Integrated Time Series Analysis and Short-Term Forecasting On Carbon Credit Trading In China

Ricardo K WU[1] and Chun Man CHAN[2]

[1]Department of Systems Engineering and Engineering Management, CUHK
[2]Department of Statistics, CUHK

1 Background

In this report, we are playing the role of a Chinese traditional factory. It is predicted that the carbon quota for our factory will be run off within 3 months, so we are facing a dilemma at the moment: 1. Purchase credit now 2. Purchase credit within 3 months. A time series analysis and forecasting are needed for us to reduce our cost by referencing the past 7 years of data.

1.1 Environmental crisis

Environmental issue has always been the talk of the town. Global warming and extreme climates are kept be spoken out; no matter which social class they come from. Look from the statistics, billions of tons of CO2 are released into the atmosphere every year as a result of coal, oil, and gas production. Greenhouse gas emissions are vent from human activities at a record high. It is believed that carbon emission plays a major role in aggregating global warming and has caused catastrophic destruction. Though countries have held conferences on the limitation of carbon emissions, the tangible effect remains to be seen. For energy-intensive industries, they need mitigating conditions for transition, some may even find it hard to get rid of emissions exceeding their carbon quota due to their business nature. Carbon credit trading is a voluntary exercise for companies to participate in.

1.2 Situation in China

Starting in 2013, China started testing in seven cities, such as Guangdong, Beijing, and Shanghai. After testing for a few years, a national carbon emission trading market under the carbon trading scheme was finally initiated in 2021. Since China is the biggest polluter in the world, the launch of an emission trading system has doubled the proportion of trading carbon emissions in the world. This action also is an important step for realizing China government's goal, which is to become carbon neutral before 2060.

In Feb 2021, China apply the scheme to 2225 power firms. As the market in China is large, the carbon emission in China has roughly 5 billion tons of carbon credit and around 8 thousand companies joined this scheme.

Guangzhou emissions exchange is the only platform for carbon emissions trading and Chine Certified Emission Reduction in Guangdong. Until the beginning of December 2022, the total turnover of the Guangzhou emissions exchange had exceeded 200 million.

1.3 New economic opportunities

Every country had its own quota for carbon emissions, which is set by the government in accordance with the emissions target committed under the international convention. Then the quota will be assigned to different companies. Each quota represents that a company could emit a ton of greenhouse gas. The company can reduce the emissions of carbon dioxide by introducing relevant technology, then they may sell the excessive quota to other companies through the carbon credit trading market. On the contrary, those firms which cannot reduce the emission of carbon dioxide could buy quotas from other enterprises through the carbon market. In other words, selling carbon permits is an opportunity for companies to increase their revenue. Tesla Carbon Credit Sales reached a new record in 2022, which is $1.78 billion. When the firm gains more profit, the government may earn more profit tax to increase financial reserves, then the government has more funding to relieve livelihood issues.

2 Introduction

2.1 Methods

Most conventional statistical models analyze different types of data in a linear way by linear and probabilistic statistical inference, they assume data coming from models and what they are doing is to find out the most likely parameters that make the model generate the observed data. Though the diagnostics can be extensive, they have many limitations on analyzing other types of data and the process requires direct intuitions about sampling distributions and statistical assumptions, while machine learning provides a totally different way.

The algorithmic and probabilistic way can extract patterns from messy data without much domain knowledge, it sometimes performs better than conventional methods, especially for non-linear and high dimensional or out-of-sample data.

In this report, both traditional and innovative methods be adopted to make predictions.

2.2 Methodology

The averaged monthly unit price of carbon credits data is first decomposed into traditional statistical parts consisting of trend, seasonality, and noises; the cyclic element is excluded due to the characteristic of the trend shown in the graph. Seasonal Extraction in ARIMA Time Series(SEATS), Seasonal and Trend decomposition using Loess(STL), and Multiple Seasonal-Trend decomposition using LOESS(MSTL) are adopted to help decompose the data and inspect the components of the data.

After decomposition, some naive forecasting methods like Moving averages(SMA), Centered moving averages(CMA), Weighted moving averages(WMA), and Exponential smoothing(EMA) are used for naive and simple forecasting, which are set to act as indicators for later modeling.

Besides, we use a trend projection model including linear regression and polynomial regression. In linear We first find out the coefficient of the regression line and fit the data to do the forecasting. For polynomials, we first find out the best degree of the model and then do the forecasting.

We also use Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model. We first test the ARCH effects, then select the best model and do the forecasting.

But as this paper is mainly for machine learning and deep learning, other parts will be taken off from it but leave the performance for comparison.

3 Time series decomposition

> As an emerging carbon trading market, we wonder about its performance.

3.1 Exploratory Time Series Analysis

There are 1797 daily raw data defined as $Y = \{Y_1, Y_2, \ldots, Y_{1797}\}$ collected from Guangzhou Emission Exchange in the format shown below.

Date	Type	Opening	Closing	Max	Min	%Change	Volume	Amount

As we aim to analyze the unit price per month and forecast the future trend due to missing data from the low monthly trading frequency, we do the operation to get the unit price by summing up Volume$_{day}$ and Amount$_{day}$ of the month to get the averaged price in RMB per volume defined as $\frac{Volume_{day}/Amount_{day}}{n_{month}}$. After the operation, the raw data are transformed into 96 monthly data defined as $X = \{X_1, X_2, \ldots, X_{96}\}$, the visualization of time series X and its summary are shown in Figure 1 and Figure 2

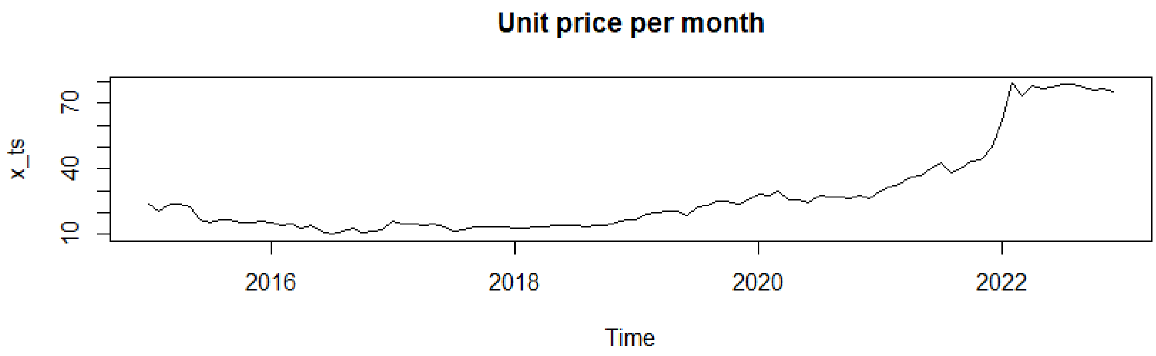


Figure 1: GDEA monthly unit price

```
     Date           Unit.price.per.month
 Length:96          Min.   :10.03
 Class :character   1st Qu.:14.21
 Mode  :character   Median :20.76
                    Mean   :27.92
                    3rd Qu.:29.68
                    Max.   :78.99
```

Figure 2: GDEA monthly unit price summary

The series seemingly doesn't show a standard linear trend but is more likely a polynomial trend. To make the prediction later more accurate and precise, the series will be decomposed in two directions under traditional statistical decomposition.

There are two models for time series decomposition, named additive and multiplicative, the formulas are shown in the formula (1) and (2) respectively. Each model suggested data at time t contains three components: Trend(T), Seasonality(S), and Error(N). Formula (2) can be transformed into the additive model by taking natural logarithms on both sides as Formula (4)

$$X_t = T_t + S_t + N_t \qquad N_t \sim N(0, \sigma^2) \tag{1}$$

$$X_t = T_t * S_t * N_t \qquad N_t \sim N(0, \sigma^2) \tag{2}$$

$$\ln(X_t) = \ln(T_t) + \ln(S_t) + \ln(N_t) \qquad N_t \sim N(0, \sigma^2) \tag{3}$$

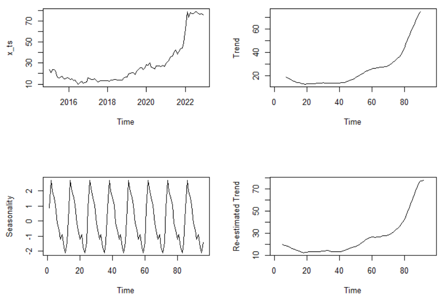


Figure 3: Decomposition under (1)

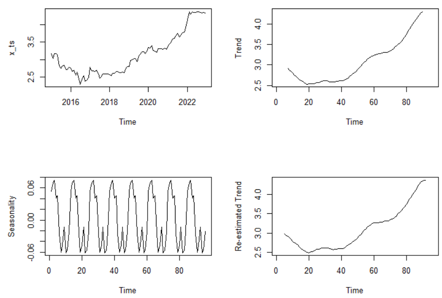


Figure 4: Decomposition under (4)

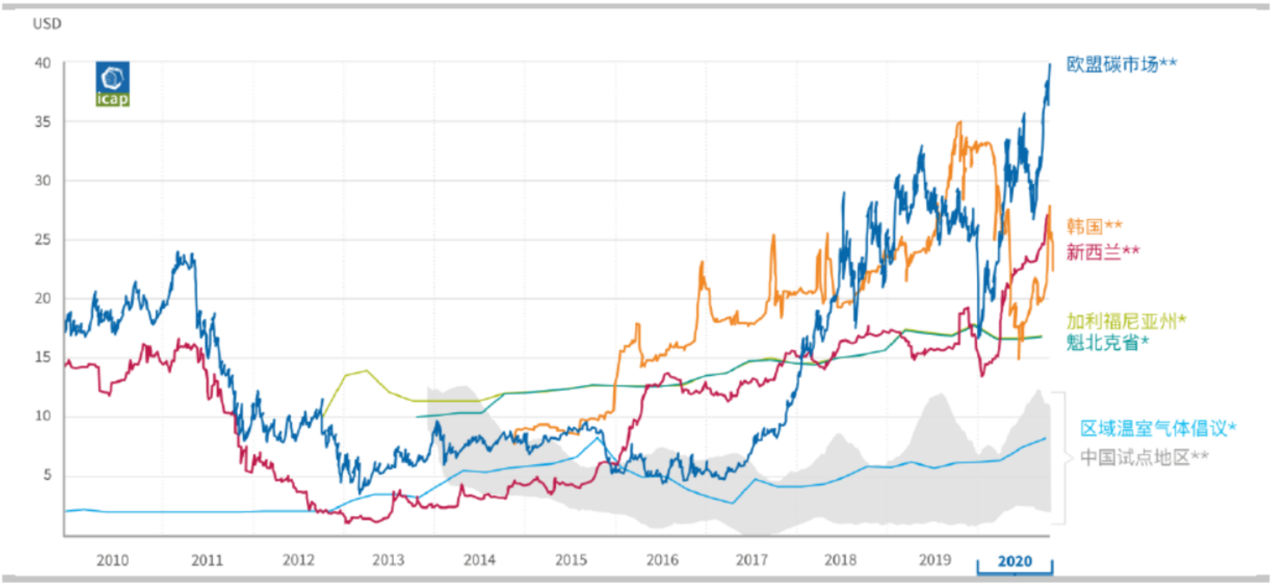
However, the graph doesn't show a strong seasonality effect, Which is weird as carbon trading should have rapid seasonal fluctuation; such a phenomenon is attributed to the trial stage of carbon trading in China (Figure 5). Therefore the deseasonalized models are considered as shown in the Formula (4) and (5)

$$X_t = T_t + N_t \qquad N_t \sim N(0, \sigma^2) \tag{4}$$

$$X_t = T_t * N_t \qquad N_t \sim N(0, \sigma^2) \tag{5}$$

To further analyze data, there is prolonged research based on stationary time series data, whose properties are put in Definition 3.2.

图 8:　近十年一级和二级市场碳价走势（USD/tCO_2）



数据来源：ICAP　华泰期货研究院

Figure 5: Carbon credit price

Definition .1

A Time Series is stationary if it has the following conditions:

$$1.\quad \mathbb{E}[X_t] = \mu \tag{6}$$

$$2.\quad Var(X_t) = \sigma^2 \tag{7}$$

$$3.\quad Cov(X_t, X_{t-k}) = \gamma_k \tag{8}$$

□

3.2 Decomposition

As the data doesn't show a strong seasonality but exhibit a quadratic-like shape; therefore, its time dependency is checked through Auto-Correlation Function (ACF)

Definition .2

The autocorrelation function (ACF) reveals how the correlation between any two values of the signal changes as their separation changes

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{\sqrt{Cov(X_t, X_t) * Cov(X_{t-k}, X_{t-k})}} \tag{9}$$

□

The ACF of X in Figure 6 and the Partial ACF of X in Figure 7 shows that: there is a strong relationship between time t and $t-k$ for $|k| \geq 1$ and $k \in \mathbb{Z}$. It is reasonable to assume time series Y is not stationary which is not suitable for prediction, so some techniques need to be used to decompose it. Now we take the multiplicative model such that $X = \{\ln X_1, \ln X_2, \ldots, \ln X_{96}\}$.

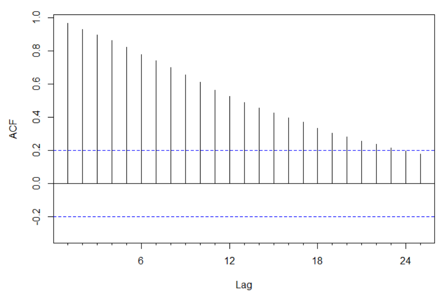


Figure 6: ACF result of X

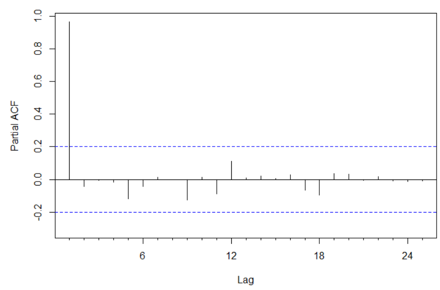


Figure 7: Partial ACF result of X

There are 3 common techniques to decompose:

1. Differencing 2. Filtering 3. Least square fitting

As there is a strong correlation between time, so differencing is introduced with regard to this situation.

Definition .3

Backshift operator $\mathcal{B}$:

$$\mathcal{B}X_t \triangleq X_{t-1} \quad \Rightarrow \quad \mathcal{B}^k X_t \triangleq X_{t-k} \tag{10}$$

Usage: to remove p-degree polynomial trend by p times of differencing
$\square$

Before differencing, it is helpful if some techniques can help us to decide the degree of Backshift operator. The built-in function *ndiffs()* suggests a second-degree differencing.

```
> ndiffs(x_ts)
[1] 2
```

The result in Figure 8 shows that one degree of differencing is enough for X to become stationary.

Then the ACF of $X_t(1-\mathcal{B})$ is checked, and the *Augmented Dickey-Fuller* test is used to check whether the differenced data is stationary or not.

```
Augmented Dickey-Fuller Test

data:  x1diff
Dickey-Fuller = -4.0855, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(x1diff) : p-value smaller than printed p-value
```

As the output shows that the p-value is smaller than $\alpha = 0.05$, we cannot reject the null hypothesis. Thus, $X_t(1-\mathcal{B})$ is stationary.

The trend is then got by calculating X_t - $X_t(1-\mathcal{B})$, the smoothed trend is taken by applying filter $[1/24, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/24]$

3.2.1 Noise analysis

So far we can use two methods to analyze the noise:

1. Differencing 2. Filtering

For the first method, the stationary time series (Noises) made by 1^{st} degree differencing has $\mathbb{E}[N] = 0.012$ & $Var(N) = 0.008$. While the second method also gives similar values with $\mathbb{E}[N] = -0.005$ &

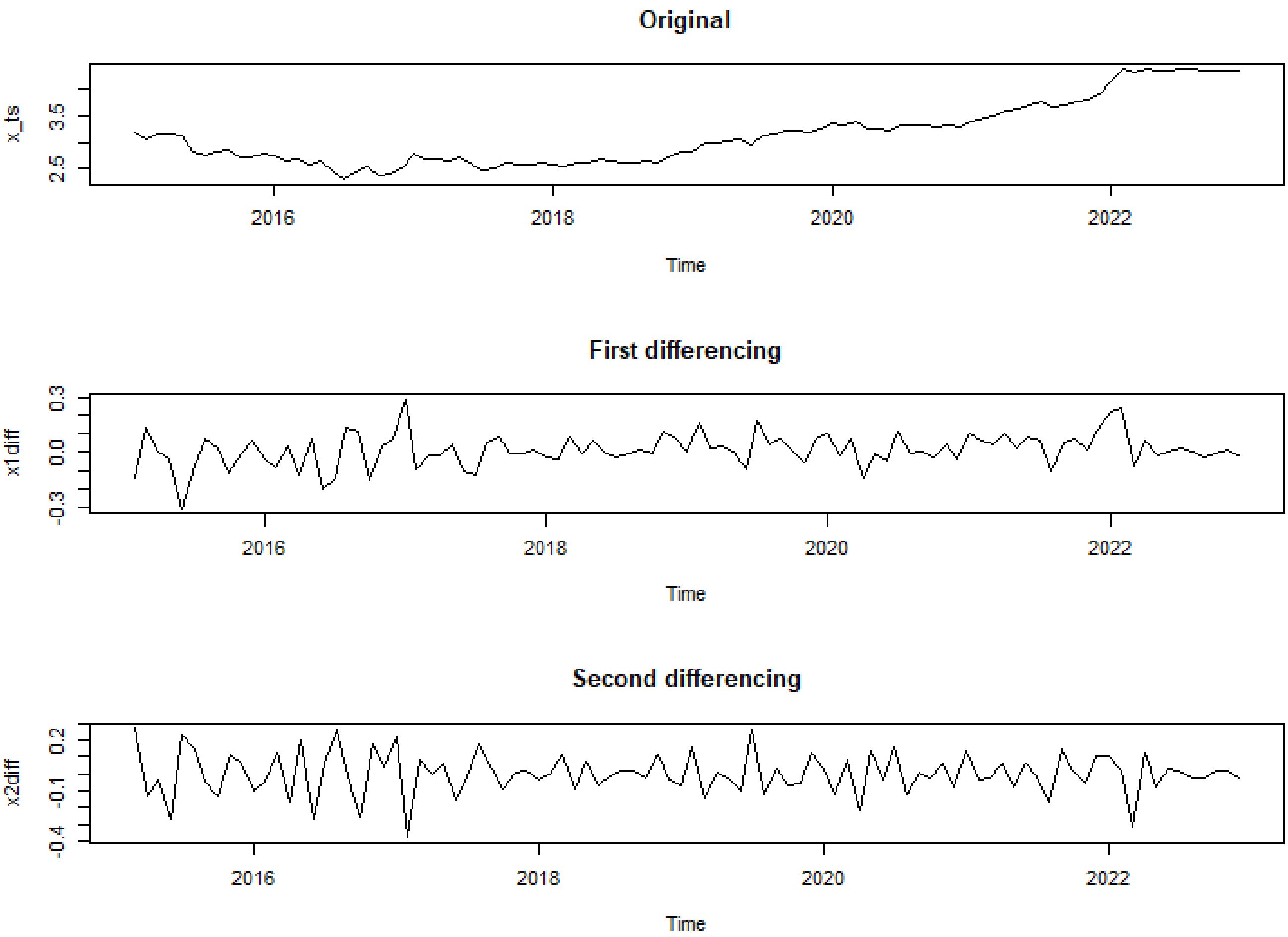


Figure 8: Differencing

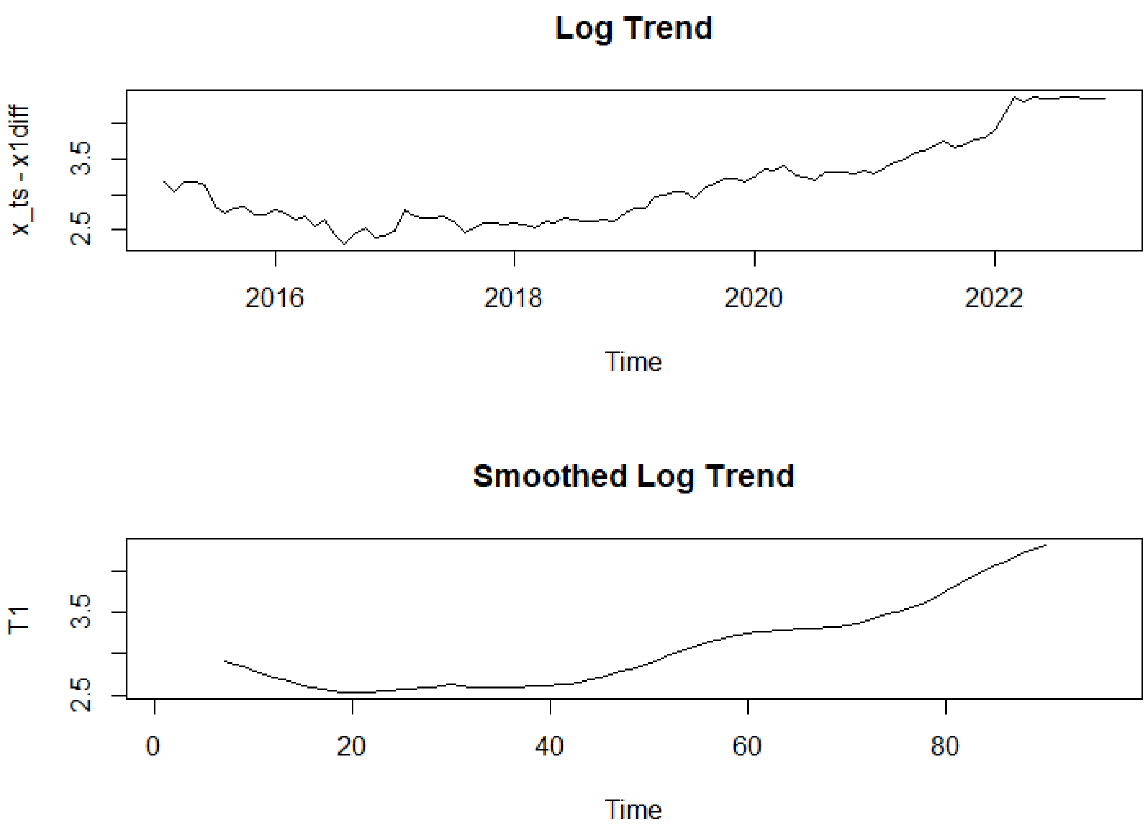


Figure 9: Log trend and smoothed log Trend

$Var(N) = 0.007$. The time series plot and QQ plot for each method are shown below.

From the above information, $Cov(X_t, X_{t-1})$ is high, it also shows the stationary residue with normal distribution $N_t \sim N(0, 0)$. Therefore, the random walk assumption for dataset X is reasonable.

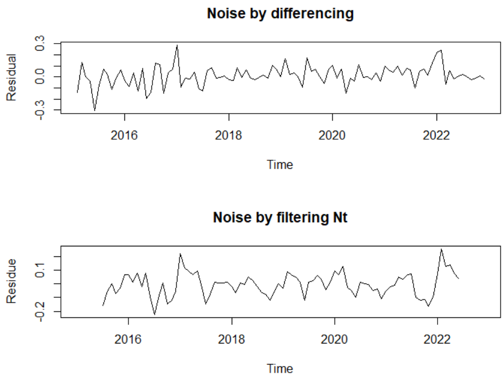


Figure 10: Residue plot

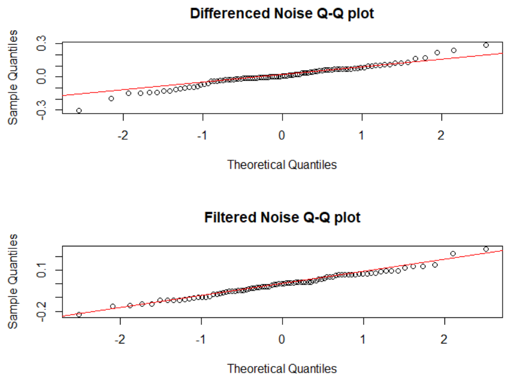


Figure 11: Residue QQ plot

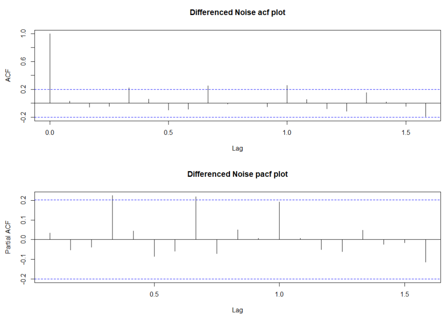


Figure 12: ACF and PACF of differenced noise

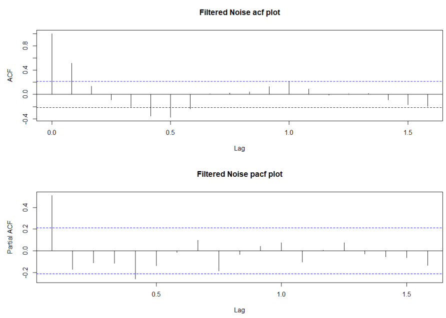


Figure 13: ACF and PACF of filtered noise

3.2.2 Preparation for forecasting - HMA

Before we move further, there is a non-negligible error to be solved: the filtering method we used to smooth the trend will cause missing data issues, especially for the most recent data which is much more important as it started levelling off. However, the smoothed trend does not show such an inflexion; an alternate smoothing method should be used. Among Kalman filter, Particle filter, and other low-pass filters, taking complexity and accuracy into consideration, Hull Moving Average(HMA) is selected.

The HMA, developed by Alan Hull, is an extremely fast and smooth-moving average. The HMA almost eliminates lag altogether and manages to improve smoothing at the same time. It is commonly adopted in stock analysis as an advanced industry smoothing method.

A long-period HMA may be used to identify trends. If the HMA is rising, the prevailing trend is rising, indicating it may be better to enter long positions and vice versa. While a short period, HMA may be used for entry signals in the direction of the prevailing trend. A long entry signal, when the prevailing trend is rising, occurs when the HMA turns up, and a short entry signal, when the prevailing trend is falling, occurs when the HMA turns down.

Considering our situation that we need to long carbon credits within three months with a large volume, while the unit price seems to become flat and may decrease, the smoother should be able to catch the inflexion signal to reduce the cost. Thus, A HMA with a short period is carefully chosen. However, this paper is mainly for the neural network models, so HMA-adopted models will only be put into comparison but not shown explicitly.

Definition .4

Weighted Moving average(WMA) for data X with k period
$\triangleq \mathrm{WMA}^{(k)}(X)$:

$$X_t = \vec{W} \cdot \vec{X} \tag{11}$$

While

$$\vec{W} = (W_{t-1}, W_{t-2}, \cdots, W_{t-k}) \;\; \& \;\; \vec{X} = (X_{t-1}, X_{t-2}, \cdots, X_{t-k})$$

□

Definition .5

Hull Moving average(HMA) $\triangleq \mathrm{HMA}^{(k)}(X)$:

$$\mathrm{WMA}^{\sqrt{k}}(2 * \mathrm{WMA}^{(\frac{k}{2})}(X) - \mathrm{WMA}^{(k)}(X)) \tag{12}$$

While WMA is defined in Formula (11)
□

4 Data centric models

This chapter is for machine learning models and other modern models in complement to the conventional statistical models.

4.1 Naive machine learning

4.1.1 Cross validation

In this part, the data is split into 3 parts: training set, cross-validation set, and test set. Short data splitting is chosen to forecast a short period of data. For convenience, a Python library called Sktime is introduced in this section. It is an open-source Python toolbox for machine learning with time series funded by the UK Economic and Social Research Council, the Consumer Data Research Centre, and The Alan Turing Institute. It extends the sci-kit-learn API to time series tasks. The function temporal train test split with 6 splits is used to split the data which means the last 6 data are used for validation. The split data is shown in Figure 14. ExpandingWindowSplitter(initial window=22, step length=13,fh=6) is used as a cross-validation splitter.

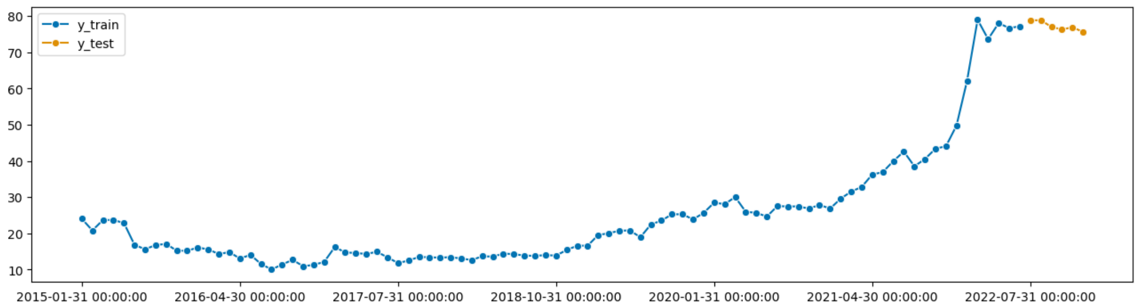


Figure 14: Temporal train test split (6 splits)

Then, Naive forecaster(sp=5) is used as the base model, the plot with 90% confidence in Figure 15 shows the in-sample prediction for cross-validation.

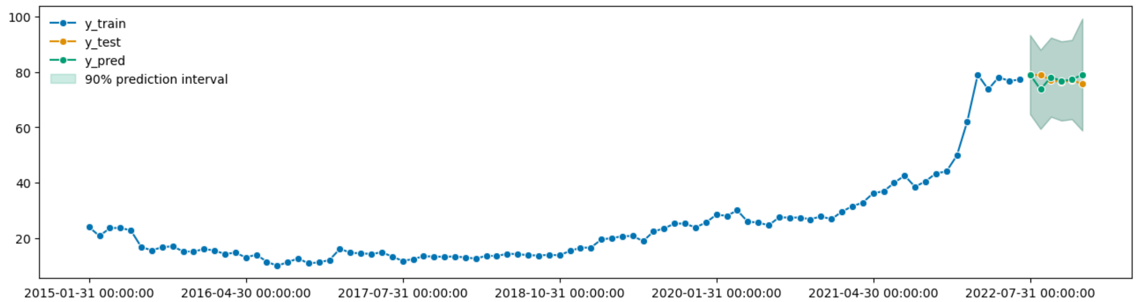


Figure 15: Naive forecaster(sp=5,CI=0.9) for cv

After fitting, AutoETS is the best basic model in Sktime with the performance shown in Figure 16, the plot is bounded by 90 confidence interval.

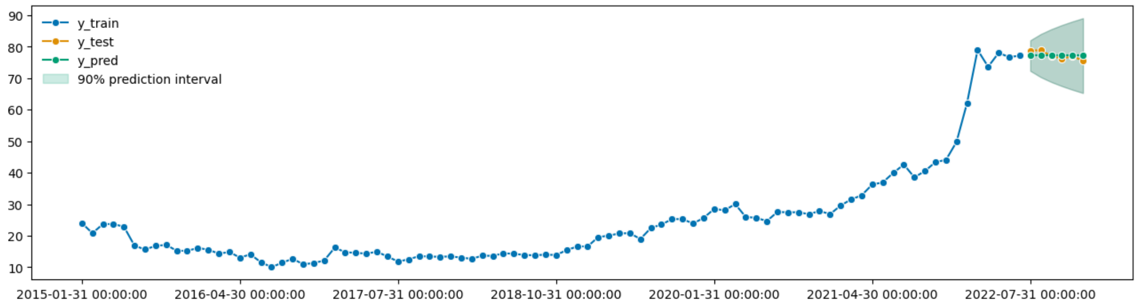


Figure 16: AutoETS (CI=0.9) for cv

Table 1: MSE of base models

Metric	Naive	AutoETS
MSE	6.587	1.456

AutoETS outperforms Naive forecaster for the selected dataset in cross-validation.

4.1.2 Prediction

After cross-validation and hyperparameters tuning, the prediction for the out-of-sample 6 data by naive forecaster and AutoETS is shown in Figure 17 and Figure 18.

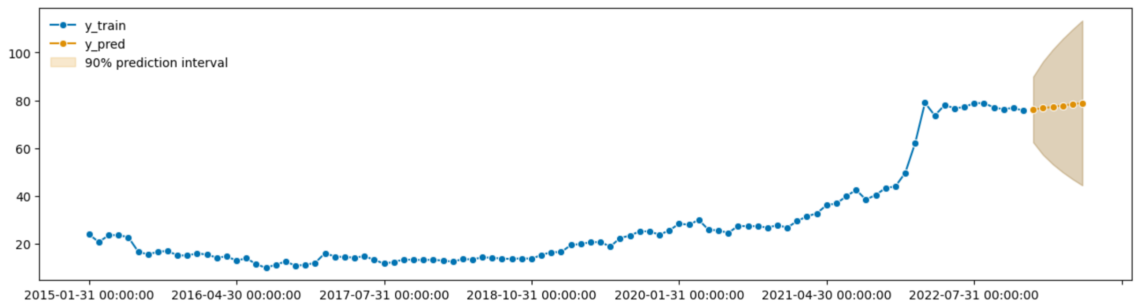


Figure 17: Naive forecaster(sp=5,CI=0.9) for cv

Apparently, AutoETS perform better with short prediction boundary and small error. The MSE for predicted data is shown below.

However, AutoETS shows a flat trend with the same output for both January and February data.

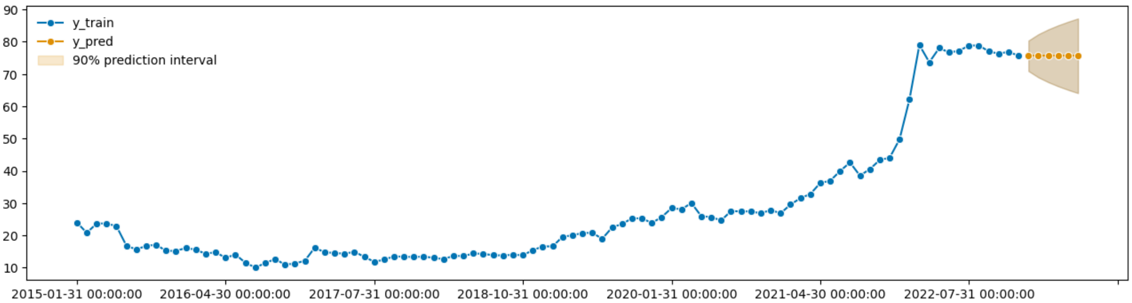


Figure 18: AutoETS (CI=0.9) for test

Table 2: MSE of base models

Model	MAE	MSE
Naive	3.85	8.16
AutoETS	0.479	0.12

4.2 Neural network

For this part, LSTM and GRU are chosen to check their performance on the selected dataset. Due to the complexity and poor interpretability, only the training performance is shown here.

We use the min-max scalar to do feature scaling so that the algorithm can converge faster with less loss. The parameters for LSTM is: input_dim = 1, hidden_dim = 32, num_layers = 4, output_dim = 1, num_epochs = 100, loss = MSE, learning rate = 0.005. The error scores are Train Score: 5.11 MSE, CV Score: 4.58 MSE. The performance graph is shown in Figure 19.

The in-sample cv dataset prediction is shown in Figure 20.

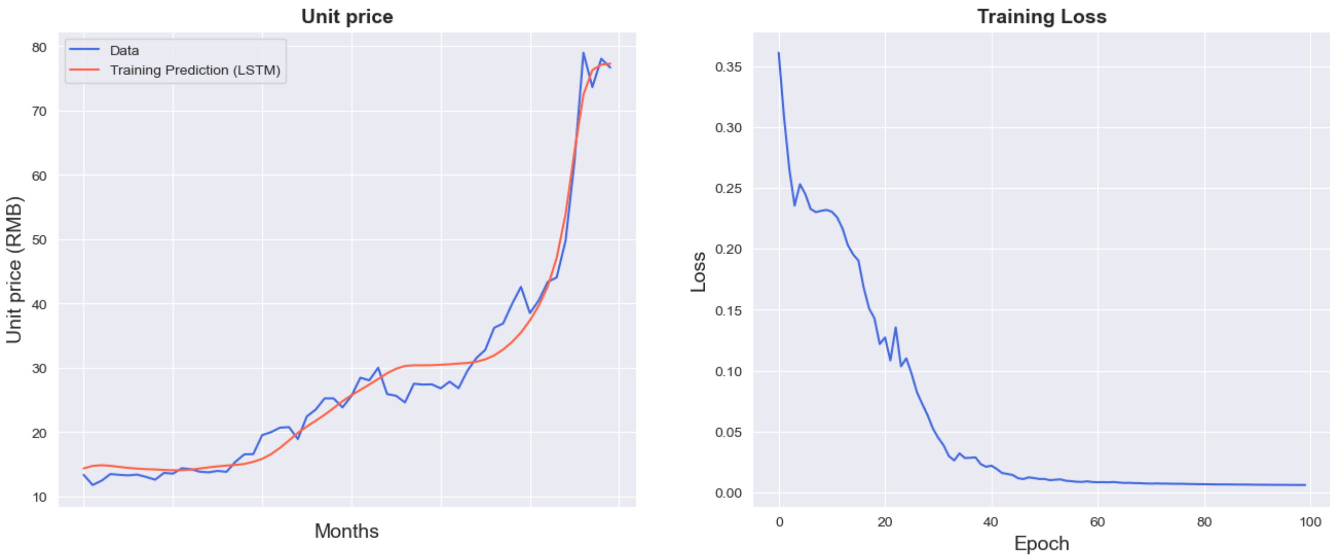


Figure 19: LSTM training performance

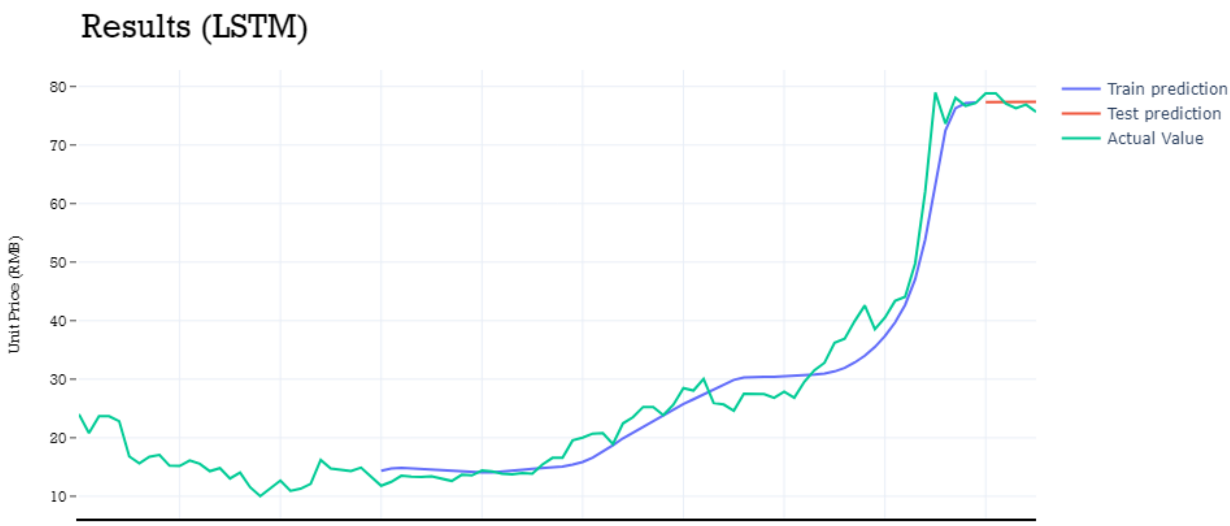


Figure 20: LSTM performance

In 19, the prediction shows a flat trend. By checking the values, there are minor differences between them.

Then We use the same hyperparameters to check the GRU model, which is also a neural network model for sequential data with a simpler architecture than LSTM. The performance is shown in Figure 21.

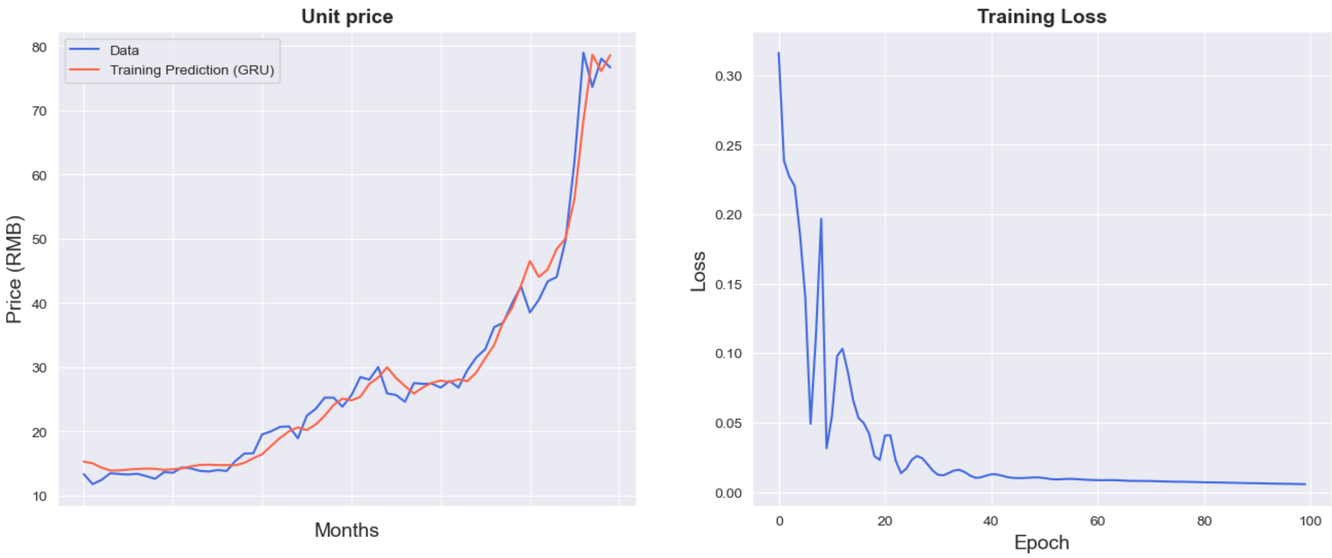


Figure 21: GRU training performance

The in-sample cv dataset prediction is shown in Figure 22.

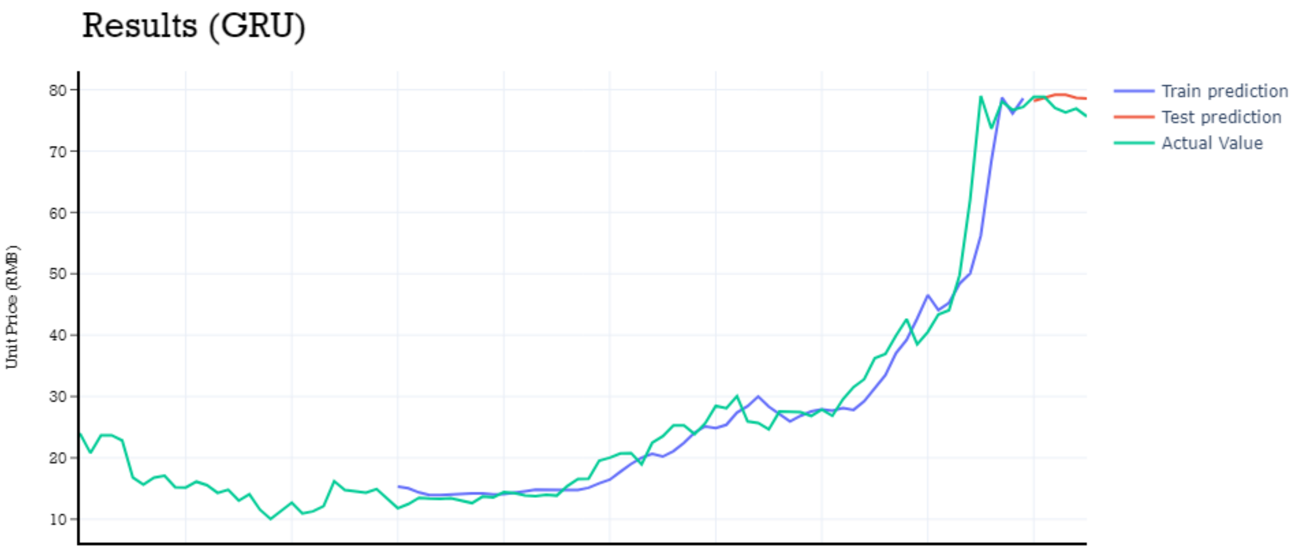


Figure 22: GRU performance

5 Conclusion

After comparison among the models, the models with outstanding performance are listed in Figure 23.

Though models' performance differs in smooth and raw data, It is believed the fluctuation is mainly due to the smooth original trend of the raw data. Unlike other countries, the selected data cannot show the seasonal fluctuation and rapid trend exhibited by other markets. However, the models still can give a satisfying result in this analysis.

Model	MAE
Linear_Raw	13.22 (AIC:8.943341,BIC:16.375)
GARCH (Raw)	0.80243 (AIC:4.299, BIC:4.496)
ARIMA_Smooth	0.8102 (AIC:304.57,BIC:314.44)
Exponential alpha=0.9	0.3058
AUTOETS, LSTM	0.479

Figure 23: Final Comparison

References

The situation in China:
https://www.bbc.com/news/business-22931899

https://www.cnemission.com/article/gywm/201907/20190700001675.shtml

https://www.bbc.com/zhongwen/trad/science-57867157

https://www.qianzhan.com/analyst/detail/220/210721-00170013.html

Carbon-trading-market:
https://asia.nikkei.com/Spotlight/Environment/Climate-Change/China-s-national-carbon-trading-market-eyes-June-debut-in-Shanghai

polynomial Regression:
https://www.statology.org/polynomial-regression-r/

Linear Regression:
https://www.scribbr.com/statistics/linear-regression-in-r/

https://online.stat.psu.edu/stat501/lesson/4/4.3

https://www.rose-hulman.edu/class/ma/inlow/Math485/ftests.pdf

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2018). Short-Term Forecasting: When Do Statistical Models Outperform Judgment? International Journal of Forecasting, 34(1), 3-16. doi: 10.1016/j.ijforecast.2017.06.00

New Economic Opportunities:
https://www.bbc.com/zhongwen/trad/science-57867157

https://carboncredits.com/tesla-carbon-credit-sales-reach-record-1-78-billion-in-2022/

Tan, S. T. Applied Mathematics for the Managerial, Life, and Social Sciences, latest edition, Brooks / Cole, Cengage Leaming

Anderson, D.R., Sweeney, D. J., Williams, T. A. Quantitative Methods for Business, edition, Thomson South-Western.

GARCH:
https://rstudio-pubs-static.s3.amazonaws.com/346563_b404c507f20549f196f43c57d363595f.html
https://rpubs.com/Sharique16/garch
https://rpubs.com/Naishad/forecasting-bitcoin-v1

ML:
http://www.sktime.net/en/latest/

Deep:
https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html
https://zh.d2l.ai/chapter_recurrent-modern/gru.html