Exports of China Line plot (1960-2022)

```
series = pd.read_csv('china-exports.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
plt.figure(figsize=(6, 4))
plt.plot(series['date'], X)
plt.ylabel('Billions of US $')
plt.xlabel('Years')
plt.grid()
plt.show()
```

Figure 44: Exports Density plot

```
series = pd.read_csv('china-exports.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
pyplot.figure(1)
pyplot.subplot(211)
X.hist()
plt.xlabel('Values')
plt.ylabel('Frequency')
pyplot.show()
```

Figure 45: Exports of China Box and Whisker plot (1960-2022)

```
series = pd.read_csv('china-exports.csv', skiprows=16)
```

```
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
series['10-Year Interval']=series['date'] // 10 * 10
plt.figure(figsize=(6, 4))
sns.boxplot(x='10-Year Interval', y=X, data=series)
plt.xlabel('10-year Interval')
plt.ylabel('Billions of US $')
plt.grid()
plt.show()
```

Exports Manual model construction

Dickey Fuller

```
series = pd.read_csv('china-exports.csv', skiprows=16)
X = series[' Billions of US $']result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

Figure 46: Exports ACF and PACF

```
series = pd.read_csv('china-exports.csv', skiprows=16)
series['date'] = pd.to_datetime(series['date'])
series['date'] = series['date'].dt.year
X = series[' Billions of US $']
pyplot.figure()
pyplot.subplot(211)
plot_acf(X, lags=20, ax=pyplot.gca())
pyplot.subplot(212)
plot_pacf(X, lags=20, ax=pyplot.gca())
pyplot.show()
```

Figure 47: Exports predictions on train data with ARIMA (5, 2, 1) model

```
train_size = int(len(y) * 0.88)
train, test = y[0:train_size], y[train_size:len(y)]

order = (5, 2, 1)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel = ARIMAmodel.fit()
y_pred = ARIMAmodel.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)

plt.figure()
plt.plot(dates[0:train_size], train, color="black", label="Train Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Billions of US $')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()
```

Automatic model for Exports

```
model = auto_arima(train, suppress_warnings=True)
print("Best ARIMA Order:", model.order)
model_fit = ARIMA(train, order=model.order)
model_fit_result = model_fit.fit()
print("\nARIMA Model Parameters:")
print(model_fit_result.summary())
```

Figure 48: Exports predictions made with the training set with ARIMA (0, 2, 3)

```
order = (0, 2, 3)
ARIMAmodel = ARIMA(train, order=order)
ARIMAmodel_fit = ARIMAmodel.fit()
y_pred = ARIMAmodel_fit.get_forecast(steps=len(test))
y_pred_df = y_pred.conf_int(alpha=0.05)
y_pred_df["Predictions"] = ARIMAmodel_fit.predict(start=y_pred_df.index[0],
end=y_pred_df.index[-1])
plt.plot(dates[0:train_size], train, color="black", label="Train Data")
plt.plot(dates[train_size:], test, color="red", label="Test Data")
plt.ylabel('Billions of US $')
plt.xlabel('Years')
plt.plot(dates[train_size:], y_pred_df["Predictions"], color='green',
label='ARIMA Predictions')
plt.legend()
plt.show()
arima_rmse = np.sqrt(mean_squared_error(test, y_pred_df["Predictions"]))
print("RMSE:", arima_rmse)
absolute_percentage_errors = np.abs((test - y_pred_df["Predictions"]) / test)
* 100
mape = absolute_percentage_errors.mean()
print(f'MAPE: {mape:.2f}%')
```

Figure 49: Exports Forecast with ARIMA (0, 2, 3) model

```
model = ARIMA(X, order=(0, 2, 3))
results = model.fit()
forecast_steps = 10
forecast_values = results.get_forecast(steps=forecast_steps).predicted_mean

plt.figure(figsize=(10, 6))
plt.plot(dates, X, label='Original Data')
plt.plot(forecast_index[:2], forecast_values[:2], label='2-year
Forecast',linestyle='dashed', color='red', marker='o')
plt.plot(forecast_index[2:], forecast_values[2:],  label='Next 8 Forecasts',
linestyle='dashed', color='black', marker='o')
plt.text(pd.to_datetime('2022-01-01'), 3400,  '3712.18', rotation=0,
fontsize=11, color='red')
plt.text(pd.to_datetime('2019-01-01'), 4200,  '3920.44', rotation=0,
fontsize=11, color='red')
plt.text(pd.to_datetime('2025-01-01'), 5400,  '5392.44', rotation=0,
fontsize=11, color='black')
plt.title('0,2,3 model')
plt.xlabel('Years')
plt.ylabel('Billions of US $')
plt.xticks()
plt.legend()
plt.grid()
plt.show()
```

Figure 50: Forecasted Values of China (2023-2032)

```
years = np.array([2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031,
2032]).reshape(-1, 1)
forecast_gdp = np.array([12600.801514, 14678.500482, 15295.853520,
14910.495259, 16695.318436,
                         17791.912086, 17323.365909, 18710.138270,
20190.148504, 19812.701847]).reshape(-1, 1)
forecast_inflation = np.array([3.115434, 2.279638, 2.279638, 2.279638,
2.279638, 2.279638,
                               2.279638, 2.279638, 2.279638,
2.279638]).reshape(-1, 1)
forecast_imports = np.array([3275.085283, 3412.576458, 3550.067632,
3687.558806, 3825.049981,
                             3962.541155, 4100.032329, 4237.523504,
4375.014678, 4512.505852]).reshape(-1, 1)
forecast_exports = np.array([3712.180218, 3920.447852, 4104.448050,
4288.448247, 4472.448445,
                             4656.448642, 4840.448839, 5024.449037,
5208.449234, 5392.449432]).reshape(-1, 1)
forecast_fdi = np.array([180.166881, 180.166881, 180.166881, 180.166881,
180.166881,
                         180.166881, 180.166881, 180.166881, 180.166881,
180.166881]).reshape(-1, 1)
combined = np.concatenate([forecast_gdp, forecast_inflation, forecast_imports,
                                forecast_exports, forecast_fdi], axis=1)
scaler = MinMaxScaler()
scaled_data = scaler.fit_transform(combined)
forecast_gdp, forecast_inflation, forecast_imports, forecast_exports,
forecast_fdi = np.hsplit(scaled_data, 5)
plt.figure(figsize=(10, 6))
plt.plot(years, forecast_gdp, label='Forecast for GDP Per Capita', marker='o')
plt.plot(years, forecast_inflation, label='Forecast for Inflation Rate',
linewidth= 2, marker='o')
plt.plot(years, forecast_imports, label='Forecast for Imports', marker='o')
plt.plot(years, forecast_exports, label='Forecast for Exports', marker='o')
plt.plot(years, forecast_fdi, label='Forecast for FDI', marker='o')
plt.xlabel('Years')
plt.ylabel('Forecasted Values (Scaled)')
plt.legend()
plt.grid()
plt.show()
```

Appendix D – Certificate of Ethics Approval

Coventry University

Certificate of Ethical Approval

Applicant:	Alexandra Espialidou
Project Title:	A time series analysis of the Chinese economy

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:	09 Oct 2023
Project Reference Number:	P164785