

Laporan Pengerjaan Kuis 1

Metode Peramalan Deret Waktu

Nama : Muhammad Ikhsan Ananda

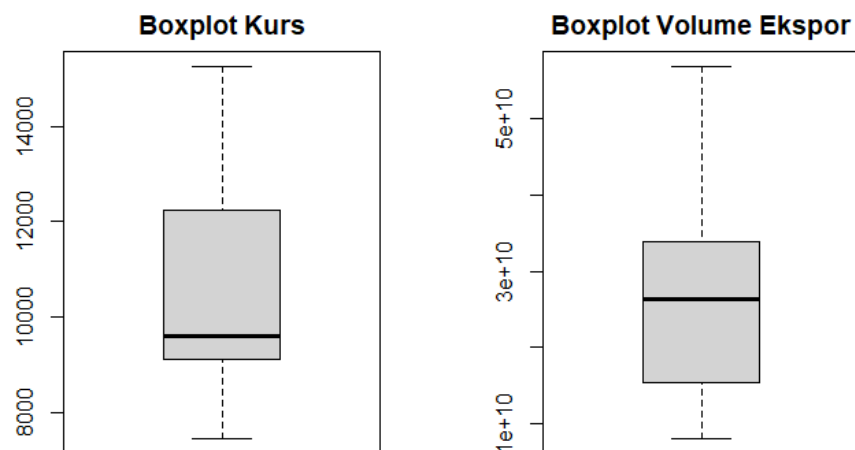
NIM : G64190032

▪ Eksplorasi Data

Tahap pertama adalah eksplorasi data. Tahapan ini diperlukan untuk mengetahui karakteristik dari data yang saya miliki. Berikut adalah hasil summary dari eksplorasi data yang telah dilakukan :

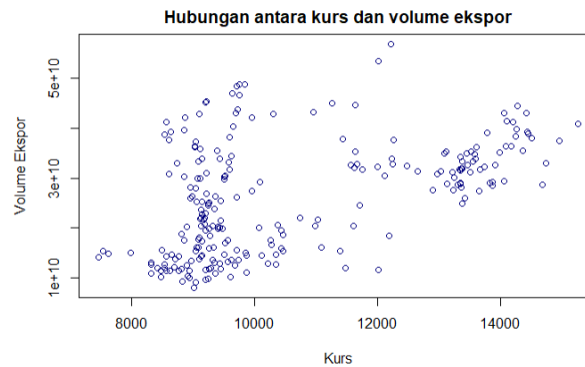
Parameter	Tahun (int)	Kurs (int)	Volume Ekspor (double)
N data	237	237	237
Null/NA	0	0	0
Min	2000	7457	8.131×10^9
Median	2009	9597	2.630×10^{10}
Mean	2009	10593	2.611×10^{10}
Max	2019	15259	5.686×10^{10}

Selanjutnya adalah upaya untuk mengetahui apakah terdapat outlier dalam data. Pendeteksian outlier dilihat melalui visualisasi boxplot. Outlier yang diamati adalah outlier pada kolom kurs dan volume ekspor, sebab kedua kolom tersebut adalah kolom utama yang dijadikan pengamatan. Hasilnya tidak ada outlier yang ditemukan seperti terlihat pada boxplot berikut :

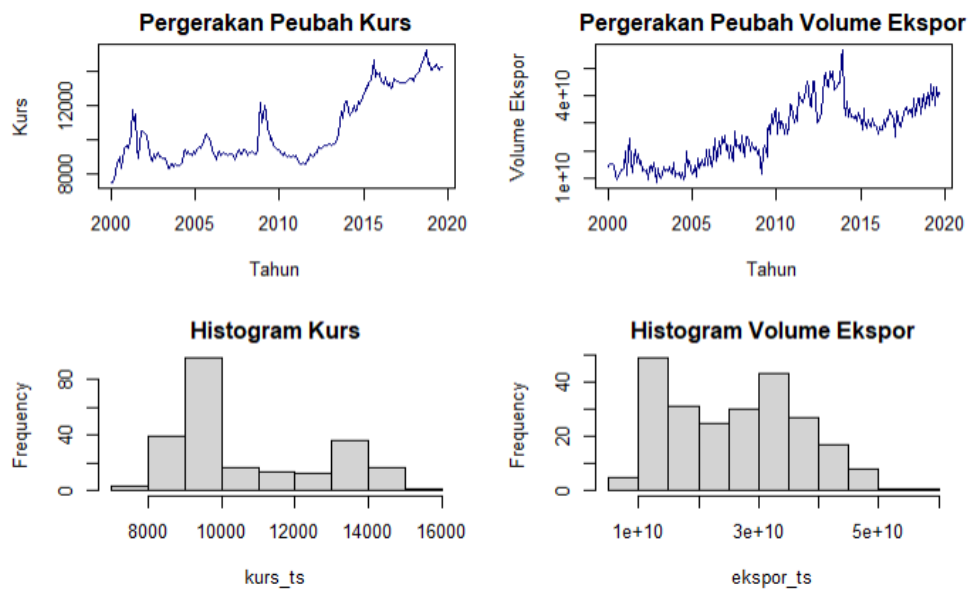


Berikutnya adalah upaya untuk mengetahui apakah peubah kurs dan volume ekspor memiliki keterkaitan serta mengetahui koefisien korelasi dari kedua peubah tersebut. Hipotesis yang saya gunakan adalah H_0 kedua peubah saling bebas dan H_1 kedua peubah tidak saling bebas. Hasilnya, p-value menggunakan metode two sample t-test bernilai

4.259872×10^{-99} dan koefisien korelasi pearson sebesar 0.476046. Melalui p-value yang lebih kecil dari 0.05 dan koefisien korelasi pearson, saya menyimpulkan bahwa kedua peubah saling bebas dan korelasinya termasuk kedalam *moderate possitive correlation*. Pada visualisasi berikut juga menunjukkan bahwa memiliki ketimpangan yang tinggi sehingga grafik menjadi tidak representatif walaupun secara umum diketahui pola sebarannya adalah linear.



Selanjutnya adalah melihat pergerakan pada peubah kurs dan volume ekspor. Pergerakan data kedua peubah divisualisasikan seperti berikut :

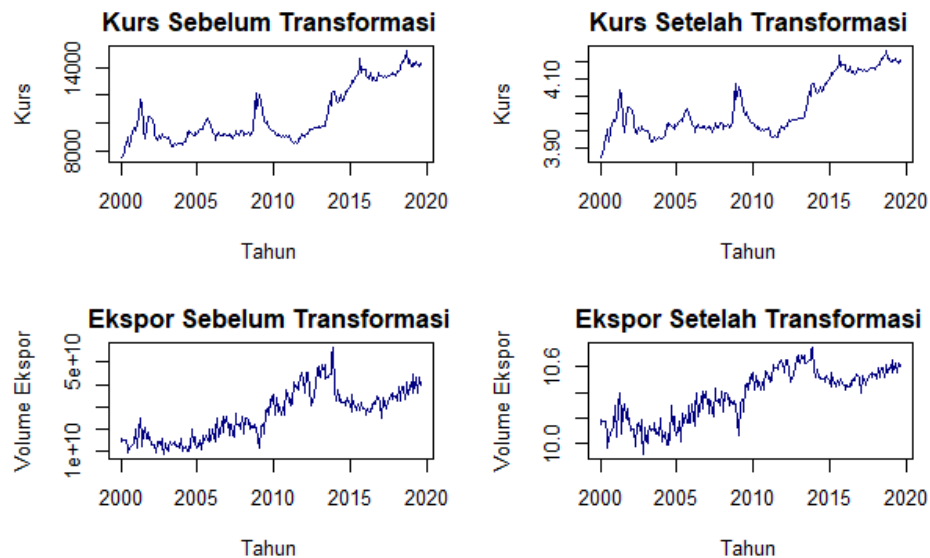


Berdasarkan visualisasi diatas saya menyimpulkan beberapa hal sebagai berikut :

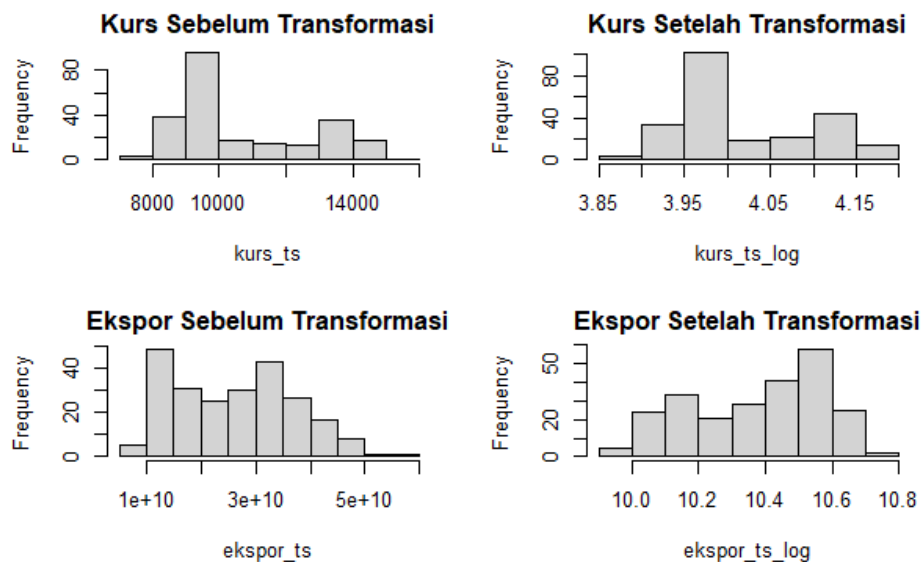
- Pada visualisasi pergerakan peubah kurs terlihat bahwa pola data deret waktu yang terjadi adalah pola data siklik. Selain itu, pada histogram kurs distribusi yang terjadi adalah positif skew.
- Pada visualisasi pergerakan peubah volume ekspor terlihat bahwa pola data deret waktu yang terjadi adalah pola data trend positif. Selain itu, pada histogram volume kurs distribusi yang terjadi adalah multimodal.

- Transformasi Logaritma

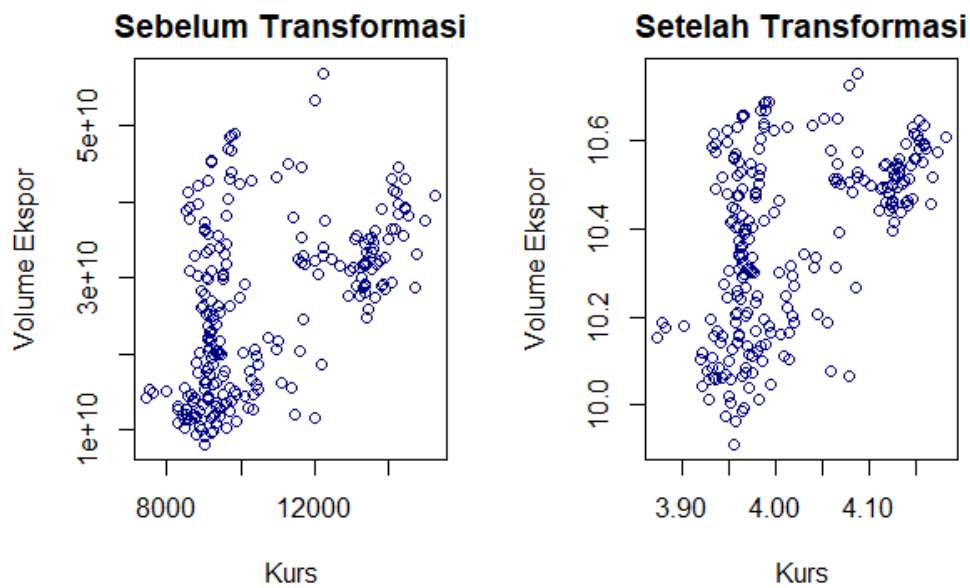
Transformasi logaritma dilakukan sebagai akibat dari visualisasi peubah yang terlihat tidak linear. Transformasi logaritma dilakukan terhadap dua peubah yaitu kurs dan volume ekspor sehingga didapatkan hasil sebagai berikut :



Berdasarkan visualisasi diatas saya menyimpulkan bahwa range data berubah ketika data telah dilakukan transformasi logaritma. Namun, apabila dilihat dari pola data deret waktu antara data sebelum dan sesudah transformasi tidak terlihat perbedaan yang signifikan sehingga pola data deret waktu tetap sama baik sebelum maupun sesudah transformasi.



Berdasarkan histogram diatas saya menyimpulkan bahwa range data berubah ketika data telah dilakukan transformasi logaritma. Namun, apabila dilihat dari distribusi antara data sebelum dan sesudah transformasi tidak terlihat perbedaan yang signifikan sehingga distribusi data tetap sama baik sebelum maupun sesudah transformasi.



Berdasarkan visualisasi diatas saya menyimpulkan bahwa tidak terdapat perubahan yang berarti pada hubungan kedua peubah sebelum dan sesudah transformasi. Pola yang terlihat adalah kedua peubah menghasilkan korelasi yang positif dan terlihat linear walaupun tidak termasuk *perfect linear*. Selain itu, terlihat juga bahwa ketika nilai kurs semakin besar, maka volume ekspor perlahan akan meningkat.

■ Pembagian Data Training dan Data Testing

Data training dan testing dibagi dengan presentase untuk training sebesar 80% dan testing sebesar 20% sehingga menghasilkan training sebesar 190 data pertama dan testing sebesar 47 data terakhir.

	Tahun	Bulan	Kurs.Rp.US.	Volume.ekspor.neto.kg.
1	2000	Jan	3.872564	10.154017
2	2000	Feb	3.877199	10.186408
3	2000	Mar	3.882069	10.174216
4	2000	Apr	3.901840	10.179314
5	2000	May	3.937117	10.168106
6	2000	Jun	3.942851	10.062591
7	2000	Jul	3.955928	9.962903
8	2000	Aug	3.920228	10.042412
9	2000	Sep	3.945074	10.070235
10	2000	Oct	3.974374	10.111277

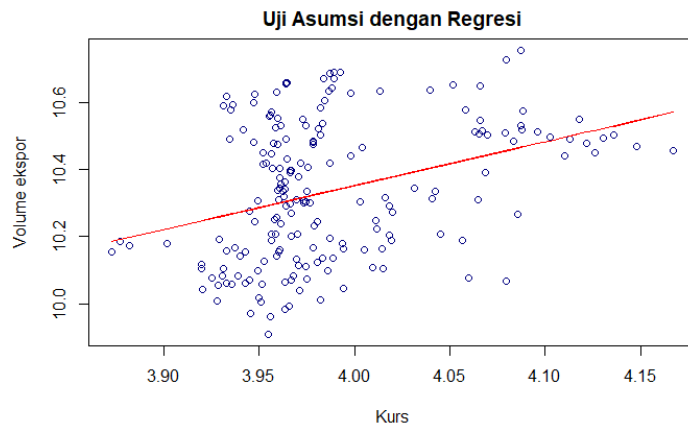
Data training

	Tahun	Bulan	Kurs.Rp.US.	Volume.ekspor.neto.kg.
191	2015	Nov	4.142139	10.46501
192	2015	Dec	4.140728	10.46238
193	2016	Jan	4.142327	10.45367
194	2016	Feb	4.127979	10.41367
195	2016	Mar	4.124113	10.45861
196	2016	Apr	4.121757	10.43981
197	2016	May	4.135037	10.43917
198	2016	Jun	4.120969	10.49506
199	2016	Jul	4.118132	10.46080
200	2016	Aug	4.124895	10.50241

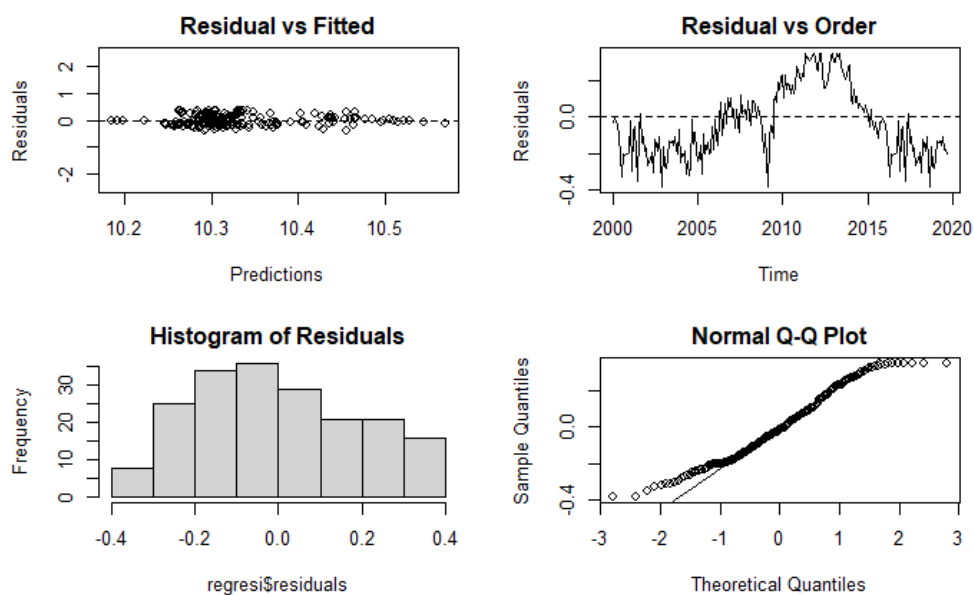
Data testing

■ Analisis Regresi Untuk Mengetahui Autokorelasi Sisaan

Analisis regresi sederhana pada data training yang telah dilakukan tranformasi logaritma memberikan hasil persamaan $\hat{y} = 5.1355 + 1.3040x + 0.1916$ dan menghasilkan r square sebesar 0.122.



Selanjutnya, saya mencoba untuk melihat apakah ada autokorelasi pada sisaan sehingga menghasilkan visualisasi sebagai berikut :

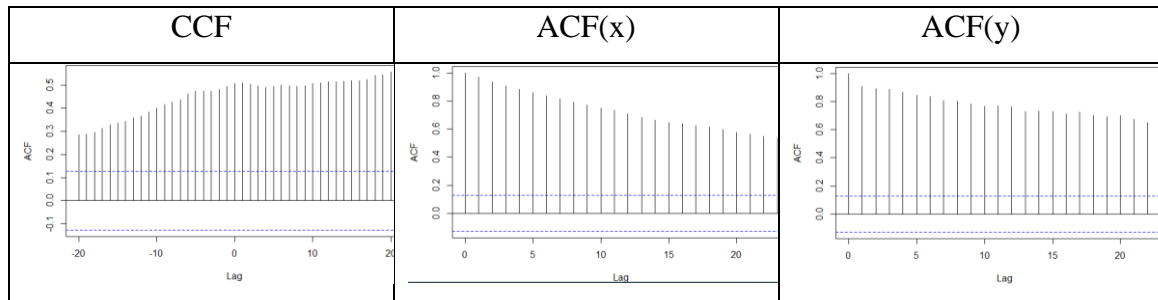


Berdasarkan visualisasi diatas terdapat auto korelasi positif pada residual vs order. Kemudian untuk melihat apakah sisaan menyebar normal. Pengujian dilakukan dengan metode Uji Jarque Bera dengan H_0 adalah sisaan menyebar normal dan H_1 adalah sisaan tidak menyebar normal. Hasilnya adalah nilai p-value sebesar 0.022 sehingga lebih kecil dari 0.05 dan tolak H_0 (sisaan tidak menyebar normal). Selain itu, pada Uji Durbin Watson menghasilkan nilai dw sebesar 0.2349551 yang dimana nilai tersebut lebih kecil dari 2 dan mendekati 0 sehingga terdapat autokorelasi positif.

Berikutnya untuk menghilangkan autokorelasi pada data saya menggunakan prosedur cochrane-ortcutt sehingga memberikan hasil persamaan $\hat{y} = 11.42847 - 0.26820x + 0.087$ dengan r square sebesar 0.0023 dan dw sebesar 2.68737. Nilai dw yang dihasilkan telah mendekati 2 sehingga sudah tidak ada autokorelasi.

- Analisis Regresi dengan Peubah Lag dan Nilai Akurasi Model

Hal pertama yang saya lakukan adalah melakukan visualisasi cross correlation function. Visualisasi tersebut untuk mengukur koherensi antara kedua peubah ketika salah satu peubah digeser dalam waktu relatif terhadap peubah lainnya. Visualisasi tersebut menghasilkan lag optimum 5 pada ccf, 1 pada acf (x), dan 1 pada acf(y).



Berikutnya adalah proses penentuan model dan pengujian model. Penentuan model dilakukan dengan membandingkan antara model Koyck, Polynomial, dan Autoregressive, sementara untuk pengujian hasil terbaik adalah dengan membandingkan nilai MAPE yang seluruhnya terlihat sebagai berikut :

Model	Persamaan	MAPE
Kyock	$y_t = 0.687 + 0.895Y.1 + 0.101X.t + 0.087$	0.00534
Polynomial	$y_t = 4.316 + 0.967X.t0 - 1.184X.t1 + 0.245X.t2 + 0.193$	0.00465
Autoregressive	$y_t = 0.722 + 0.899Y.1 - 0.315X.t + 0.397X.1 + 0.087$	0.00646

Ukuran keakuratan model yang digunakan adalah MAPE sebab nilai MSD dan MAD sudah berbanding lurus dengan MAPE. Nilai MAPE yang semakin kecil menandakan model yang terbaik. Oleh karena itu, saya memilih model polynomial sebab menghasilkan nilai MAPE terkecil dibandingkan model Kyock dan Autoregressive yaitu sebesar 0.00465 atau 0.4%.

- Dokumentasi

Dokumentasi pengerjaan laporan ini terdapat pada link berikut : [ikhsanananda/Kuis_MPDW \(github.com\)](https://github.com/ikhsanananda/Kuis_MPDW)