Time Series Analysis
Australian Quarterly Electricity Production 1950-2010

Shengyang Zhuang

Group 7
Ashot Hovsepyan, Kevin Johansson, Mohammed Rafid, Vlado Grujcic

KTH Royal Institute of Technology
szhuang@kth.se

May 19, 2022

Overview

Introduction to the data set

The chosen data set used in this project contains the values for the quarterly Australian electricity production from year 1956 to 2010.

The analysis was based on **R**. The source of the data comes from **key2stats**.

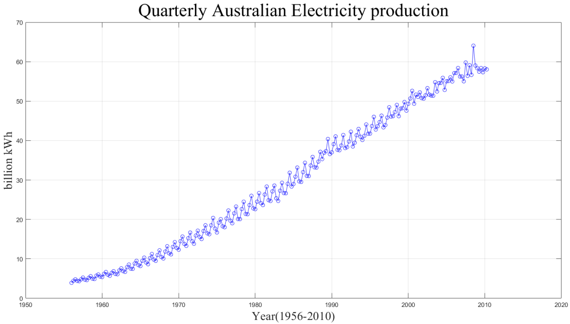
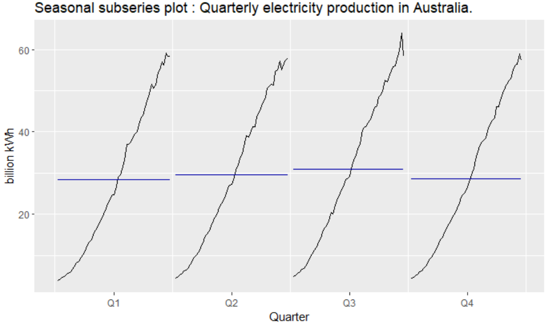


Figure: The quarterly Australian electricity production (kWh) from year 1956 to 2010

Preliminary analysis: trend and seasonal plot

A significant approach to emphasize the trend and seasonal patterns is to collect the data for each season (quarter) in separate mini time plots.



- Apparent increasing trend
- Strong seasonality $\Longrightarrow$ $\boxed{\text{Classical decomposition model}}$
- No abrupt changes and biased data

Trend and seasonality removal

Non-stationary data with seasonality $\Rightarrow$ Take a $\boxed{\text{seasonal difference}}$.
By applying a differencing operator with lag d = 4 we can obtain:

$$\nabla_4 X_t = X_t - X_{t-4} = (1 - B^4)X_t = m_t - m_{t-4} + Y_t - Y_{t-4}$$

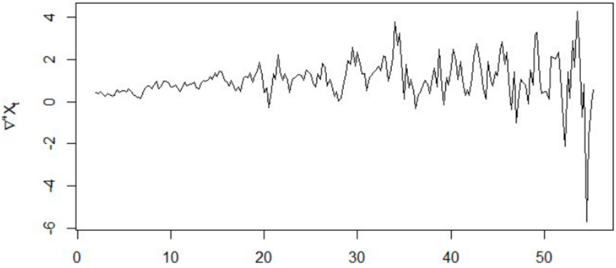


Figure: Time series of the lag-4 difference of the quarterly electricity production in Australia (Seasonality is removed)

Trend and seasonality removal

Still non-stationary due to trend component

Original data shows a linear increase $\Rightarrow$ Take an additional $\boxed{\text{first difference}}$

$$\nabla\nabla_4 X_t = \nabla(m_t - m_{t-4}) + \nabla(Y_t - Y_{t-4})$$

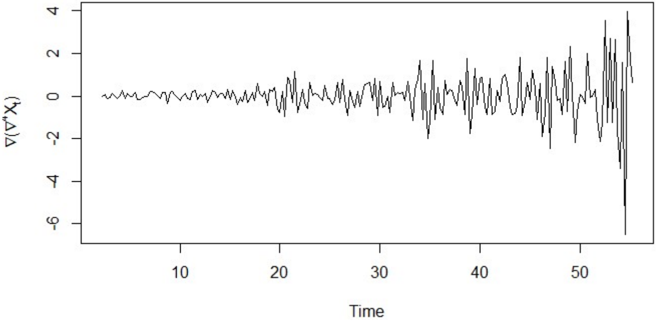


Figure: Time series of the first-difference of the lag-4 difference of the quarterly electricity production in Australia.

Trend and seasonality removal

To verify that the removal of trend and seasonality is successful, we use the Augmented Dickey-Fuller test (ADF test) to check the stationarity of the time series.

- Stationary if the test yields a p-value smaller than 0.05

```
Dickey-Fuller = -12.097, Lag order = 5, p-value = 0.01
alternative hypothesis:  stationary
```

Figure: ADF test result generated from RStudio

The result shows stationarity of $\hat{X}_t$, since the p-value is less than 0.05

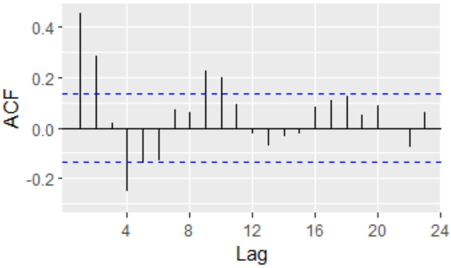
- $\hat{X}_t = \nabla\nabla_4 X_t$

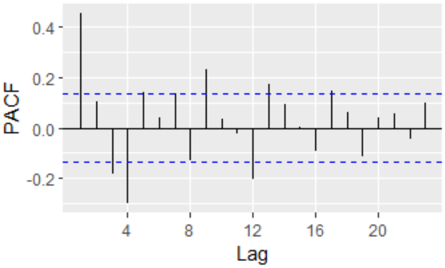
To find the fittest model, we mainly used two different approaches.

- Mannual method
- Automatic algorithm

SARIMA: Fitting model parameters manually

Before fitting the stationary time series $\hat{X}_t$ with a standard model, we have to first calculate both ACF and PACF.



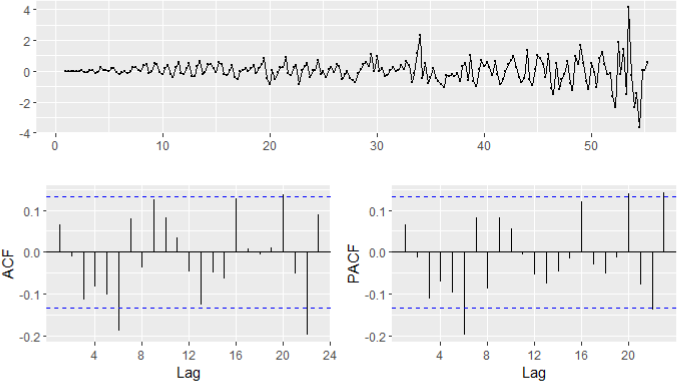


We begin with an ARIMA(0,1,1)(0,1,1)[4] model, since:

- Spike at lag 1 $\Longrightarrow$ suggests a non-seasonal MA(1) component
- Spike at lag 4 $\Longrightarrow$ suggests a seasonal MA(1) component

SARIMA: Fitting model parameters manually

Residuals from ARIMA(0,1,1)(0,1,1)[4] model:



- Spike at lag 6 $\Longrightarrow$ additional non-seasonal terms are needed

SARIMA: Fitting model parameters manually

The estimation of p, q and P, Q by ACF and PACF are based on empirical data.
Other models with nearby parameter values still need to be considered.

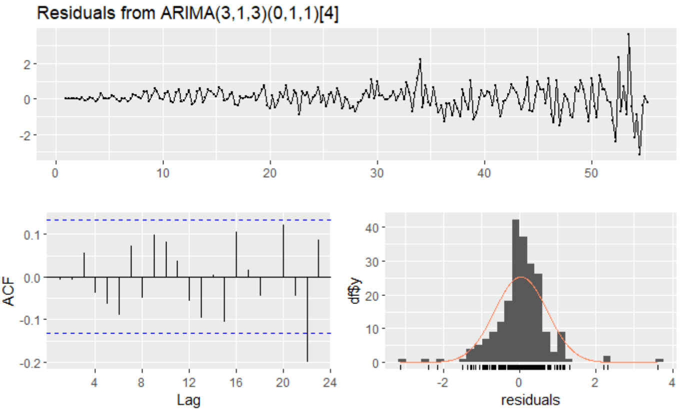
Model Structure	**AICc**	**BIC**	**Model Structure**	**AICc**	**BIC**
ARIMA(0,1,1)(0,1,1)[4]	483.48	493.45	ARIMA(2,1,1)(0,1,1)[4]	478.77	495.29
ARIMA(0,1,2)(0,1,1)[4]	482.08	495.34	ARIMA(2,1,2)(0,1,1)[4]	480.67	500.43
ARIMA(0,1,3)(0,1,1)[4]	479.36	495.88	ARIMA(2,1,3)(0,1,1)[4]	/	/
ARIMA(1,1,1)(0,1,1)[4]	479.74	492.43	ARIMA(3,1,1)(0,1,1)[4]	480.5	500.26
ARIMA(1,1,2)(0,1,1)[4]	479.65	496.17	ARIMA(3,1,2)(0,1,1)[4]	481.63	504.61
ARIMA(1,1,3)(0,1,1)[4]	480.35	500.11	ARIMA(3,1,3)(0,1,1)[4]	472.63	498.82

Table: Model structure with different parameters

Therefore, we should choose ARIMA(3,1,3)(0,1,1)[4] model.

SARIMA: Fitting model parameters manually

Residuals from ARIMA(3,1,3)(0,1,1)[4] model:



- still one spike left in ACF (correlated residuals)
- Ljung-Box test result: p-value=0.02615 (fail)

SARIMA:Fitting model parameters with `auto.arima()`

The `auto.arima()` function uses:

- `nsdiffs()` to determine D (the number of seasonal differences to use)
- `ndiffs()` to determine d (the number of ordinary differences to use)

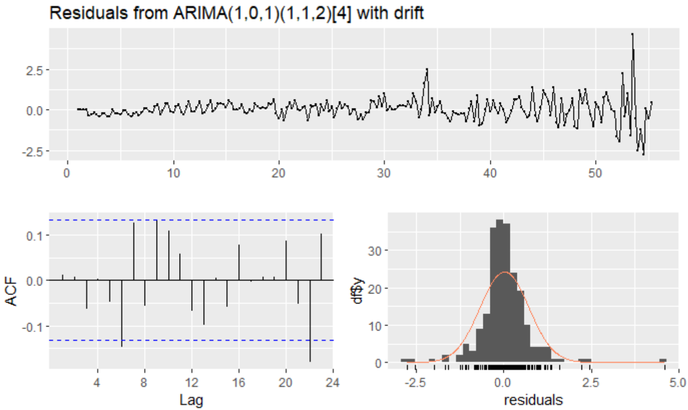
The selection of the other model parameters p, q and P, Q are all determined by minimizing the AICc values.

ARIMA(1,0,1)(1,1,2)[4] with drift						
Parameters	**ar1**	**ma1**	**sar1**	**sma1**	**sma2**	**drift**
Estimate	0.9158	−0.4411	0.8928	−1.6921	0.7737	0.2299
s.e.	0.0467	0.0942	0.1420	0.1425	0.0954	0.0456
sigma^2 = 0.5082: log likelihood = -230.58						
AIC=475.16 AICc=475.71 BIC=498.72						

Therefore, we should choose ARIMA(1,0,1)(1,1,2)[4] model.

SARIMA:Fitting model parameters with `auto.arima()`

Residuals from ARIMA(1,0,1)(1,1,2)[4] model:



- still several tiny spikes left in ACF (correlated residuals)
- Ljung-Box test result: p-value=0.003089 (fail)

SARIMA

Truth: None of the models considered here pass all of the residual tests!

In practice, we would normally use the best model we could find, even if it did not pass all of the tests.

ARIMA(3,1,3)(0,1,1)[4]

1. AICc = 472.63
2. p-valule = 0.02615
3. one spike, fail LB test

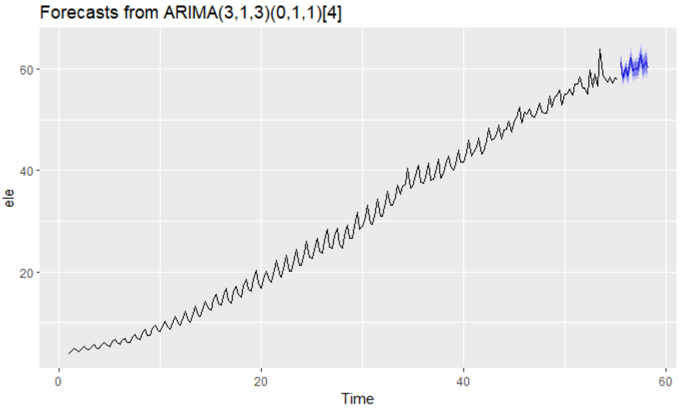
ARIMA(1,0,1)(1,1,2)[4]

1. AICc = 475.16
2. p-valule = 0.003089
3. several spikes, fail LB test

$\Longrightarrow$ Choose ARIMA(3,1,3)(0,1,1)[4].
The model can still be used for forecasting, but the prediction intervals may not be accurate due to the correlated residuals.

Sometimes it is just not possible to find a model that passes all of the tests!

Forecasting SARIMA process

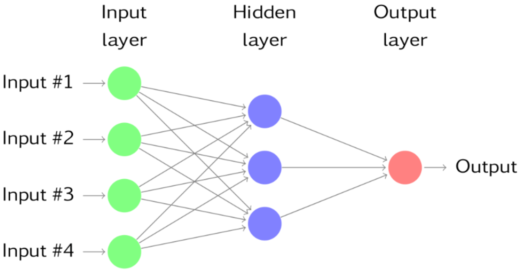


The forecast follows the historical pattern and seems to be relatively stable and the prediction intervals allow for the data to trend upwards and downwards during the forecast period.

Conclusion: Difficulties and further discussion

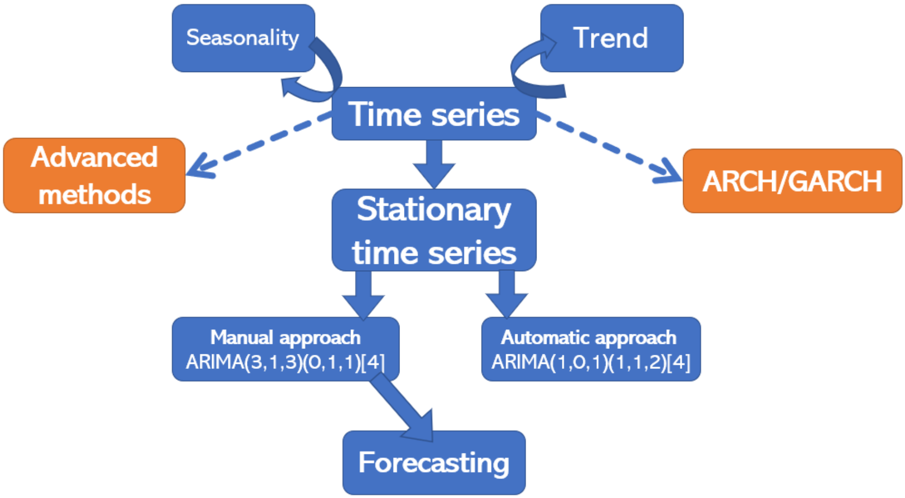
Truth: the model with the smallest AICc value can not pass all the tests.
$\Longrightarrow$ We might have to consider **ARCH or GARCH models**.

Alternatively, **Neural network models**. Ex: NNAR model



Lagged values of the time series $\Rightarrow$ inputs of the neural network

Conclusion: results



References

Hyndman, R.J., Athanasopoulos, G. (2018)
Forecasting: principles and practice, 2nd edition

Holmes, E. E., M. D. Scheuerell, and E. J. Ward. (2021)
Applied time series analysis for fisheries and environmental data

P.J. Brockwell and R.A. Davis
Introduction to Time Series and Forecasting, Springer

The End