Table of Content

Introduction

Hong Kong is an international city with a population of 7 million, and the night view of Victoria Harbour is famous worldwide. In this research report, we would like to investigate the backbone of this city - electricity. We are adopting a time series analysis method to the electricity data to predict future electricity consumption.

There are two main companies providing electricity in Hong Kong: HK Electric and CLP. HK Electric provides service to Hong Kong Island and Lamma Island, and CLP is responsible for the electricity to all other areas. The two companies have six main power plants in Hong Kong. While most electricity is generated from non-renewable sources, and only 30% is generated from renewable sources. At the same time, the fee for Hong Kong electricity is around 0.160 0.160 U.S. Dollars per kWh for households which is higher than the global average of 0.14 U.S. Dollars per kWh and was 90 Rank in international.

This report will adopt various time series models to find the best suited for Hong Kong's electricity data. Including ARMA, SARIMA, ARIMA, SARIMAX, Linear Regression and Holt Winter. While adopting a regression model to investigate the dataset further.

Methodology and Results and Discussion

Descriptive Analysis

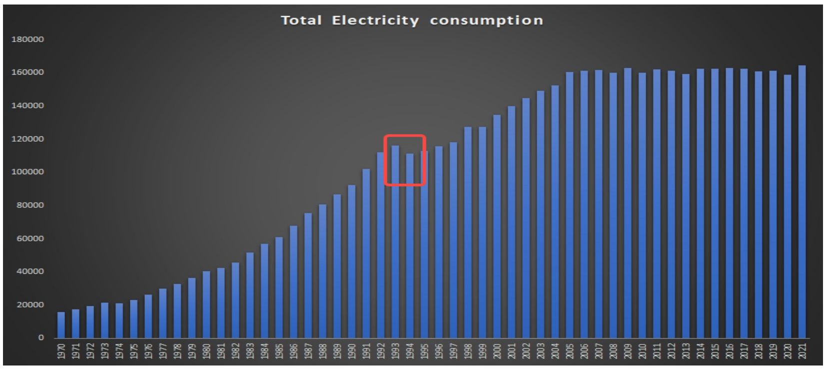


Table 1: Total electricity consumption from 1970 to 2021

Year	Domestic	Commercial	Industrial	Street lighting	Exports to China
1993	24092	53131	22309	278	16201
1994	25827	57508	21437	282	6327

Table 2: Total electricity consumption in 1970 and 2021

From table 2, the electricity consumption had increased from 16023 to 164578 Terajoule with a 927% increase. The yearly electricity consumption is shown by the histogram above. There is a slight decrease in 1994. That is because there was a sharp decrease in export to China by 10000 Terajoule in 1994. Overall the increasing trend is noticeable. Electricity consumption has been increasing since 1970. After 2005, it maintained a level of around 160000 Terajoule per year. And people's demand for electricity is growing in different areas like domestic and commercial use.

Year	Domestic	Commercial	Industrial	Street lighting	Exports to China
1970	20.83%	37.60%	41.12%	0.45%	0.00%
2021	28.84%	64.18%	6.78%	0.19%	0.00%

Table 3: Total electricity consumption proportion in 1970 and 2021

From 1970 to 2021, the amount of commercial electricity used is the highest proportion of the total electricity usage. The ratio has increased from 37.6% to 64.18% in commercial areas.

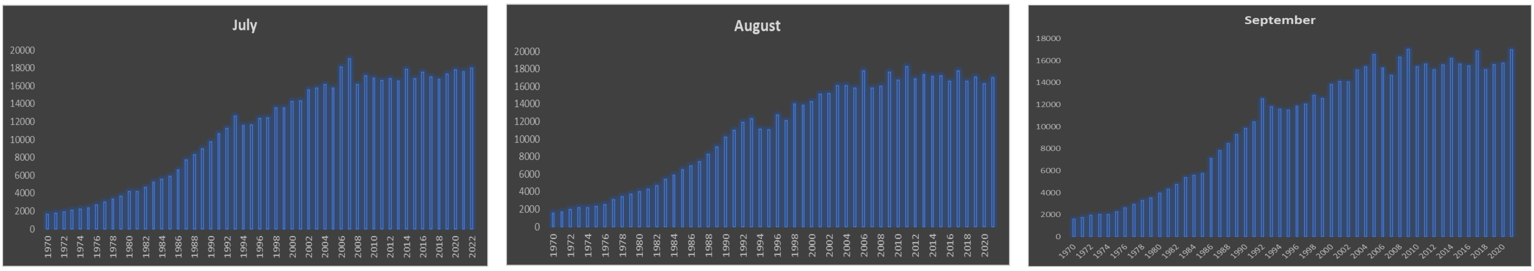


Table 4: Total electricity consumption in July, August and September

On top of that, July to September are the three months with the highest electricity consumption compared with other months. There may be various factors leading to this scenario. For one thing, seasonal factors and technological advancement influence people's electricity usage as these three months are the hottest months in a year. Domestic and commercial usage is getting higher during these months.

Month	Average of Domestic	Average of Commercial	Average of Industrial	Average of Exports to China	Average total electricity consumption
Jan	1470.00	4179.23	1106.06	525.70	7205.70
Feb	1486.96	3775.92	927.47	355.14	6501.23
Mar	1513.06	4250.85	1083.42	520.74	7292.42
Apr	1559.38	4623.92	1218.55	462.93	7808.23
May	2028.30	5359.66	1343.00	494.45	9163.04
Jun	2590.92	5753.53	1372.68	527.32	10175.70
Jul	3173.68	6049.75	1429.04	572.09	11148.94
Aug	3220.53	5950.26	1424.15	600.59	11115.09
Sep	2991.62	5748.47	1389.81	552.66	10611.13
Oct	2267.54	5320.19	1324.87	475.65	9329.17
Nov	1543.65	4718.00	1240.44	497.40	7936.77
Dec	1373.00	4391.71	1179.48	606.98	7470.21
Average	**2103.32**	**5011.07**	**1253.27**	**516.19**	**8815.83**

Table 5: Total electricity consumption in different month

From the table above, Jul has the highest electricity usage throughout the year. And commercial usage accounts for the most significant proportion of total electricity consumption. Besides, domestic usage is higher than industrial use and export. One thing that should be noted is that export to China has ended since October 2018. That's because Hong Kong has transformed into a pure electricity import region ever since.

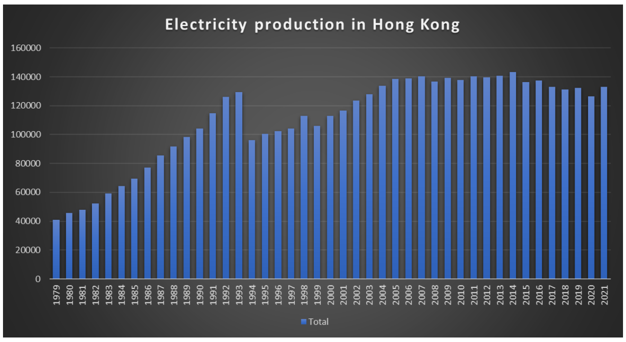


Table 6: Total production from 1979 to 2021

Apart from the electricity consumption, we gather data on local electricity production to determine the relationship with the consumption. A repeated rising trend was shown above and a sharp decrease in electricity production in 1994 as the demand decreased. After 2004, the electricity consumption has maintained a certain level which might indicate the Internal consumption is close to saturation.

Correlation between different industries						
	Domestic	Commercial	Industrial	Exports to China	Total consumption	Total production
Domestic	1.0000					
Commercial	0.9888	1.0000				
Industrial	-0.6626	-0.6824	1.0000			
Exports to China	0.1814	0.2508	0.1068	1.0000		
Total consumptio	0.9762	0.9874	-0.5780	0.3701	1.0000	
Total production	0.9066	0.9187	-0.4140	0.5231	0.9648	1.0000

Table 7: Correlation between electricity consumption and production

From table 7, domestic and commercial electricity use is highly correlated with total electricity consumption and production. Match our findings before. Meanwhile, the industrial use of electricity is negatively correlated, and export to China has a moderate relationship with the total consumption. Total consumption is highly associated with production as they are in a demand-and-supply relationship.

ARMA Model

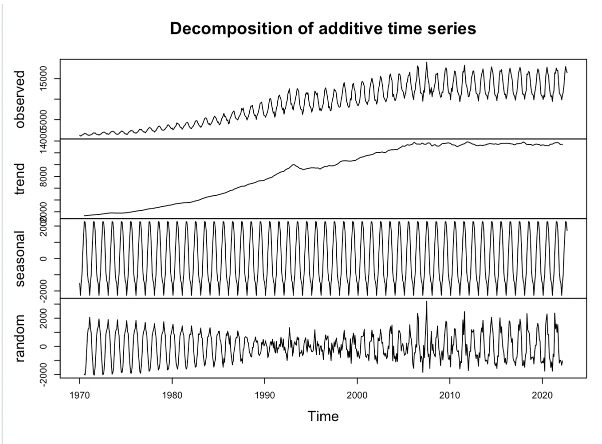
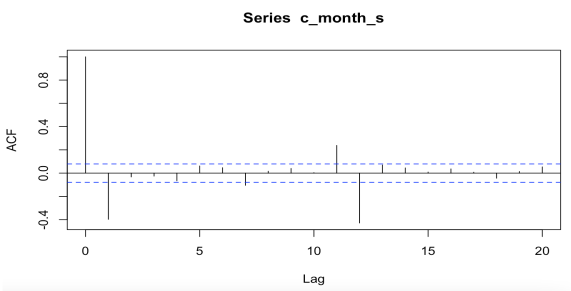


Table 8: Decomposition of additive time series

To predict the local energy consumption by building an ARMA(p,q) model. With the decompose function, it is clear to see there is an increasing trend, an increase in fluctuation, and a seasonal pattern on the time series data.

Then, we transform the data to stationary data by taking log and differencing to remove the increasing trend, increase in fluctuation and the seasonal pattern. We check the stationarity with an Augmented Dickey-Fuller test. From the table in appendix 1[1], the p-value is very small, which is smaller than 0.01. The test result shows that the transformed data is stationary.



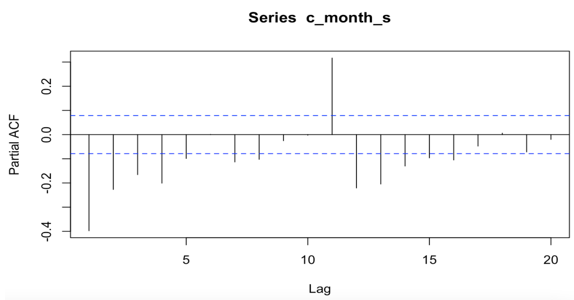


Table 9: ACF(left) and PCAF(right) plot for electricity consumption

Then, an ACF plot was plotted above. There is a clear cut-off pattern on the graph, which suggests that the data follow the moving average (MA) process. We plot a PACF graph to check again. We can see that the data follow an exponential pattern, which also shows that the data follow the MA process.

model	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)
AIC	-1720.74	-1729.81	-1738.06	-1745.47	-1743.82

Table 10: AIC table for different MA models

After that, we try various MA models from MA(1) to MA(5) to fit the dataset and choose the model with the smallest AIC. The result of testing MA(1) to MA(5) models is shown above.

We have found that a MA (4) model is the best model to fit the data[2]. The model is:

$$X_t = \mu - 0.002\epsilon_{(t)} - 0.6533\epsilon_{(t-1)} - 0.1153\epsilon_{(t-2)} - 0.1178\epsilon_{(t-3)} - 0.1136\epsilon_{(t-4)}$$

Although the best MA model is found, it does not necessarily mean the model is an excellent model to fit the data. Therefore, we have done a Box-Pierce test and a Ljung-Box test to test whether the residuals are independent identically distributed (IID). As the p-values are very small, smaller than 0.05[3]. Hence, rejecting the null hypothesis of the data is IID, which implies that the model is not a good model to fit the data.

Hence, splint the data in 1995 and transform the data after 1995 to stationary data by taking log and differencing for further processing.

---

[1] Appendix 1: Differencing table and adf.test result
[2] Appendix 2: Result of ARIMA model
[3] Appendix 3: Box-Pierce test and a Ljung-Box test for arma

ARMA Model (Cont.)

There are three main auto selection methods for time series model selection. They are:

- ESACF: Determine the model with the least AR term possible
- MINIC: Determine the model with the least BIC
- SCAN: Determine the model with the highest SSC

While these 3 methods suggested 5 models:

- ESACF: Suggested ARMA (0,1),(1,1),(2,1)
- MINIC: Suggested ARMA(5,0)
- SCAN: Suggested ARMA(0,2)

While the summary result of the ARMA models are as the following table

	ARMA(0,1)	ARMA(1,1)	ARMA(2,1)	ARMA(5,0)	ARMA(0,2)
Number of AR term	0	1	2	5	0
Number of MA term	1	1	1	0	2
AIC	5205.91	5181.99	5183.92	5226.27	5184.89
log likelihood	-2599.96	-2586.99	-2586.96	-2606.14	-2588.45
Sigma^2	655989	605971	605896	692585	611148

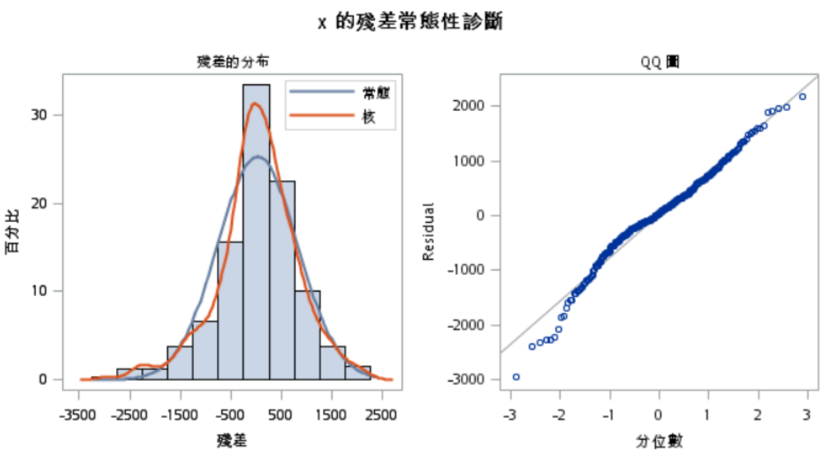
Table 11: ARMA model properties and result

From the table we could see that the ARMA(1,1) model has the best performance as it has the lowest AIC.

The ARMA (1,1) model is:

$$[(X_t + 0.6763) - 0.2887(X_{t-1}) + 0.67628\ ] = \epsilon_{(t)} + 0.9834\epsilon_{(t-1)}$$

This model is also supported by the residual and noise. As there is no relationship between residual and the noise is unpredictable. We could say that the ARMA(1,1) model is a good fit with the model after removing the seasonal effect.




```
> Box.test(zzz, type = "Box-Pierce", lag = 12)

        Box-Pierce test

data:  zzz
X-squared = 870.73, df = 12, p-value < 2.2e-16

> Box.test(zzz, type = "Ljung-Box", lag = 12)

        Box-Ljung test

data:  zzz
X-squared = 879.89, df = 12, p-value < 2.2e-16
```

Autocorrelation Check of Residuals									
To Lag	卡方	DF	Pr > ChiSq	自相關					
6	2.47	4	0.6500	-0.006	0.035	-0.036	-0.039	-0.023	0.054
12	87.31	10	<.0001	-0.055	0.085	0.003	0.065	0.138	-0.469
18	98.26	16	<.0001	-0.121	-0.078	0.082	0.009	0.064	-0.025
24	107.86	22	<.0001	0.132	0.015	0.042	-0.071	-0.058	0.003
30	131.47	28	<.0001	0.127	0.211	0.002	0.068	-0.042	0.021
36	137.87	34	<.0001	-0.085	-0.020	0.008	0.045	0.001	0.090
42	149.19	40	<.0001	-0.031	-0.143	-0.021	-0.068	0.066	-0.015
48	154.78	46	<.0001	0.093	0.036	0.027	0.021	0.023	-0.058

Table 12: Result of arma model (1,1), Autocorrelation check for Residuals(Top), Distribution of the model(bottom left), Q-Q plot of the model(Bottom middle), Box test of the model (Bottom right)

SARIMA

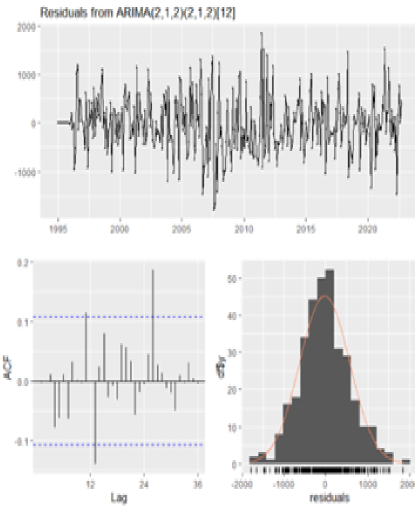
From the ARMA model, an autocorrelation relationship is found according to the randomness of data. In order to study the seasonal effect on energy consumption, we would like to construct a SARIMA model.

```
Series: x
ARIMA(2,1,2)(2,1,2)[12]

Coefficients:
          ar1     ar2      ma1      ma2     sar1     sar2     sma1     sma2
      -0.6510  0.2592  -0.0497  -0.8182  -0.8011  -0.2144  -0.0473  -0.3746
s.e.   0.1036  0.0589   0.0920   0.0819   0.3461   0.0769   0.3491   0.2595

sigma^2 = 378252:  log likelihood = -2511.3
AIC=5040.61   AICc=5041.19   BIC=5074.52

Training set error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE
Training set -26.90329 595.314 458.8136 -0.4439065 3.601778 0.7314076
                     ACF1
Training set -0.003181653
```



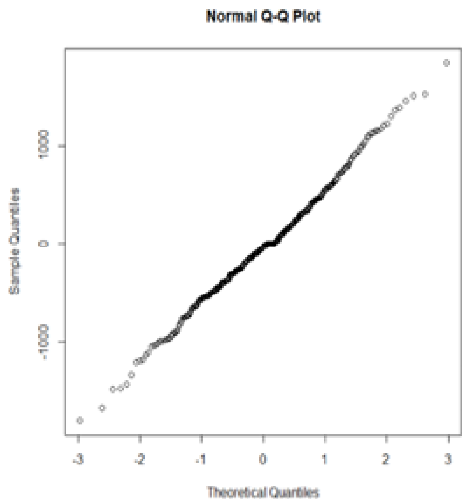


Table 13: Result of SARIMA model

After investigation, ARIMA(2,1,2)(2,1,2)[12] would be our final model. This is supported by the training set error. The model accuracy is high with 96.4%(1-MAPE*0.01), and MAE is relatively low compared to the dataset value). According to the q-q plot, all values in the ACF bar stay between the blue dotted line. This tells residuals tend to be normally distributed. Showing the model has no bias in estimating energy consumption and is a suitable model.

SARIMA Model:

$$Y_t - Y_{t1} - Y_{t12} = \omega - 0.651 * W_{t-1} + 0.2592 * W_{t-2} - 0.0473 * W_{t-12} - 0.3746 * W_{t-24}$$

$$- 0.0497 * \varepsilon_{t-1} - 0.8182 * \varepsilon_{t-2} - 0.0473 * \varepsilon_{t-12} - 0.3746 * \varepsilon_{t-24} + \varepsilon$$

ARIMA(2,1,2)(2,1,2)[12]states our model has 2 AR terms, 1 difference, 2 MA terms, 2 seasonal AR terms, 1 Seasonal difference and 2 seasonal MA terms and the length of seasonal would be 12 (data are presented monthly). Because both trend and seasonal have shown an increasing trend, we need to include a difference in both trend and seasonal parts. The time series for the trend is dependent on the value and the error of the past 2 months and the seasonal effect is dependent on the value and the error of the past 12 and 24 months.

Prediction using SARIMA Model

As ARIMA(2,1,2)(2,1,2)[12] is a fitted model, we would like to use this model to estimate the value in 2022 to 2025 by monthly.

> fore

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Oct 2022	13769.428	12911.639	14627.22	12457.553	15081.30
Nov 2022	11595.754	10872.009	12319.50	10488.882	12702.63
Dec 2022	10993.923	10306.341	11681.51	9942.356	12045.49
Jan 2023	10921.711	10237.147	11606.27	9874.761	11968.66
Feb 2023	9815.686	9199.003	10432.37	8872.551	10758.82
Mar 2023	11261.582	10552.286	11970.88	10176.807	12346.36
Apr 2023	11780.706	11036.733	12524.68	10642.897	12918.51
May 2023	14678.115	13748.538	15607.69	13256.449	16099.78
Jun 2023	16204.265	15174.954	17233.58	14630.070	17778.46
Jul 2023	17420.677	16310.586	18530.77	15722.939	19118.42
Aug 2023	16924.559	15842.471	18006.65	15269.648	18579.47
Sep 2023	16086.004	15053.905	17118.10	14507.544	17664.46
Oct 2023	13774.791	12857.768	14691.81	12372.325	15177.26
Nov 2023	11600.270	10825.205	12375.34	10414.910	12785.63
Dec 2023	10998.205	10260.571	11735.84	9870.091	12126.32
Jan 2024	10925.965	10190.261	11661.67	9800.804	12051.13
Feb 2024	9819.509	9155.566	10483.45	8804.095	10834.92
Mar 2024	11265.969	10500.930	12031.01	10095.943	12435.99
Apr 2024	11785.294	10981.390	12589.20	10555.829	13014.76
May 2024	14683.831	13677.532	15690.13	13144.829	16222.83
Jun 2024	16210.576	15094.264	17326.89	14503.324	17917.83
Jul 2024	17427.462	16221.327	18633.60	15582.838	19272.09
Aug 2024	16931.150	15753.280	18109.02	15129.754	18732.55
Sep 2024	16092.268	14966.753	17217.78	14370.941	17813.59
Oct 2024	13780.344	12779.399	14781.29	12249.531	15311.16
Nov 2024	11604.946	10757.505	12452.39	10308.896	12901.00
Dec 2024	11002.638	10194.753	11810.52	9767.084	12238.19
Jan 2025	10930.369	10123.240	11737.50	9695.971	12164.77
Feb 2025	9823.467	9093.848	10553.09	8707.611	10939.32
Mar 2025	11270.509	10428.406	12112.61	9982.623	12558.40
Apr 2025	11790.044	10903.717	12676.37	10434.523	13145.56
May 2025	14689.749	13578.490	15801.01	12990.225	16389.27
Jun 2025	16217.109	14982.411	17451.81	14328.801	18105.42
Jul 2025	17434.485	16098.363	18770.61	15391.062	19477.91
Aug 2025	16937.973	15631.170	18244.78	14939.390	18936.56
Sep 2025	16098.753	14848.168	17349.34	14186.148	18011.36

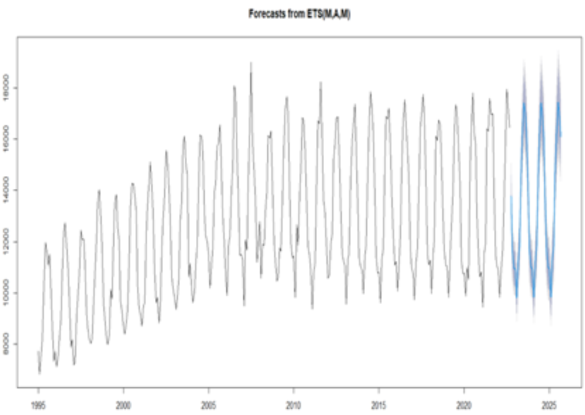


Table 14: Prediction of SARIMA model

From the time series plot, we have found that energy consumption has a stable increasing trend and a seasonal effect. This makes the time series model have high accuracy and efficiency. Thus we have found that the range inside the 95% confidence is relatively narrow due to the low standard deviation, which means the estimated value should be reliable. The blue line is the predicted value and we will find the shape of it is similar to the original data value in the time series. And the monthly value increases annually.

Linear Regression

First, in the linearity analysis, we can find that R^2 is very large and F=822, so we can conclude that there is a positive relationship between production and consumption. According to the model Y=-440.23+1.1873*x[4], we find that there is also a close relationship between production and energy consumption directly, so we need to carry out further verification on this basis.

Next we need to test whether it has a time series effect between energy consumption and production. The result was placed in the reference[5]. After comparison, we found that both of them are stationary and have similar data trends. Since the significance threshold was set at .05 is all that is required and the p-value is .01 which is lower than .05.So we conclude that there is a time series effect between energy consumption and production, and the two are strongly related.

[4] Appendix 4: Result of the linear regression model
[5] Appendix 5: Linear regression Result

At the same time, we found that both the Partial Autocorrelation Function and autocorrelation function are so similar to the model of similar time series effects on both energy consumption and production. Therefore, these two time series are highly correlated, which means that production is closely related to energy consumption, and a large amount of energy is consumed in the process of production.

SARIMAX

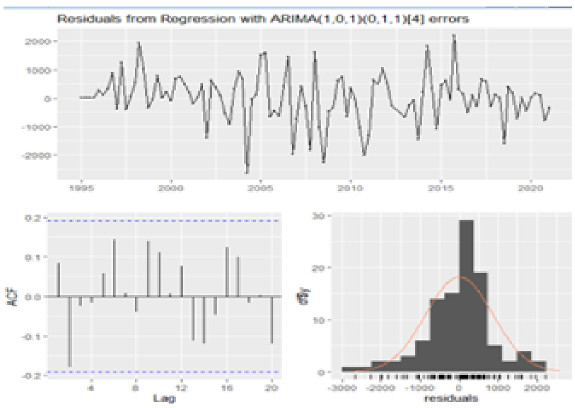
In order to estimate the relationship between energy consumption and production, an ARIMAX model has been selected to predict and forecast energy consumption by the value of production.

```
Series: gpl
Regression with ARIMA(1,0,1)(0,1,1)[4] errors

Coefficients:
         ar1      ma1     sma1    xreg
      0.9895  -0.8059  -0.5697  0.4125
s.e.  0.0135   0.0643   0.0902  0.0390

sigma^2 = 795254:  log likelihood = -828.52
AIC=1667.03   AICc=1667.66   BIC=1680.11

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE
Training set 4.516379 857.1248 625.2643 0.02138329 1.694969 0.4688008
                   ACF1
Training set 0.08305697
```



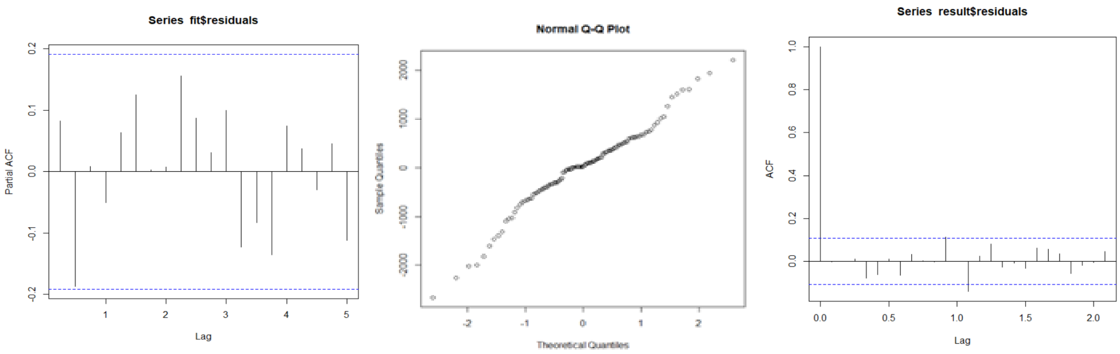


Table 15: Result of SARIMAX model

We have found that ARIMA(1,0,1)(0,1,1)[4] would be our final model. As we can see from the training set error, the model accuracy is very high with 98.31% (1-MAPE*0.01, MAE is relatively low compared to the dataset value). Residuals tend to be diagonal from the q-q plot. All of the value in the ACF bar stays between the blue dotted line, this tells us the residuals tend to be normally distributed. This means that our models have no bias in estimating the amount of energy consumption and it is a very good model.

Model:

$$Y_t - Y_{t12} = \omega - 0.9885 * W_{t-1} - 0.8059\varepsilon_{t-1} - 0.5697\varepsilon_{t-12} + 0.4125 * X + E_t$$

ARIMA(1,0,1)(0,1,1) states our model has 1 AR term,0 differences, 1 MA term, 0 seasonal AR terms, 1 Seasonal difference, 1 seasonal MA term, and the length of seasonal would be 4 (data are presented in quartiles). Because energy consumption and production have similar increasing trends, there is no need to include the differences. Using energy production to estimate energy consumption, the time series for the trend depends on the value and the error of the past quartile. Because the seasonal effect shows increasing trends, the seasonal difference is required to add to the model to make the data stationary. The seasonal effect mostly depends on the previous quartile for the seasonal error.

Data Trend Analysis

In terms of studying data trends, we did a simple linear regression to study the relationship between time and electricity, where time was set to year. In the chart, the red line is energy consumption after removing the trend, which shows almost straight growth and then horizontal growth after 2007.[6]

Therefore, we conclude that the overall trend of its data is a steady growth trend. According to the formula : $Y=1275+23.729x+\epsilon$, the trend of the electricity consumption increases by 23.7 Terajure per month.

Then we need to explore the impact of seasonal changes on energy consumption data. First, we define November to February as winter, March and April as spring, May to August as summer, and September to October as fall. Then the following table is obtained by analyzing the general data trend in the histogram.

Season	Value	Percentage
Spring	15100.64	0.138
Summer	41602.77	0.38
Autumn	19940.3	0.182
Winter	32723.4	0.299

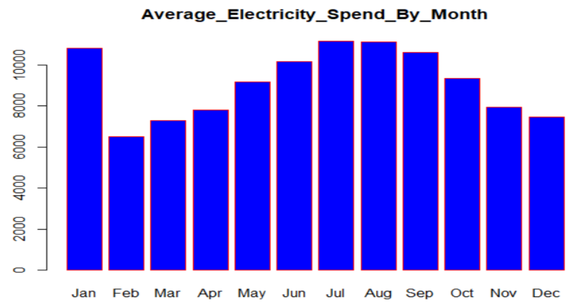


Table 16: Table of Seasonal Electricity usage (left), Bar chart of Average electricity spend by month(right)

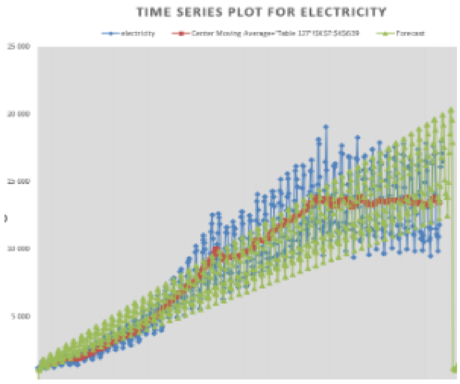
From the chart, we find that the electricity consumption in summer and winter is the largest, accounting for 38% and 29.9% of the total electricity consumption in the whole year respectively. From which it can be concluded that people will consume more electricity in winter and summer to maintain a relatively comfortable environment.

Next, we need to make a forecast for the future, and after adding seasonal effects we get the time series analysis in the following figure. Where the green line is the seasonal influence on the electricity consumption prediction, where the seasonal influence is different in different periods. Overall, the forecasted values should be roughly higher than the actual values.

The graph below shows the forecasted energy consumption by season for 2023 after adding seasonal effects, where we find that summer and winter are still the most consumed seasons. The overall energy consumption growth trend is steadily increasing and is within our forecast range. This means that there is no need to build additional power plants in the short term, and that the increase in energy consumption can be reduced through energy conservation and the promotion of energy-efficient appliances.

[6]Appendix 5:Linear regression Result

2023 Season	Electricity spend
Spring	26364.5
Summer	64469
Autumn	40695
Winter	74306.2



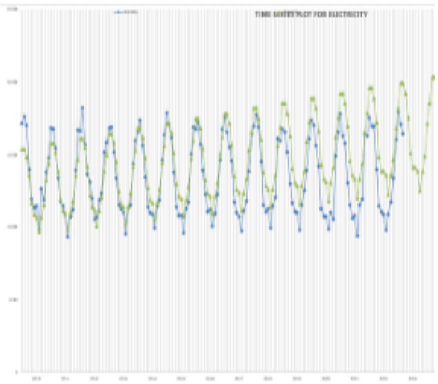


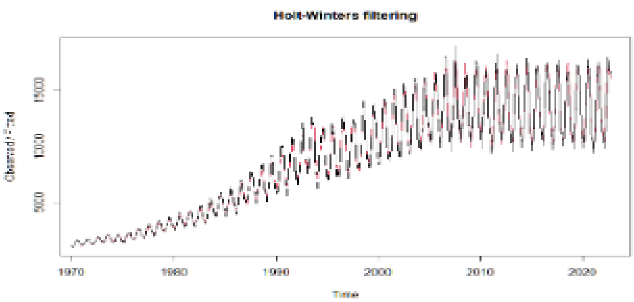
Table 17: Table of prediction of 2023 electricity consumption (top), Time series plot for electricity consumption all time(bottom-left) and Time series plot for electricity consumption from 2010-2028(bottom-right)

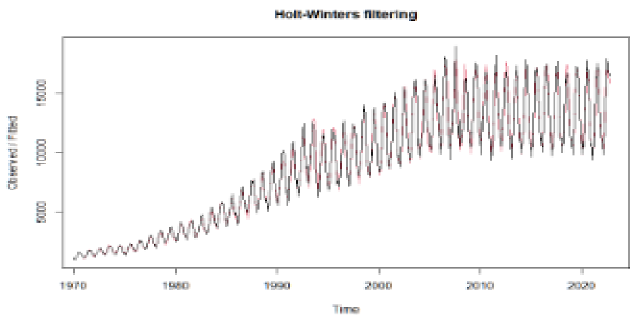
Holt Winters

We adopted both the additive seasonal form and the multiplicative seasonal form in the Holt winters method. and the results are as follows:

- Additive seasonal form: SSE is 174256325
- Multiplicative seasonal form: SSE is 162191209

The multiplicative seasonal form shows a lower SSE and its prediction line is closer to the





real data thus it is a better model.

Table 18:Additive seasonal form line chart (left) and Multiplicative seasonal form line chart (right)

Coefficient	Value	Coefficient	Value
a	13081	s6	0.8649846
b	3.44274	s7	0.9053494
s1	1.05087	s8	1.128393
s2	0.8857	s9	1.24521
s3	0.84114	s10	1.336962
s4	0.83756	s11	1.294574
s5	0.75305	s12	1.231044

Table 19: Results of Multiplicative Holt Winters

The prediction formula : $13081 + 3.44 \bullet k + s_{(n+k-p)}$

Conclusion

First of all, in terms of major consumer electrical products, most household electrical appliances in Hong Kong mainly focus on TV, air conditioning, refrigerator, washing machine and microwave oven, which can effectively improve the quality of people's life. Secondly, by comparing the total social electricity consumption of other major cities, we selected several cities with developed commerce for comparison. Compared with other major cities, the electricity consumption of Hong Kong is relatively high.

We found that electricity consumption in Chengdu and Shenzhen is two to three times that of Hong Kong. However, this is related to the industrial structure of the cities because these cities still have a certain amount of secondary industry. At the same time, Hong Kong is almost dependent on the tertiary sector, which consumes far less electricity. Therefore, we also chose Shanghai and Singapore, with similar commercial development levels to Hong Kong for comparison. We found that the electricity consumption of Hong Kong is three to six times that of Singapore and Shanghai. Therefore, we can conclude that electricity consumption in Hong Kong is relatively high.

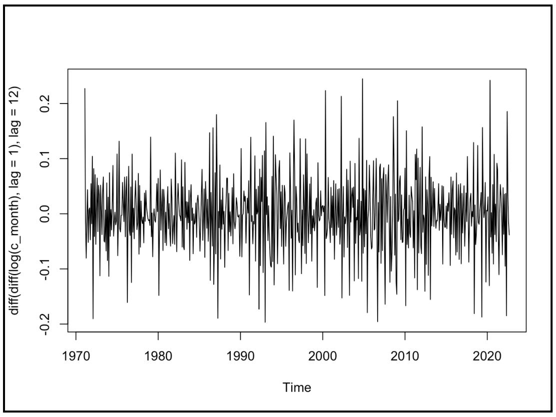
In terms of limitations and further suggestions, the factors affecting electricity consumption also include temperature, humidity and rainfall. Therefore, we can make our research more precise and accurate by adding the following two types of data. The first is the population data, which contains two small categories, namely the data on population flow and the data on Hong Kong's population growth. Because Hong Kong, as a tourist city, has visitors from all over the world on particular months and days, this can also cause a surge in electricity consumption. At the same time, power consumption in Hong Kong is also related to population growth.

The second is meteorological data. We believe that has a significant impact on electricity consumption in Hong Kong. We found that in addition to seasonal climate changes, special climate phenomena can also bring about changes in electricity consumption. Such as annual typhoons, El Nino, and even the urban heat island effect also impact electricity consumption.

Finally, after several models we constructed, we came to the following two conclusions. We found there is a strong relationship between electricity consumption and production. Another finding is that both consumption and production tend to have a stable increasing trend.

References and Appendix

1. DATA.GOV.HK.(n.d.). https://data.gov.hk/tc-data/dataset/hk-censtatd-tablechart-energy /resource/50c98976-b7e5-4f21-bf0b-ec4af7a4a61e
2. https://data.gov.hk/tc-data/dataset/hk-censtatd-tablechart-energy/resource/cb31d820-4e5c-4dd5-b775-cbad5a95bc6d
3. https://www.hkelectric.com/en
4. CLP Power Hong Kong. (n.d.). https://www.clp.com.hk/en/index
5. https://www.emsd.gov.hk/energyland/tc/energy/energy_use/energy_scene.html
6. https://www.globalpetrolprices.com/Hong-Kong/electricity_prices/




```
> adf.test(c_month_s)

	Augmented Dickey-Fuller Test

data:  c_month_s
Dickey-Fuller = -11.986, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary

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

Appendix 1: Differencing table and adf.test result

```
> res = arima(c_month_s, order=c(0,0,4))
> res

Call:
arima(x = c_month_s, order = c(0, 0, 4))

Coefficients:
          ma1      ma2      ma3      ma4  intercept
      -0.6533  -0.1153  -0.1178  -0.1136     -2e-04
s.e.   0.0412   0.0476   0.0493   0.0350      0e+00

sigma^2 estimated as 0.003408:  log likelihood = 878.73,  aic = -1745.47
```

Appendix 2: Result of ARIMA model

```
> Box.test(res$residuals, type = "Box-Pierce", lag = 20)

	Box-Pierce test

data:  res$residuals
X-squared = 137.49, df = 20, p-value < 2.2e-16

>
> Box.test(res$residuals, type = "Ljung-Box", lag = 20)

	Box-Ljung test

data:  res$residuals
X-squared = 140.4, df = 20, p-value < 2.2e-16
```

Appendix 3:Box-Pierce test and a Ljung-Box test for arma

摘要輸出

迴歸統計	
R 的倍數	0.95937
R 平方	0.92039
調整的 R	0.92027
標準誤	1277.22
觀察值個	633

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	1.19E+10	11901189560	7295.6049	0
殘差	631	1.029E+09	1631282.084		
總和	632	1.293E+10			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	1275	101.65001	12.5430582	2.224E-32	1075.3888	1474.615293	1075.3888	1474.615
t	23.7291	0.2778114	85.41431319	0	23.18352691	24.2746206	23.183527	24.27462

Appendix 4: Result of the linear regression model

```
> summary(ress)

Call:
lm(formula = datt1 ~ datt2$total)

Residuals:
    Min      1Q  Median      3Q     Max
-8651.1 -1775.7    98.2  1747.5  6472.0

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -440.2297  1351.3924  -0.326    0.745
datt2$total    1.1873     0.0414  28.677   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2553 on 106 degrees of freedom
Multiple R-squared:  0.8858,	Adjusted R-squared:  0.8847
F-statistic: 822.4 on 1 and 106 DF,  p-value: < 2.2e-16
```

Appendix 5:Linear regression Result

Veriguide report