S.T. Yau High School Science Award

Research Report

The Team

Name of team member: Junting Wang
School: The Peddie School
City, Country: Hightstown, NJ, United States

Name of supervising teacher: Chengguang Zhu
Job Title: Assistant Professor
School/Institution: Shanghai Jiao Tong University
City, Country: Shanghai, China

Title of Research Report

Modeling and Regulation of Carbon Dioxide Emissions with Green Finance - Using Beijing, Chongqing, and Shanghai as a Case Study

Date

August 20, 2023

Modeling and Regulation of Carbon Dioxide Emissions with Green Finance - Using Beijing, Chongqing, and Shanghai as a Case Study

Junting Wang

junting.w@outlook.com
Peddie School

August 20, 2023

Abstract

The global greenhouse effect has attracted significant attention in recent years. Most of the current models regarding carbon dioxide (CO2) emission control apply multiple linear regression analyses on population, gross domestic product (GDP), and energy consumption, disregarding Green Finance Index (GFI) as a driving factor. However, the GFI measures how well a region's financial activities align with environmental goals, which should have an important impact on carbon emissions. We aim to conduct a comprehensive analysis of carbon emissions in China's big cities, highlighting the impact of GFI on carbon emissions and modeling the excessive emissions to be represented in monetary values. Specifically, we picked three cities to serve as case studies—Beijing, Chongqing, and Shanghai.

This study establishes an autoregressive integrated moving average (ARIMA) model to predict future values of the four driving forces (resident population, GDP, energy consumption, GFI); a backpropagated neural network (BPNN) model to predict carbon emissions; and a cost model to predict the cost related to excessive carbon emissions based on a fictitious scenario inspired by the US's emission goals.

The result shows that GFI significantly and negatively impacts carbon emissions. Therefore, increasing the GFI is an effective measure to ensure the realization of peak carbon emissions before 2030, which lowers the cost caused by the risk of excessive carbon emission. This study proposes a carbon emission and cost model that can provide a reference for the control of carbon emissions in Beijing, Chongqing, and Shanghai.

Keywords: peak carbon emissions, green finance index, autoregressive integrated moving average model, backpropagated neural network, cost model

Contents

1 Introduction

Greenhouse gas (GHG) emissions, especially carbon emissions, is causing the global warming effect. CO2 is especially concerning because they last in the atmosphere for an extremely long time, which means it can continuously cause global warming effect even thousands of years after being emitted USEPA (2023). Moreover, development in industry and transportation has caused increasing amounts of carbon emissions.

To address the over-emission of CO2 into the atmosphere, the global community set goals to reduce carbon emissions. The Paris Agreement was adopted by 196 countries at the UN Climate Change Conference (COP21) in Paris on December 12, 2015. The agreement went into effect in late 2016 and aims to reduce GHG emissions so that global temperature increase in the 21st century is limited to 2°C, and push for controlling it at 1.5°C UNFCCC. Each country is required to submit a Nationally Determined Contribution (NDC) report every 5 years to outline the actions to reach the goal. By 2020, lead GHG emission countries have all set goals specific for themselves regarding carbon emissions. The EU and US had declared to achieve peak carbon emissions by 2025 and net-zero carbon emissions by 2050 EU; CSO. China has set the goal of peaking carbon emissions by 2030 and reaching net-zero carbon emissions by 2060 Tay (2022).

China is currently the world's largest emitter of CO2, accounting for 30.9% of total global CO2 emissions in 2021 Ritchie (2019). China's annual CO2 emissions have been increasing since the 1970s and increased in pace in the 21st century. China's annual CO2 emissions in 2000 were 3.64 billion tons while the figure in 2021 is 11.47 billion tons, a 215% increase Ritchie and Roser (2020). To counter climate change, China sets specific goals such as lowering carbon emissions per unit of GDP by over 65% from the 2005 level and increasing the total installed capacity of wind and solar power to over 1.2 billion kW Chi (2021).

In addition to establishing targets, the global community introduced the concept of the green finance index (GFI) to link environmental issues with finance. The GFI assesses the effectiveness of financial markets and institutions in promoting environmentally sustainable investments, in essence measuring how well a region's financial activities align with environmental goals. GFI is related to many novel financial concepts such as the trading of carbon credit and the introduction of carbon tax, both provide incentives for corporates to lower carbon emissions. Therefore, regulating GFI has an important impact on urban carbon emissions.

Current approaches to lowering carbon emissions usually come with costs related to purchasing high-end technology or promoting the use of renewable energy sources. This research tries to link the incentive of lowering carbon emissions with monetary values, creating a fictitious scenario where a region need to "pay" for excessive carbon emissions above a certain target. We assume that each region needs to purchase its excessive carbon emissions based on market price. We refer to this payment amount as the cost related to excessive carbon emissions.

The US submitted an Intended Nationally Determined Contribution (INDC) to the UNFCCC in 2015, which outlines the target to cut carbon emissions by 26-28% below 2005 levels by 2025 IND (2015). This is an ambitious target set by President Obama's administration but the US had already reached peak carbon emissions in 2007, so the target is still feasible.

This research calculates the cost related to excessive carbon emissions in the framework of a fictitious scenario in China based on the US's INDC. Considering that China has not reached peak carbon emissions yet, we set the target for each of the three cities (Beijing, Chongqing, and Shanghai) to cut carbon emissions by 25% below 2019 levels by 2030.

Instead of focusing on a specific area like most current research, this paper combined many of the previous categories. It investigates the driving forces of carbon emissions, predicts carbon emissions for 11 years, and calculates costs related to the emissions. This research utilizes some of the same methods that were included in previous research such as ARIMA but applies them to different kinds of data. This research also implements certain methods that were rarely used before such as BPNN and cost analysis.

The research process in this research is shown in Figure 1. The first phase is the data preparation phase, where I obtained raw data and "cleaned" it for later analysis. The second phase is the data visualization phase, where the processed data are graphed without applying analytical algorithms. The third phase is the data analysis phase, where varying algorithms are applied to predict values of driving factors of CO2 emissions, the amount of CO2 emissions, and the cost needed to reach carbon emission goals. The fourth phase is the result analysis phase, where we analyze the results from different algorithms together and outline the policy implications.

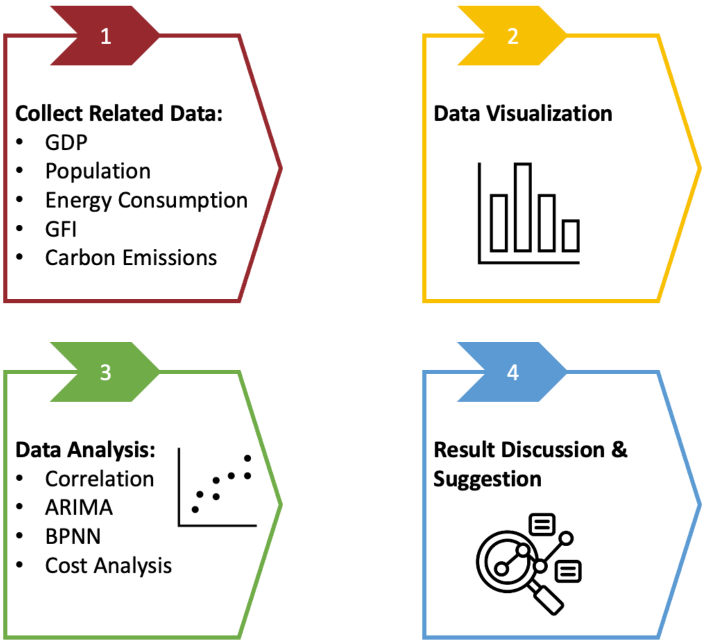


Figure 1: The overall research process

The remainder of this paper is organized as follows: Section 2 outlines related works; Section 3 introduces the data sources and different analytical approaches; Section 4 outlines the results; Section 5 presents the conclusion and related policy proposals.

2 Related Work

Current research related to the goal of this paper can be divided into four categories: predicting carbon emissions, analyzing the effect of green finance (GF) on carbon emissions, predicting costs related to carbon emissions, and miscellaneous.

2.1 Carbon Emissions Prediction

Some research uses traditional methods to predict carbon emissions. Zhou et al. applied linear regression and driving force models, such as IPAT and STIRPAT, to identify the key driving forces and forecast carbon emissions, achieving an $R^2 \approx 0.77$. They found that the 6 key driving forces are: renewable energy development, market demand changes, energy industry regulations, industrial structure reforms, industrial technology innovation, and accidental events.

Most research, on the other hand, uses machine learning (ML) algorithms. Hou et al. used Pearson correlation to identify 8 groups that have the highest effect on CO2 emissions. They then generated predictions with BPNN and optimized it with the whale algorithm, achieving an $R^2 \approx 0.96$. They also predicted China's carbon peak to be in 2033 with total emission of 10404.045 million tons. Serafeim and Caicedo achieved higher prediction accuracy with the Adaptive Boosting ML algorithm than linear regression or other supervised ML algorithms. Nguyen et al. developed a two-step framework that applies a Meta-Elastic Net learner to combine predictions from multiple base learners to predict corporate carbon emissions. They compared their proposed model to base learners such as OLS, Neural Networks, and KNN and achieved up to 30% better mean absolute error (MAE). Amarpuri et al. used a hybrid model of Convolutional Neural Network and Long Short Term Memory Network (CNN-LSTM) to predict carbon emissions in India from 2018 to 2020. Yao and Zhao used ML to identify structural breaks (notable changes in the trend or level of carbon emission over time) in the top 20 global emitters. Specifically, they implemented both unconditional and conditional analysis alongside kaya identity, which expresses total carbon emissions as a product of four factors (population, GDP per capita, energy intensity, and carbon intensity). They concluded that most structural breaks occur due to changes in the economic structure instead of climate policy and urges policymakers to seek better ways to control carbon emissions.

2.2 The Effect of GF on Carbon Emissions

This category contains a wide variety of models. All of their results are similar in that they identified a strong negative correlation between GF development and carbon emissions.

Chen et al. tested the impact of GF on carbon emissions in 30 Chinese provinces from 2005-2018 with the spatial durbin model (SDM) and found that carbon emissions between regions show a significant spatial positive correlation, meaning that a change in a particular region significantly impacts the neighboring regions in the same direction. Zhang et al. applied a Slack-based model (SBM), which evaluates the efficiency of decision-making units by considering their input and output variables, on data from 27 provinces in China from 2008-2017 to measure carbon emissions efficiency. They then implemented the Torbit model to study the impact of GF on carbon emissions efficiency. They concluded that carbon emissions efficiency is generally low in China and displays a pattern where East China has higher efficiency than West China. Sun collected a large number of relevant economic data from the internet and created a model using big data technology and ML algorithms. Fang et al. used data envelopment analysis (DEA) to construct the energy efficiency index of G7 countries and used panel data model technique to examine the relationship between GF, energy efficiency, and carbon emissions. They found that not only do GF and energy efficiency have a negative impact on carbon emissions, but also implementing effective

GF policies and energy efficiency measures can contribute to economic growth. Cao utilized cross-sectional dependency and slope heterogeneity test to determine if different panels are correlated, Dumitrescu-Hurlin (D-H) panel causality test to validate results, and Common Correlated Effects Mean Group estimator using the Autoregressive Distributed Lag approach (CS-ARDL) to correct for the presence of cross-sectional dependence in panel data. They recommended that developed countries increase cooperation to create more appropriate research and development programs. Xiong and Sun implemented fuzzy set qualitative comparative analysis to panel data of 34 Chinese provinces to analyze the effect of GF on carbon emissions and proposed that the Chinese government should improve the GF system, optimize GF structure, and create incentives for enterprises to participate in GF activities. Guo et al. analyzed data from the Yangtze River Economic Belt from 2006-2019. They applied unit root test and multicollinearity test to prove that data are stationary, stepwise regression model to explain the mediation role of technology innovation in GF and carbon emissions, SDM to measure relations among variables, and endogeneity test to demonstrate the reliability of the model. They proposed to set up a GF development alliance and an information cooperation network. Li et al. examined data from 30 Chinese provinces from 2008-2019 and found that not only does GF promote carbon emission reduction, but it also limits the real estates scale of disorderly expansion.

Some research focused on bibliometric analysis. Zhang et al. reviewed research on GF and carbon emissions reduction based on literature from 2010–2021 in the Web of Science core database. They applied Driving-Pressure-State-Impact-Response (DPSIR) framework to understand the interaction between the environment and society. They summarized that carbon emissions and GF are only recently popular topics and much previous research focused on deforestation and climate change. They also identified a discrepancy in green standards between nations and sectors, which contributed to a majority of researchers studying separately.

2.3 Predicting Cost Related to Carbon Emissions

Research in this category explores the cost related to carbon emissions in different situations and using different methods.

Zagheni and Billari used a stochastic representation of the IPAT equation to explain trends in carbon emissions and estimate the cost related to reducing carbon emissions. The study was conducted under the assumption that the overall cost to reduce emissions is directly related to the amount of emissions, countries will receive an economic incentive that corresponds with the volume of their emissions, and the GDP of a country spent on environmental quality is constant over time.

Qin et al. designed a Stackelberg game to model manufacturer-retailer dynamics under various scenarios, taking into account the presence of green financing and cost-sharing. Their findings reveal that the impact of green financing interest rates on manufacturers' carbon emission reductions is not uniformly negative. Their study emphasizes the importance of banks to make appropriate interest rates for GF to incentive manufacturers to reduce carbon emissions.

Zhang and Wen utilized the deep neural network model TCN-Seq2Seq to predict carbon pricing. The model can handle parallel training for fewer parameters and achieved higher directional accuracy (DA) along with lower mean absolute percentage error (MAPE) and root mean square error (RMSE) when compared with benchmark models such as ARIMA and LSTM.

2.4 Miscellaneous

This category includes research that doesn't fall into any of the three categories mentioned above.

York et al. refined STIRPAT by developing the concept of ecological elasticity (EE), which means the responsiveness/sensitivity of environmental impacts to a change in a driving force. They found that population and affluence increase carbon emissions and energy footprint.

Ip et al. aimed to analyze the impact of GF and urbanization on tourism. They used augmented mean group (AMG) and common correlated effect mean group (CCE-MG) to identify factors that affect Chinas tourism business and used D-H panel causality test to double-check if the factors have an impact on tourism. They found that GF, income, and renewable energy use have a substantial positive effect on the tourism business.

Ghoddusi et al. reviewed more than 130 articles published in high-impact journals from 2005-2018 and analyzed the most used ML models and the advantages and limitations of applying ML on topics around energy economics. The advantages include but are not limited to higher prediction accuracy and the ability to uncover complex relationships. The limitations include but are not limited to over-fitting and the black-box nature of ML algorithms.

Sarfraz et al. analyzed the relationship between carbon emissions and COVID-19 in India and concluded that the COVID-19 lockdown significantly decreased carbon emissions.

Xu et al. constructs a spillover index and analyzes whether there is a relationship between carbon allowance price returns and stock returns of carbon-intensive industries. They applied Multifractal Cross-Correlation Analysis (MFCAA), Detrended Cross-Correlation Coefficient, and modified Time-lagged Detrended Cross-Correlation. They found that there are stronger spillovers in more market-oriented carbon emission trading environments and that allowance allocation plays an important role in deciding the direction of the correlations.

3 Methodology

This research first collects, cleans, and visualizes data. Secondly, the correlation coefficient is computed. Thirdly, ARIMA is applied to predict future values of the four driving forces of all three cities from 2020–2030. Fourth, BPNN is applied to predict carbon emissions of all three cities from 2020–2030. Finally, the cost model is applied to approximate the cost related to controlling carbon emissions to the target value. This process is illustrated in Figure 2.

3.1 Data Collection and Visualization

This paper analyzes many different aspects related to carbon emissions so it requires many different kinds of data. This research also covers a case study of three cities (Beijing, Chongqing, and Shanghai) so it requires local data for each one. Data are obtained through the statistical yearbooks published by both the local and national governments (national bureau of statistics, beijing municipal bureau of statistics, chongqing municipal bureau of statistics, shanghai municipal bureau of statistics). The data include population, GDP in 100 million Yuan, total energy consumption in 10 thousand tons, carbon emissions in 10 thousand tons, and green finance index

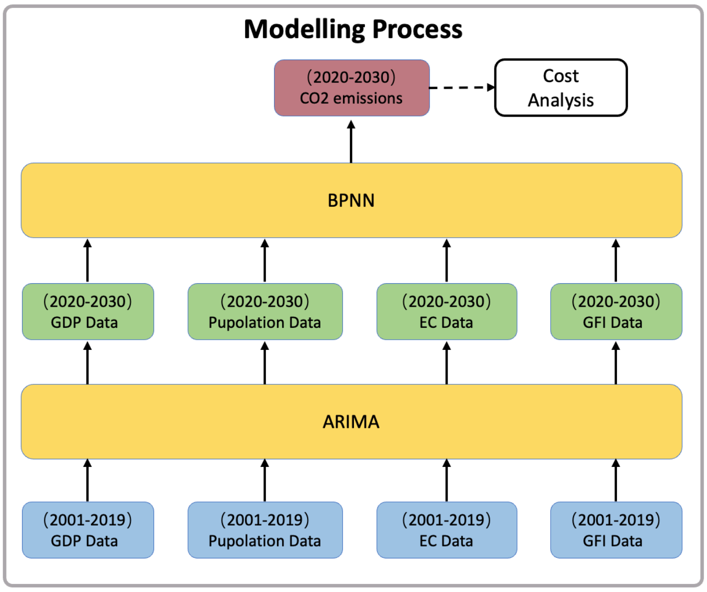


Figure 2: The modeling process (EC is short for energy consumption)

(value between 0 and 1, higher means more developed green financing). The dataset, however, has missing data for certain types of data and certain years. To make sure that data for each city and each category match in all the years used for analysis, the years between 2001–2019 are selected and considered as past data in this research.

The cleaned data is visualized with the years 2001–2019 on the x-axis and the variables on the y-axis. Each graph contains three lines, each representing values from a city. The population is displayed in Figure 3a in terms of 10 thousand; GDP is displayed in Figure 3b in terms of 100 million RMB; energy consumption is displayed in Figure 3c in terms of 10 thousand tons; carbon emissions are displayed in Figure 3d in terms of 10 thousand tons; GFI is displayed in Figure 3e as a value between 0 and 1.

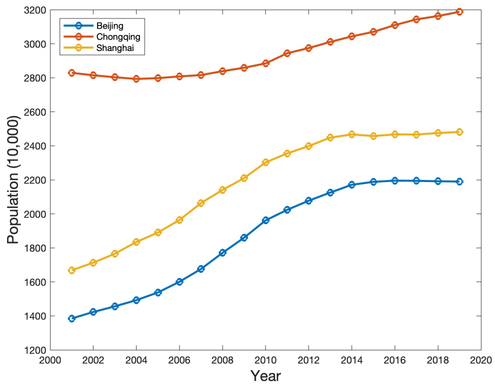
3.2 Correlation Coefficients

To observe the correlation between carbon emissions and four different variables including population, GDP, energy consumption, and GFI, the correlation coefficients are calculated with Equation 1 and shown in Table 1. It is shown from the table that different cities show different correlation behaviors. For both Beijing and Chongqing, CO2 emission has negative correlations with population, GDP, energy consumption, and GFI. For Shanghai, however, CO2 emission has positive correlations with population, GDP, energy consumption, and GFI.

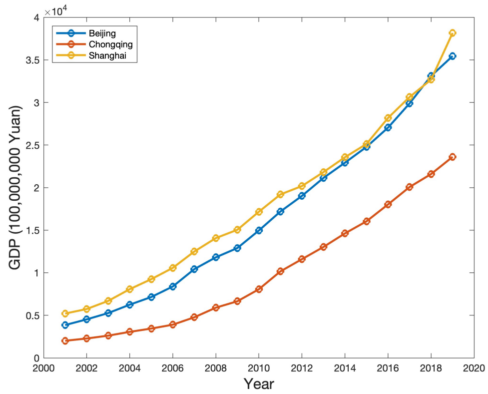
$$\frac{\sum_{i=1}^{n}\left[\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}} \tag{1}$$

	Population	GDP	Energy Consumption	GFI
Beijing	-0.743	-0.697	-0.723	-0.694
Chongqing	-0.473	-0.488	-0.566	-0.485
Shanghai	0.936	0.770	0.956	0.777

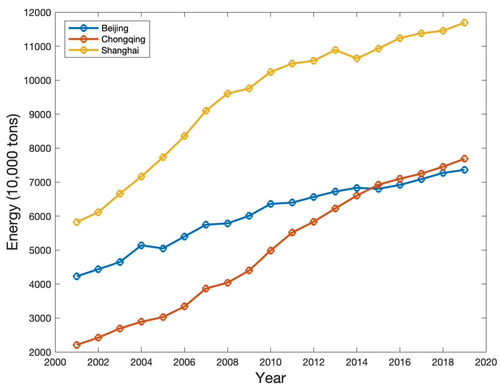
Table 1: The correlation factor between carbon emissions and the four driving forces



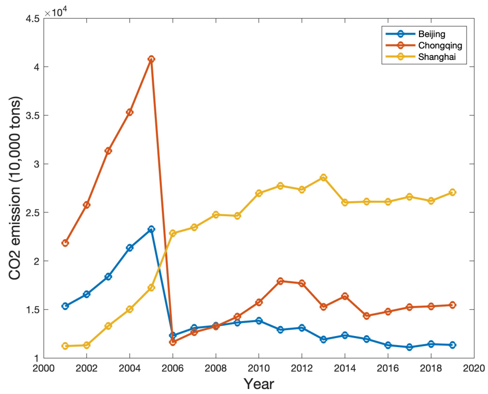
(a) Population of Beijing, Chongqing, and Shanghai from 2001 to 2019



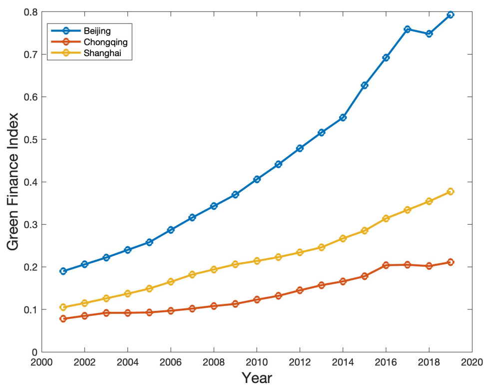
(b) GDP of Beijing, Chongqing, and Shanghai from 2001 to 2019



(c) Energy consumption of Beijing, Chongqing, and Shanghai from 2001 to 2019



(d) Carbon emission of Beijing, Chongqing, and Shanghai from 2001 to 2019



(e) Green Finance Index of Beijing, Chongqing, and Shanghai from 2001 to 2019

Figure 3

3.3 Predicting Driving Forces using ARIMA

3.3.1 ARIMA Introduction

The ARIMA model can be easily understood by breaking the name into three parts: AR stands for autoregression, meaning that the model regresses on its prior values; I stands for integrated, meaning that data values are replaced by the differences between two consecutive values; MA stands for moving average, meaning the model considers the error or noise in the lagged observation.

Matlab provides a specific function to build an ARIMA model: `arima(p,d,q)`, which uses the typical parameters of an ARIMA model. The p value is known as the lag order and determines the number of autoregressive (AR) terms that the model considers. This value determines how reliant the model is on past data points. Essentially, a p value of 1 means that the model's output for time t directly relies on time $t-1$. A higher p value indicates that the prediction directly relies on more past data points. The d value represents the order of differencing, which represents the 'I' piece of the model. The d value should be equal to the number of times a difference calculation needs to be applied to the time series to result in a stationary series. Usually, for a time series with a linear trend, $d = 1$ should be used. The q value is known as the order of moving average and determines the number of moving average (MA) terms that the model considers. As a result, $q = 1$ means that the output for time t is directly related to the error or noise calculation of time $t-1$. A higher q value indicates that the prediction is directly related to the error or noise calculation of more past data points.

Given that X_t represents the series created from the original dataset and is non-stationary, differencing X_t for d times can result in Y_t, a stationary series. According to the `ARIMA(p,d,q)`, Y_t can be represented by Equation 2 and Equation 3.

$$Y_t = \nabla^d X_t \tag{2}$$

$$Y_t = c + \varphi Y_{t-1} + \varphi Y_{t-2} + \cdots + u_t + \theta u_{t-1} + \theta u_{t-2} + \cdots + \theta_q u_{t-1} \tag{3}$$

Box Jenkins Method This method is usually used to determine the p, d, and q values for an ARIMA model and check if the model is a good fit. The next paragraph explains a simple way to use this method.

To determine the parameter values, one should first use the Dickey-Fuller test to identify if the time series is stationary. In case when the time series is non-stationary, it is necessary to difference the series d times to make it stationary, thus the d value is obtained. The p and q values are related to two functions, autocorrelation (ACF) and partial autocorrelation (PACF), and can be determined by visualizing the ACF and PACF graphs of the stationary time series. Usually, if the ACF graph trails off while the PACF graph has a hard cutoff after a lag, the model is AR with p set to the lag of PACF before the cutoff. On the other hand, if the PACF graph trails off while the ACF graph has a hard cutoff after a lag, the model is MA with q set to the lag of ACF before the cutoff. The model will be a mix of AR and MA if both the ACF graph and the PACF graph trail off.

To check if the model is a good fit, researchers should visualize the residuals and related information to ensure the residuals are normally distributed and uncorrelated.

Although the Box Jenkins method seems straightforward, the ACF and PACF graphs usually don't show a clear trend. Thus, the researcher needs to subjectively determine the parameter values mostly from experience. This method also doesn't provide a solution if a mix of AR and MA is used.

3.3.2 ARIMA Setup For This Research

This research took a more objective approach to obtain fitting parameters of ARIMA. For each of the four driving forces for each city, the 19-year time series is split into train and test data in a 3.75:1 ratio. The training set has data from 2001 to 2015 while the testing set has data from

2016 to 2019. Each training set is treated with 128 different ARIMA models. None of the time series are stationary in this research, so only $d = 1$ and $d = 2$ are considered. The p and q values each range from 0 to 7. A Matlab script is used to automatically run the 128 models to predict the values in the following years (2016-2019). Then, the root mean squared error (RMSE) of the predicted values and the testing set is calculated and stored in the matrix. The script keeps track of the lowest RMSE value and its corresponding p, d, q values, these parameter values are recorded for each of the driving forces for each city and will be used to predict the values of corresponding driving forces from 2020 to 2030. The script is explained in Algorithm 1.

Algorithm 1 Automated script to test 128 ARIMA cases

```
min_n ← ∞
p_n ← 0, d_n ← 0, q_n ← 0

loop (d in 1 to 2; p in 0 to 7; q in 0 to 7):
Compute RMSE for the ARIMA model with the given p,d,q
if RMSE < min_n then
    min_n ← RMSE
    p_n ← p, d_n ← d, q_n ← q
end if

return min_n, p_n, d_n, q_n
```

The result of the script displays the specific values for the ARIMA parameters and is shown in Table 2. There are two exceptions though. Firstly, if the lowest RMSE is achieved with both p and q equal to 0, then the parameters that produced the second lowest RMSE are used. Secondly, a set of p, d, q values that worked for the training set might be invalid when applying it to the entire time series to predict future data. In this circumstance, the parameters that produced the lowest RMSE and are valid are used for later implementation. Both cases are identified in the table with an * at the end of the parameter values.

	Population	GDP	Energy Consumption	GFI
Beijing	(3,1,0)	(1,1,4)	(1,1,1)	(0,1,1)*
Chongqing	(5,1,0)	(5,1,0)	(3,1,2)	(6,1,2)
Shanghai	(2,1,0)	(0,1,7)	(3,1,6)*	(0,1,2)

Table 2: The ARIMA parameters used for each of the variables in different cities

To ensure that the ARIMA model with the lowest RMSE is viable in predicting a certain time series, each RMSE value is calculated as a percentage of the mean of the testing set of the corresponding driving force. The result is shown in Table 3. The error percentages are all very low and thus prove that using an ARIMA model to predict the values of driving forces is viable.

	Population	GDP	Energy Consumption	GFI
Beijing	0.17%	0.90%	0.83%	4.10%
Chongqing	0.24%	0.34%	0.42%	4.71%
Shanghai	0.27%	6.62%	1.19%	1.00%

Table 3: RMSE as a percentage of the mean of testing set

3.4 Predicting Carbon Emissions using BPNN

3.4.1 BPNN Introduction

BPNN is a multilayer feedforward neural network known for its backpropagation process, which means that the model adjusts the weights of a neural network based on information, specifically the error rate, obtained in the previous training epoch. Figure 4 shows the basic concept behind the BPNN network used for this research. The inputs are collected from the preconnected path. Each input is then modeled by the hidden layers using different weights that are usually randomly selected. As the input reaches the output layer, the output is calculated from the previous layer's output multiplied by the weight, which allows the network to backtrack to previous layers. The gradient of the loss function is calculated by taking the derivative of the loss function by weight and the value of the previous layers are calculated through backpropagation to reduce loss. The process is repeated until the previous sum is updated. In essence, the backpropagation process increases the number of correct output nodes and loss is reduced.

In our study, specifically, the network contains three layers: the input layer includes four nodes, which represent GDP, population, energy consumption, and GFI; the hidden layer has 15 nodes; and the output layer has one node which stands for carbon emissions volume.

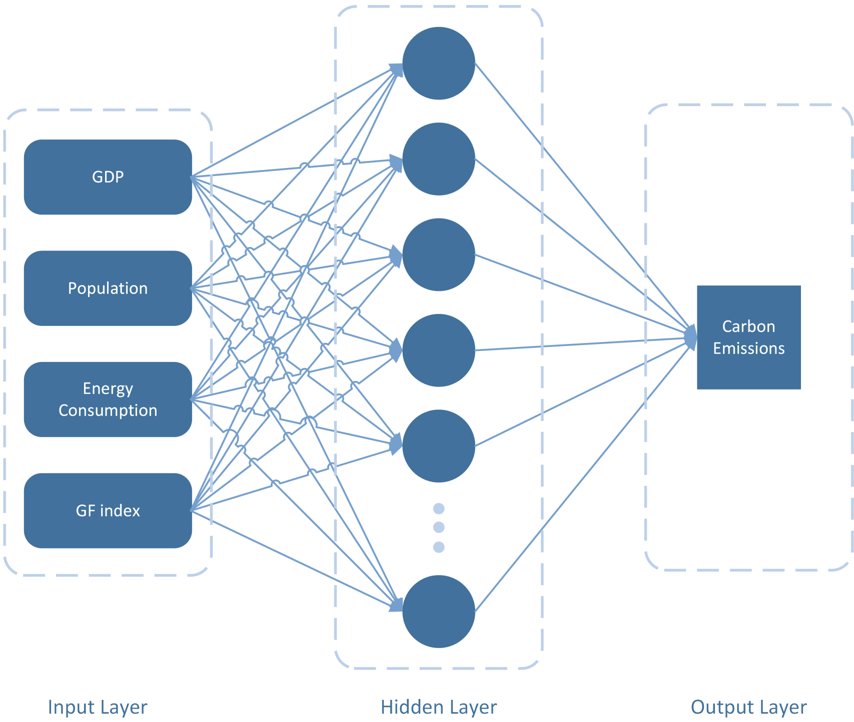


Figure 4: The BPNN network process for this research

3.4.2 BPNN Setup For This Research

In this study, the complete dataset only includes 19 years of carbon emission and driving forces data. To obtain more training data for the neural network, this research applies linear interpolation to expand the 19 data points (yearly) to 216 data points (monthly).

This research specifically utilized the BPNN network provided by Matlab. The overall network process is shown in Figure 15 in Appendix A. The network is applied to each city to estimate carbon emissions data from the predicted time series data of the four driving forces generated from the ARIMA models. To ensure that BPNN can predict the carbon emissions data accurately, the 19-year time series is also split into train and test data in a 5:1 ratio. The training set has data from 2001 to 2016 (months 1–180) while the testing set has data from 2016 to 2019 (months 181–216). The network training parameters are the same for each city and are outlined in Table 4.

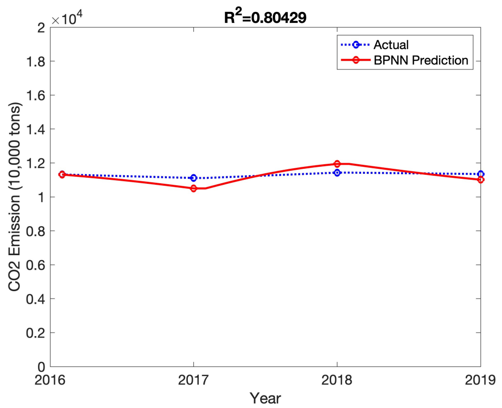
Size of hidden layers	Epochs	Target MSE	Learning Rate
15	10000	1e-8	0.001

Table 4: The BPNN parameters used

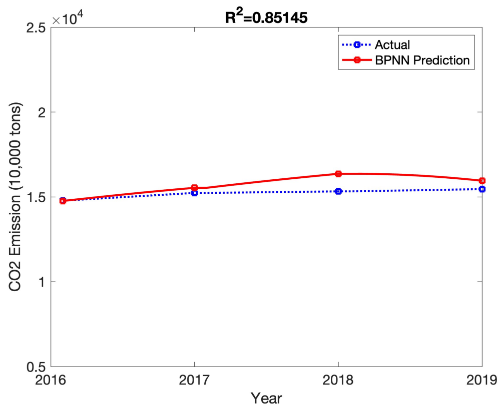
The validity and generalization performance of the trained model is tested with the relative error of the predictions in the three years of test data. The result is summarized in Table 5. The detailed comparison between the actual value and the predicted values for all three cities is shown in Figure 5. These results indicate that the trend of the predicted values for all three cities is consistent with the actual values and the relative errors are all very low. This proves that Population, GDP, Energy Consumption, and GFI are driving forces of carbon emissions and that BPNN has excellent generalization performance.

	Beijing	Chongqing	Shanghai
Maximum Relative Error	0.0554	0.0676	0.0437
Average Relative Error	0.0247	0.0364	0.0230

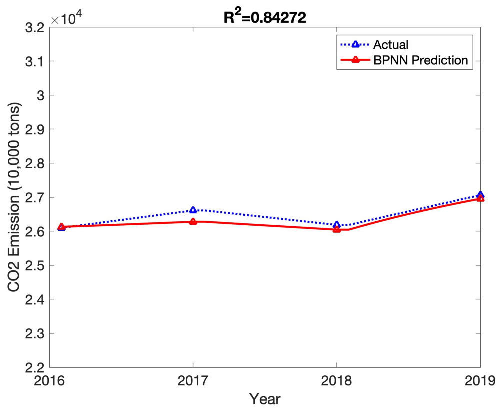
Table 5: A summary of the relative error of BPNN for all three cities



(a) Beijing



(b) Chongqing



(c) Shanghai

Figure 5: BPNN test data

The trained BPNN models for each city are later used with the predicted values of the four driving forces in 2020–2030 from the ARIMA models to predict the carbon emissions amount in 2020–2030. This 11 years of predicted data points from ARIMA also undergoes linear interpolation to expand into monthly data points before being fed to the BPNN models. To determine the effect of the GFI on carbon emissions, the prediction is conducted under 5 different conditions, depicted in Table 6.

Condition	Description
1	Use the predicted values of the driving forces without any change
2	All predicted GFI values are multiplied by 1.05
3	All predicted GFI values are multiplied by 1.10
4	All predicted GFI values are multiplied by 1.15
5	All predicted GFI values are multiplied by 1.20

Table 6: The different conditions of BPNN prediction

3.5 Cost Analysis

In this section, the cost model proposed by Zagheni and Billari is used to analyze the effect of initial carbon emission on the cost for control of carbon emissions in Beijing, Chongqing, and Shanghai.

3.5.1 Cost Factors

This cost model includes several parameters that need to be solved for, such as population and GDP. The population's rate of increase is expressed with Equation 4, which is essential in that it calculates the value of ρ and P_m, both are used in later stages of cost analysis. P represents the population at a certain time t. The left-hand side of Equation 4 represents the population's rate of increase based on the current population size, which could be calculated using linear regression with the population's growth rate at time t on the y-axis and the population size at time t on the x-axis. The function resulting from the linear regression can be expressed using Equation 5, and we can then backtrack since $\rho = a_1$ and $P_m = \frac{a_1}{a_2}$.

$$\frac{dP}{Pdt} = \rho(1 - \frac{P}{P_m}) \tag{4}$$

$$\frac{dP}{Pdt} = a_1 - a_2 P \tag{5}$$

This population equation was developed in 1994 by Dietz and Rosa as a stochastic counterpart of the IPAT and ImPACT identities. Moreover, Equation 6 was proposed to model the relation between carbon emissions and population, GDP, and energy consumption, and can be simplified to Equation 7. In 2007, Zagheni and Billari further refined the model into Equation 8.

$$I = aP^b A^c T^d e \tag{6}$$

$$\ln(I) = \ln(a) + b\ln(P) + c\ln(A) + d\ln(T) + \ln(e) \tag{7}$$

$$\ln(I) = \ln(a) + (b - c)\ln(P) + c\ln(PA) + d\ln(T) + \ln(e) \tag{8}$$

Equation 8 is fitted with data regarding carbon emissions, population, and GDP. The `lsqlin` function from Matlab solves constrained linear least square problems and returns values b and c, both are later used in the final cost analysis.

This study also analyzes a fictitious scenario based on the US's INDC proposed in 2015 IND (2015). For the following cost evaluation, each region is assumed to be required to import emission credits based on market price if they fail to lower carbon emissions by 25% of 2019 levels by 2030.

Given the above setup, the potential cost related to carbon emissions for any country can be set to $C(I,t)$, with I being the amount of carbon emissions as the underlying asset and t being the year. This function can be seen as a European option with carbon emissions as an asset. Further assuming that the risk-free rate is constant r, the cost function can be expressed as Equation 9, where $\bar{I}$ represents the carbon emission threshold (75% of a region's carbon emissions value in 2019).

$$C(I,t) = E(\alpha(I_T - \bar{I})e^{-r(T-t)}|I_{t0} = I_0 \tag{9}$$

Further combining Equation 9 with the Black-Scholes equation, Equation 10 can be achieved. $N(x)$ is a standard normal distribution accumulation function expressed in Equation 11 with $d_1(t)$ and $d_2(t)$ expressed in Equation 12 and Equation 13, respectively.

$$C(I,t) = \alpha I[\frac{\rho_0(\frac{P_m - P_0}{P_0})e^{-\rho_0 t} + 1}{(\frac{P}{P_m} - 1)e^{-\rho_o T} + 1}]^b e^{(c\mu - r)(T-t)} N(d_1(t)) - \alpha \bar{I} e^{-r(T-t)} N(d_2(t)) \tag{10}$$

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{w^2}{2}} dw \tag{11}$$

$$d_1(t) = \frac{\ln(I) - \ln(\bar{I}) + b[\ln(\frac{P_m - P_0}{P_0} e^{-\rho_0 t} + 1) - \ln(\frac{P_m - P_0}{P_0} e^{-\rho_0 T} + 1)] + c\mu(T-t) + \frac{1}{2}c^2\mu^2(T-t)}{c\sigma\sqrt{T-t}} \tag{12}$$

$$d_2(t) = d_1(t) - c\sigma\sqrt{T-t} \tag{13}$$

3.5.2 Marginal Cost

This study analyzes four aspects of marginal cost:

1. The cost at the initial moment ($t = 0$) with different initial carbon emissions I. Expressed as $C(I, 0)$.

2. The relative change in marginal cost at the initial moment based on the change in initial carbon emissions I.

$$\frac{\partial C(I,t)}{\partial I} = \alpha \left[\frac{\left(\frac{P_m - P_0}{P_0}\right) e^{-\rho_0 t} + 1}{\left(\frac{P_m - P_0}{P_0}\right) e^{-\rho_0 T} + 1} \right]^b e^{(c\mu - r)(T-t)} N\left(d_1(t)\right) \geq 0 \tag{14}$$

3. The relative change in the cost at the initial moment based on the change in volatility, represented by σ.

$$\frac{\partial C(I,t)}{\partial \sigma} = \alpha I e^{-r(T-t) - \frac{1}{2}N\left(d_2(t)^2\right)} b\sqrt{T-t} > 0 \tag{15}$$

4. The cost at time t with initial carbon emissions I. Expressed as $C(I,t)$.

All four aspects are analyzed separately based on each of the three cities (Beijing, Chongqing, Shanghai) and graphed to show the result.

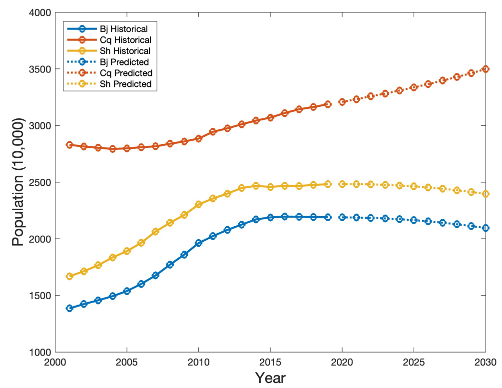
4 Result

This section analysis the results from different algorithms together and outlines policy implications.

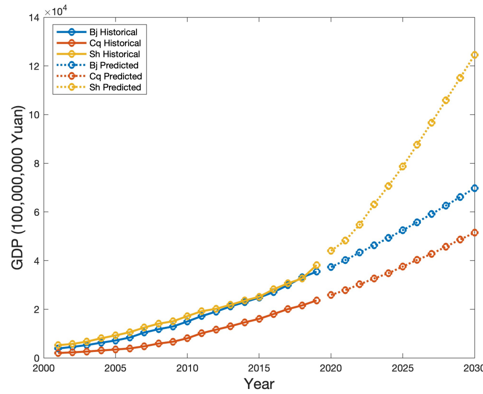
4.1 Driving Forces

The values that the ARIMA model predicted for all four driving forces from Beijing, Chongqing, and Shanghai in 2020–2030 are visualized alongside historical values in Figure 6. Specific values are summarized in Appendix A in Table 8, Table 9, and Table 10, respectively.

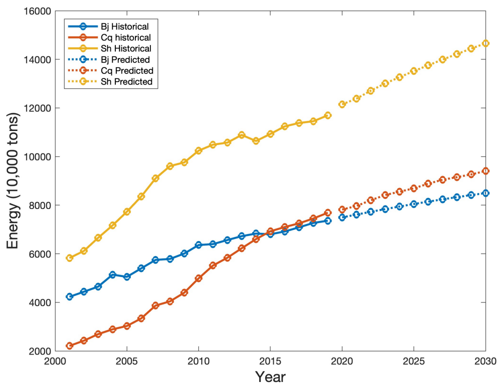
One thing to note here is that the GFI prediction for Beijing exceeds 1 for the years 2023 and later, which is impossible in real life since GFI is a value between 0 and 1. Therefore, the values for 2023 and later are adjusted to 1.00 when used with BPNN to predict carbon emissions.



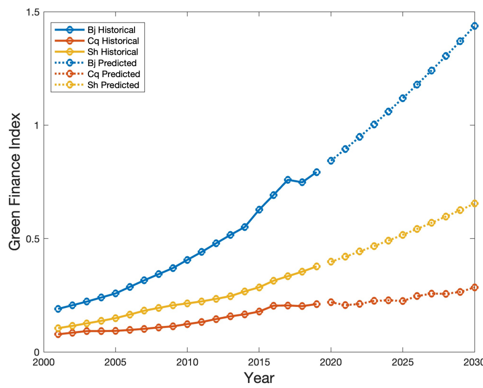
(a) Population of Beijing, Chongqing, and Shanghai from 2001 to 2030 (historical and predicted values)



(b) GDP of Beijing, Chongqing, and Shanghai from 2001 to 2030 (historical and predicted values)



(c) Energy consumption of Beijing, Chongqing, and Shanghai from 2001 to 2030 (historical and predicted values)



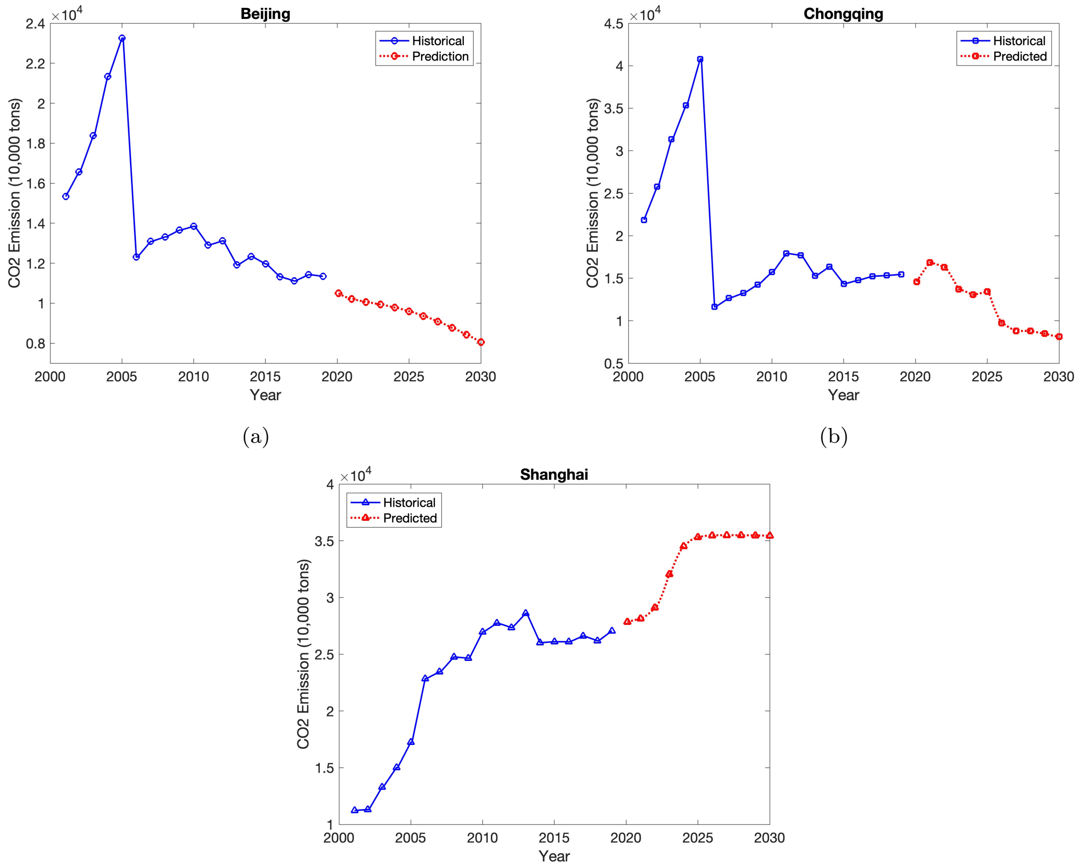
(d) GFI of Beijing, Chongqing, and Shanghai from 2001 to 2030 (historical and predicted values)

Figure 6: Visualization of ARIMA prediction values

4.2 Carbon Emissions

The predicted carbon emissions values of Beijing, Chongqing, and Shanghai using the predicted values of the four driving forces and pre-trained BPNN model are shown in Figure 7. The values from 2001 to 2019 are past values and are shown in blue. The values from 2020 to 2030 are predicted values and are shown in red.

The predicted result shows that both Beijing's and Chongqing's carbon emission values display a decreasing trend up to 2030, consistent with the goal of reaching carbon peak in 2030. However, Shanghai's predicted carbon emission values in general display an increasing trend up to 2030, inconsistent with the goal.



(a) (b) (c)

Figure 7: All three cities' carbon emissions from 2001 to 2030 (including both past and predicted values)

Green Finance Manipulations The predicted carbon emissions values of Chongqing and Shanghai including conditions with the GFI values modified are shown in Figure 8 and Figure 9. The GFI modifications are not applied to Beijing because many of the predicted GFI values of Beijing already exceed 1, which eliminates room for increasing the GFI values.

In both Shanghai and Chongqing, higher GFI values result in lower predicted carbon emissions. Figure 10 and Figure 11 show the significant decrease in predicted carbon emissions when the GFI is multiplied by 1.2. Shanghai's carbon emissions trend even changes from increasing to decreasing when GFI values are multiplied by greater than or equal to 1.15. This proves that the GFI greatly and negatively impacts carbon emissions. However, in Chongqing, the effect of a higher GFI diminishes over time, where the predicted carbon emissions values generally converge from 2027 to 2030. We suspect that this result is due to a diminishing return of high GFI values in some cases.

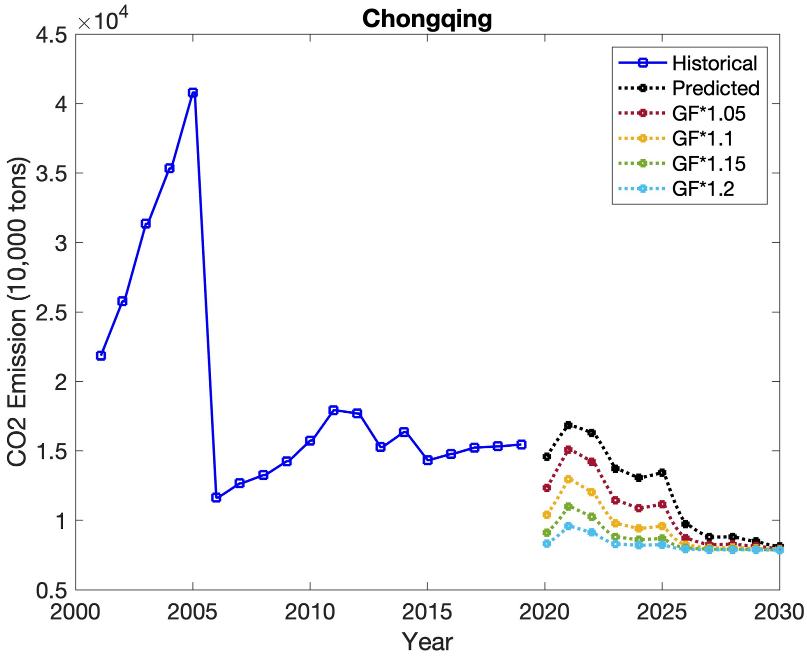


Figure 8: Chongqing's carbon emissions from 2001 to 2030 (including past values and all 6 prediction conditions)

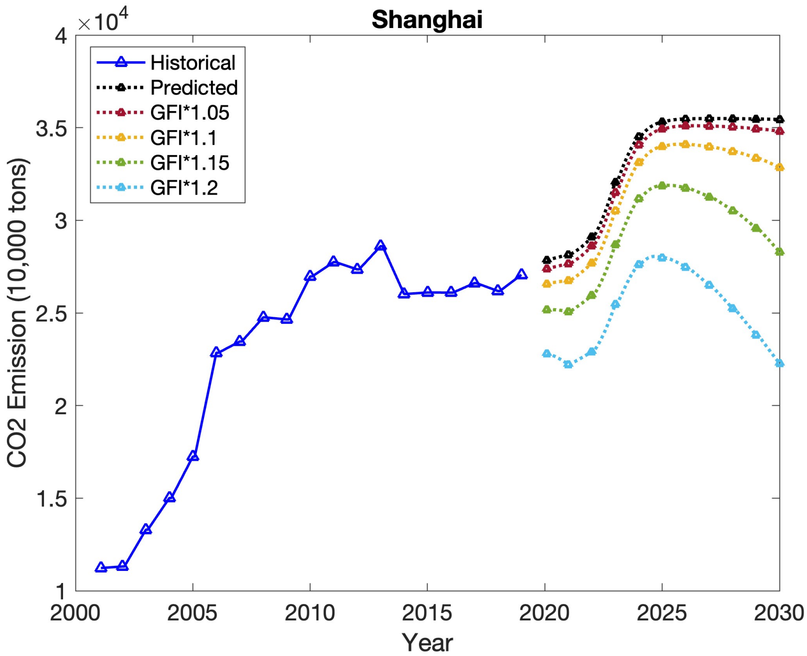


Figure 9: Shanghai's carbon emissions from 2001 to 2030 (including past values and all 6 prediction conditions)

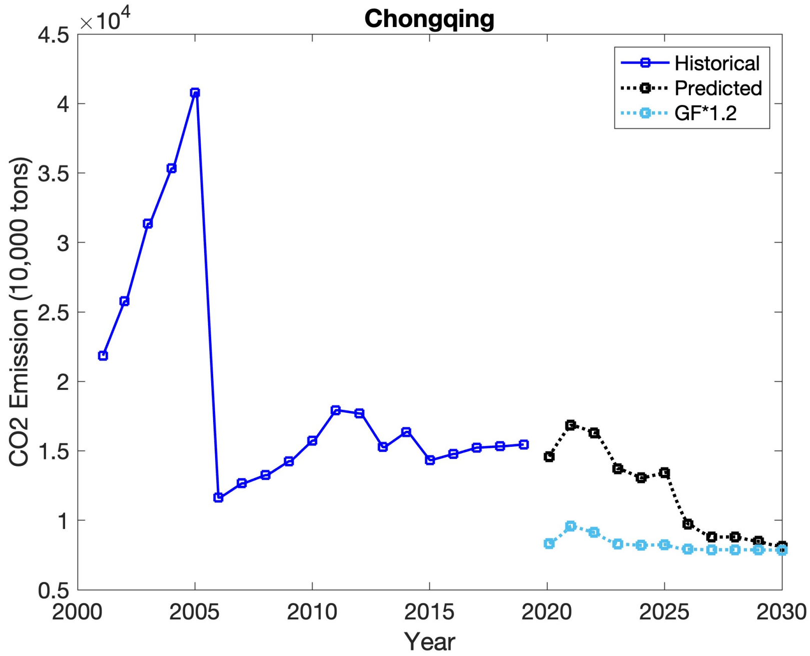


Figure 10: Chongqing's carbon emissions from 2001 to 2030 (including past values and all 6 prediction conditions)

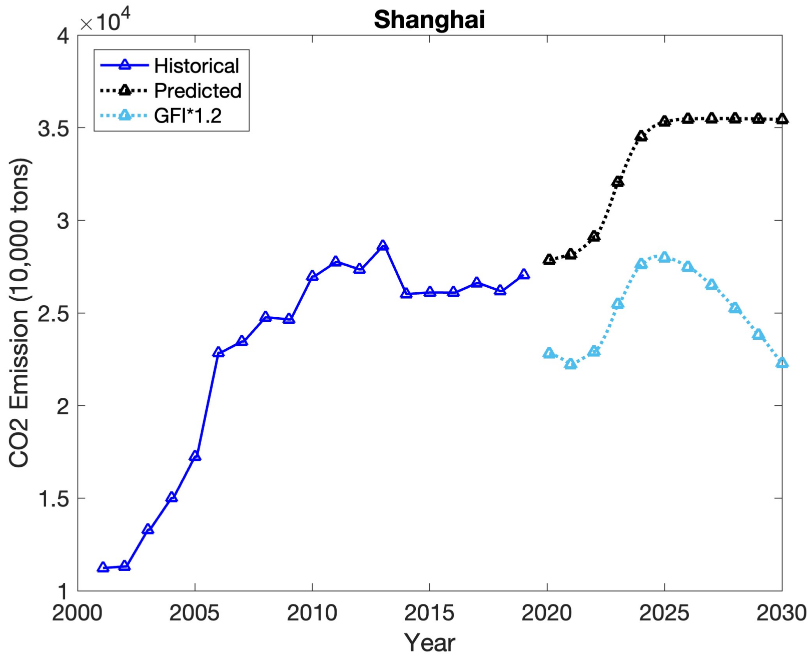


Figure 11: Shanghai's carbon emissions from 2001 to 2030 (including past values and all 6 prediction conditions)

4.3 Cost Analysis Result

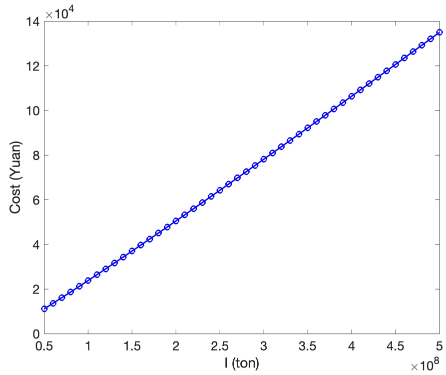
As outlined in the Methodology section, there are some equations involved in finding the value of variables in preparation for the final cost analysis. The values of the variables for each city are displayed in Table 7.

	rho	P0	Pm	mu	sigma1	sigma2
Beijing	0.0944	1.122e7	2.554e7	0.19033	0.0228	0.1439
Chongqing	-0.0628	3.098e7	2.647e7	0.18841	0.0069	0.101
Shanghai	0.1151	1.327e7	2.697e7	0.15827	0.0124	0.1105

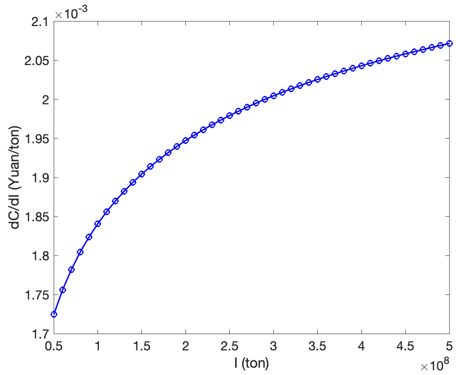
	a	b	I	$\bar{I}$
Beijing	-0.0551	15.7455	15334.14	8.5097e3
Chongqing	-0.562	-31.8888	21851.78	1.1593e4
Shanghai	-0.2516	-11.6013	11229.32	2.0295e4

Table 7: Cost model related variables

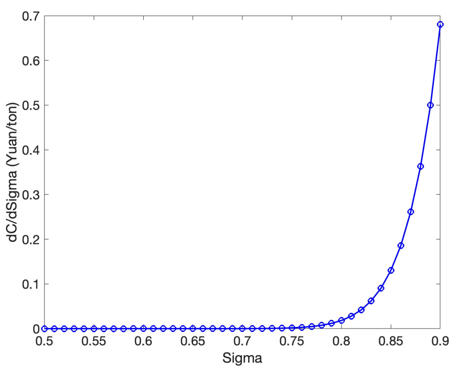
These values are used to generate the final cost analysis graphs for Beijing, Chongqing, and Shanghai, corresponding to Figure 12, Figure 13, and Figure 14, respectively. We observe that for all three cities, initial carbon emissions I is positively correlated with the cost and relative change in marginal cost, and σ is positively correlated with the relative change in the cost. In terms of time, Beijing's predicted marginal cost is negatively correlated with time T while Chongqing and Shanghai's predicted marginal cost is positively correlated with time T.



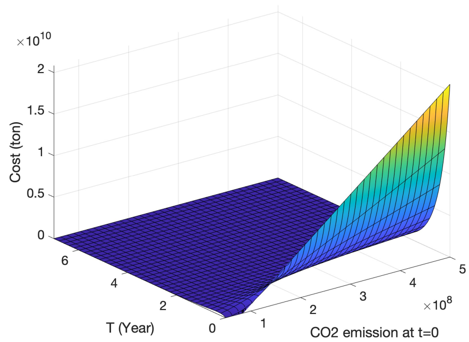
(a) The cost based on different initial carbon emissions I



(b) Relative change in marginal cost based on change in initial carbon emissions I

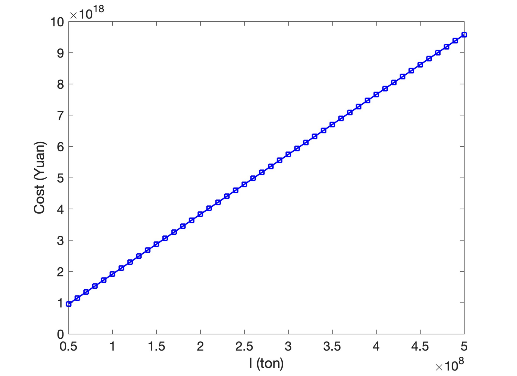


(c) Relative change in the cost based on change in σ

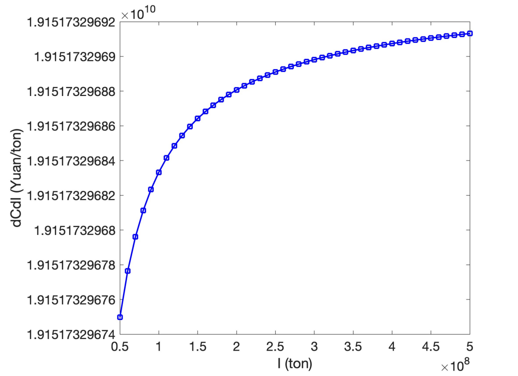


(d) The cost at time T with initial carbon emissions I

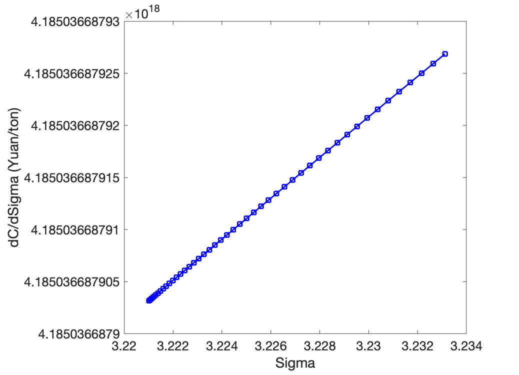
Figure 12: Cost analysis result for Beijing



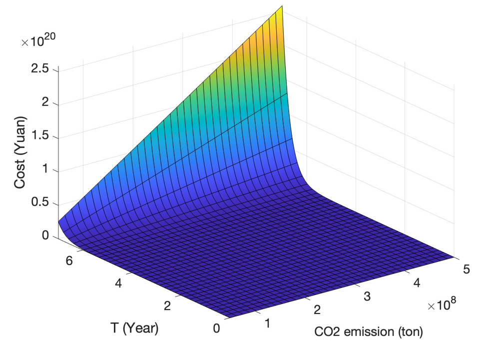
(a) The cost based on different initial carbon emissions I



(b) Relative change in marginal cost based on change in initial carbon emissions I

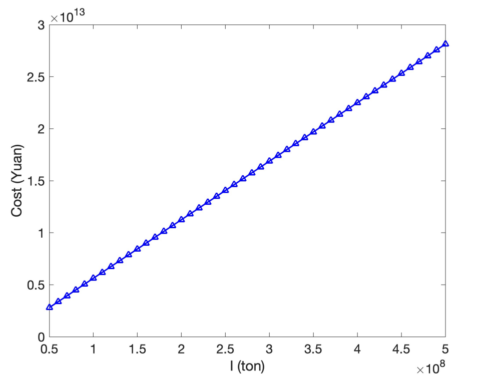


(c) Relative change in the cost based on change in σ

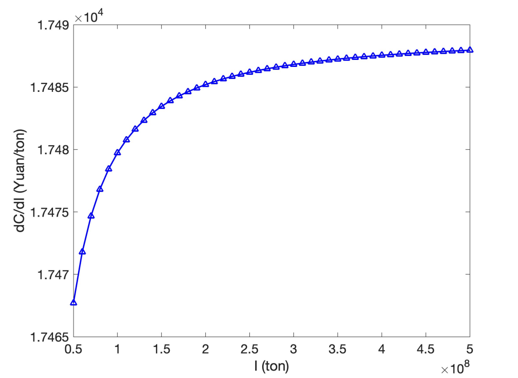


(d) The cost at time T with initial carbon emissions I

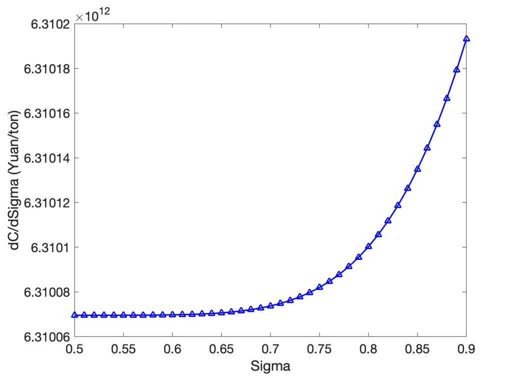
Figure 13: Cost analysis result for Chongqing



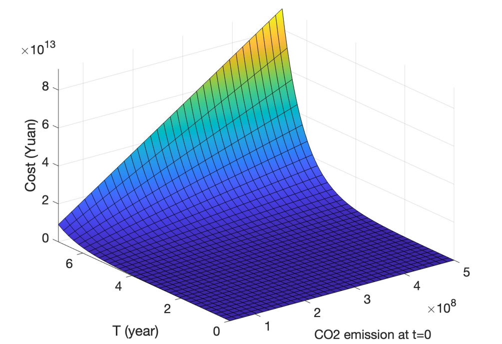
(a) The cost based on different initial carbon emissions I



(b) Relative change in marginal cost based on change in initial carbon emissions I



(c) Relative change in the cost based on change in σ



(d) The cost at time T with initial carbon emissions I

Figure 14: Cost analysis result for Shanghai

5 Conclusion

This paper explored the driving forces of carbon emissions and the prediction of future carbon emissions from the perspective of governments, specifically focusing on Beijing, Chongqing, and Shanghai as three case studies. It also identifies the cost related to excessive carbon emissions.

This study performed time series prediction using ARIMA; driving force analysis and carbon emissions prediction using BPNN; and cost analysis using equations derived from the Feynman-Kac and Black-Scholes equations. All calculations and models are performed in Matlab.

The main findings of this research can be divided into three parts.

First, our training of BPNN shows that using the four driving forces (Population, GDP, Energy Consumption, and GFI) to predict carbon emissions is viable and accurate. Our analysis further shows that GFI has a great and negative impact on carbon emissions, meaning that a higher GFI leads to significantly lower carbon emissions. However, this impact could also follow a diminishing returns curve in some circumstances where differences in higher GFI values display a small effect on carbon emissions.

Second, the carbon emissions data from 2020–2030 as predicted by our trained BPNN models provide insight into the probability of China reaching carbon peak in 2030. The prediction results from both Beijing and Chongqing show a decreasing trend while the result from Shanghai shows an increasing trend.

Third, our cost model results can act as a useful reference for the marginal cost of excessive carbon emissions in the fictitious scenario where the target is set to limit carbon emissions to 75% of 2019 levels by 2030.

This paper provides a comprehensive analysis and prediction of the driving forces and costs related to carbon emissions but has limitations in several areas. Firstly, due to the quantities of publicly available data, this study only took into account the years between 2001–2019 as past data, and only three cities are considered. Future research should expand the quantity of historical data by obtaining data from the 20th century and recent years, as well as expand the number of locations studied to many more cities or even provinces. Future research could also dive into corporate carbon emissions and incentives for corporates to decrease carbon emissions, such as carbon tax and carbon trading. Secondly, this research assumes the hypothetical condition where each city in China must pay the excessive emissions at a market price if it doesn't lower its carbon emissions to 75% of 2019 levels by 2030. Although this is a practical case, future research could look into trade agreements or limitations between China and other countries that have certain carbon emission requirements and calculate the cost of excessive carbon emissions based on these actual trade agreements.

References

(2015). FACT SHEET: U.S. Reports its 2025 Emissions Target to the UNFCCC.

(2021). China's achievements, new goals and new measures for nationally determined contributions. UNFCCC NDC Registry.

Amarpuri, L., Yadav, N., Kumar, G., and Agrawal, S. (2019). Prediction of co2 emissions using deep learning hybrid approach: A case study in indian context. *IEEE*.

beijing municipal bureau of statistics. .

Cao, L. (2023). How green finance reduces co2 emissions for green economic recovery: empirical evidence from e7 economies. *Environmental Science and Pollution Research*, 30:3307–3320.

Chen, X., Chen, Z., Sharifi, A., Feng, C., and Yang, J. (2021). Can green finance development reduce carbon emissions? empirical evidence from 30 chinese provinces.

chongqing municipal bureau of statistics. .

CSO. Net-Zero Emissions Operations by 2050, including a 65

Dietz, T. and Rosa, E. A. (1994). Rethinking the environmental impacts of population, affluence and technology.

EU. 2050 long-term strategy.

Fang, Z., Yang, C., and Song, X. (2022). How do green finance and energy efficiency mitigate carbon emissions without reducing economic growth in g7 countries? *Frontiers in Psychology*, 13.

Ghoddusi, H., Creamer, G. G., and Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, 81:709–727.

Guo, C. Q., Wang, X., Cao, D. D., and Hou, Y. G. (2022). The impact of green finance on carbon emission–analysis based on mediation effect and spatial effect. *Frontiers in Environmental Science*, 10.

Hou, D., Zhang, H., Li, L., Wang, X., Lin, Y., and Du, H. (2021). Deep learning-based comparative modeling of carbon emissions projections. pages 241–251. Institute of Electrical and Electronics Engineers Inc.

Ip, Y., Iqbal, W., Du, L., and Akhtar, N. (2023). Assessing the impact of green finance and urbanization on the tourism industry—an empirical study in china. *Environmental Science and Pollution Research*, 30:3576–3592.

Li, Z., Wu, L., Zhang, Z., Chen, R., Jiang, Y., Peng, Y., Zheng, K., and Jiang, W. (2022). The transformative impacts of green finance governance on construction-related co2 emissions. *Sustainability (Switzerland)*, 14.

national bureau of statistics. - .

Nguyen, Q., Diaz-Rainey, I., and Kuruppuarachchi, D. (2021). Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. *Energy Economics*, 95.

Qin, J., Zhao, Y., and Xia, L. (2018). Carbon emission reduction with capital constraint under greening financing and cost sharing contract. *International Journal of Environmental Research and Public Health*, 15.

Ritchie, H. (2019). Who emits the most co2 today? - our world in data.

Ritchie, H. and Roser, M. (2020). China: Co2 country profile - our world in data.

Sarfraz, M., Mohsin, M., Naseem, S., and Kumar, A. (2021). Modeling the relationship between carbon emissions and environmental sustainability during covid-19: a new evidence from asymmetric ardl cointegration approach. *Environment, Development and Sustainability*, 23:16208–16226.

Serafeim, G. and Caicedo, G. V. (2022). Machine learning models for prediction of scope 3 carbon emissions.

shanghai municipal bureau of statistics. .

Sun, C. (2022). The correlation between green finance and carbon emissions based on improved neural network. *Neural Computing and Applications*, 34:12399–12413.

Tay, A. (2022). By the numbers: China's net-zero ambitions.

UNFCCC. The Paris Agreement.

USEPA (2023). Understanding global warming potentials — US EPA.

Xiong, Q. and Sun, D. (2022). Influence analysis of green finance development impact on carbon emissions: an exploratory study based on fsqca. *Environmental Science and Pollution Research*.

Xu, L., Wu, C., Qin, Q., and Lin, X. (2022). Spillover effects and nonlinear correlations between carbon emissions and stock markets: An empirical analysis of china's carbon-intensive industries. *Energy Economics*, 111.

Yao, J. and Zhao, Y. (2022). Structural breaks in carbon emissions: A machine learning analysis, wp/22/9, january 2022.

York, R., Rosa, E. A., and Dietz, T. (2003). Stirpat, ipat and impact: Analytic tools for unpacking the driving forces of environmental impacts. *Ecological Economics*, 46:351–365.

Zagheni, E. and Billari, F. C. (2007). A cost valuation model based on a stochastic representation of the ipat equation. *Population and Environment*, 29:68–82.

Zhang, F. and Wen, N. (2022). Carbon price forecasting: a novel deep learning approach. *Environmental Science and Pollution Research*, 29:54782–54795.

Zhang, W., Zhu, Z., Liu, X., and Cheng, J. (2022a). Can green finance improve carbon emission efficiency? *Environmental Science and Pollution Research*, 29:68976–68989.

Zhang, Z., Liu, Y., Han, Z., and Liao, X. (2022b). Green finance and carbon emission reduction: A bibliometric analysis and systematic review. *Frontiers in Environmental Science*, 10.

Zhou, Y., Zhang, J., and Hu, S. (2021). Regression analysis and driving force model building of co2 emissions in china. *Scientific Reports*, 11.

Appendix A Figures and Tables

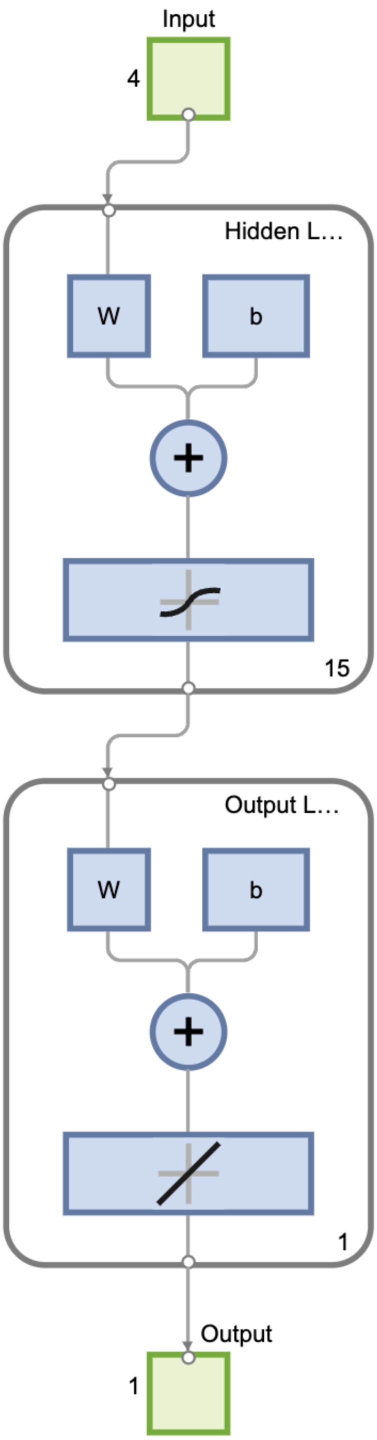


Figure 15: Overall Process of BPNN network in Matlab

Beijing				
Year	Population	GDP	Energy Consumption	GFI
2020	2188.7	37314	7497.2	0.84304
2021	2187.1	40188	7610.8	0.89478
2022	2184	43325	7727.6	0.94822
2023	2179.3	46247	7837.5	1.0034
2024	2172.7	49328	7944.3	1.0602
2025	2164.2	52495	8046.7	1.1188
2026	2153.9	55761	8145.1	1.179
2027	2141.8	59125	8239.3	1.241
2028	2127.8	62586	8329.5	1.3046
2029	2112.1	66145	8415.5	1.37
2030	2094.6	69801	8497.4	1.437

Table 8: The values of each driving force in Beijing from 2020 to 2030 as predicted by the ARIMA model.

Chongqing				
Year	Population	GDP	Energy Consumption	GFI
2020	3208.9	25879	7820	0.21917
2021	3231.8	27855	7965.8	0.20647
2022	3258.7	30239	8202	0.2115
2023	3281.4	32637	8418.2	0.22514
2024	3308.7	34800	8552.4	0.2276
2025	3336.9	37541	8693.6	0.22487
2026	3364.7	40282	8883.3	0.24615
2027	3396.9	42734	9046	0.25725
2028	3428.2	45707	9152.7	0.25655
2029	3461.7	48676	9265.7	0.26467
2030	3497.8	51475	9403.9	0.28484

Table 9: The values of each driving force in Chongqing from 2020 to 2030 as predicted by the ARIMA model.

Shanghai				
Year	Population	GDP	Energy Consumption	GFI
2020	2481.7	43989	12148	0.39767
2021	2481.5	48288	12383	0.41989
2022	2480.3	54779	12711	0.44289
2023	2476.3	63072	13009	0.46664
2024	2470.4	70678	13265	0.49116
2025	2462.8	78685	13518	0.51645
2026	2453.3	87619	13754	0.5425
2027	2441.7	96675	13992	0.56932
2028	2428.2	105850	14217	0.5969
2029	2412.8	115150	14444	0.62524
2030	2395.5	124560	14660	0.65435

Table 10: The values of each driving force in Shanghai from 2020 to 2030 as predicted by the ARIMA model.

Appendix B Code

All code used while conducting this research can be found on google drive (`https://drive.google.com/file/d/1pQ8q8VpO_AOSvbNTMl9v_qv7kmhzYq8U/view?usp=sharing`) and microsoft onedrive (`https://1drv.ms/u/s!Asi_plQOBA2UiBTu3tUfN5gABpAk?e=SKOS1C`).

Acknowledgement

The author is grateful for the guidance offered by Chengguang Zhu on modelling and coding during the research process. The author would also like to thank S.T. Yau High School Science Award for providing this opportunity.

Commitments on Academic Honesty and Integrity

We hereby declare that we

1. are fully committed to the principle of honesty, integrity and fair play throughout the competition.
2. actually perform the research work ourselves and thus truly understand the content of the work.
3. observe the common standard of academic integrity adopted by most journals and degree theses.
4. have declared all the assistance and contribution we have received from any personnel, agency, institution, etc. for the research work.
5. undertake to avoid getting in touch with assessment panel members in a way that may lead to direct or indirect conflict of interest.
6. undertake to avoid any interaction with assessment panel members that would undermine the neutrality of the panel member and fairness of the assessment process.
7. observe the safety regulations of the laboratory(ies) where we conduct the experiment(s), if applicable.
8. observe all rules and regulations of the competition.
9. agree that the decision of YHSA is final in all matters related to the competition.

We understand and agree that failure to honour the above commitments may lead to disqualification from the competition and/or removal of reward, if applicable; that any unethical deeds, if found, will be disclosed to the school principal of team member(s) and relevant parties if deemed necessary; and that the decision of YHSA is final and no appeal will be accepted.

(Signatures of full team below)

X Junting Wang

Name of team member:

X Chengguang Zhu

Name of supervising teacher:

Declaration of Academic Integrity

The participating team declares that the paper submitted is comprised of original research and results obtained under the guidance of the instructor. To the team's best knowledge, the paper does not contain research results, published or not, from a person who is not a team member, except for the content listed in the references and the acknowledgment. If there is any misinformation, we are willing to take all the related responsibilities.

Names of team members: Junting Wang

Signatures of team members

Junting Wang

Name of the instructor: Chengguang Zhu

Signature of the instructor

Chengguang Zhu

Date
August 19, 2023