

ANALISIS RUNTUN WAKTU HARGA ALPUKAT DENGAN MENGGUNAKAN METODE ARIMA DAN SEASONAL NAIVE



Disusun oleh:

Kelompok 5

1. Abdul Muhaimin (2106703052)
2. Andri Naufal Al'afif (2106725085)
3. Justin Martinus (2106724832)

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA**

2023

DAFTAR ISI

BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan.....	2
1.4 Manfaat Penelitian.....	2
BAB II METODE PENELITIAN.....	3
2.1 Data.....	3
2.2 Preprocessing Data.....	3
2.3 Pengecekan Data.....	4
2.4 Metode Analisis Data.....	8
BAB III HASIL DAN PEMBAHASAN.....	10
3.1 Uji Stasioneritas.....	10
3.2 Spesifikasi Model.....	11
3.3 Overfitting Model.....	14
3.4 Penentuan Model Terbaik.....	16
3.5 Estimasi Parameter.....	16
3.6 Model Diagnostik (Analisis Residual).....	18
3.7 Forecasting.....	22
3.8 Analisis dan Modeling Lanjutan Tanpa Outlier.....	30
3.9 Forecasting Lanjutan.....	41
BAB IV PENUTUP.....	43
4.1 Kesimpulan.....	43
DAFTAR PUSTAKA.....	45
LAMPIRAN.....	46

BAB I PENDAHULUAN

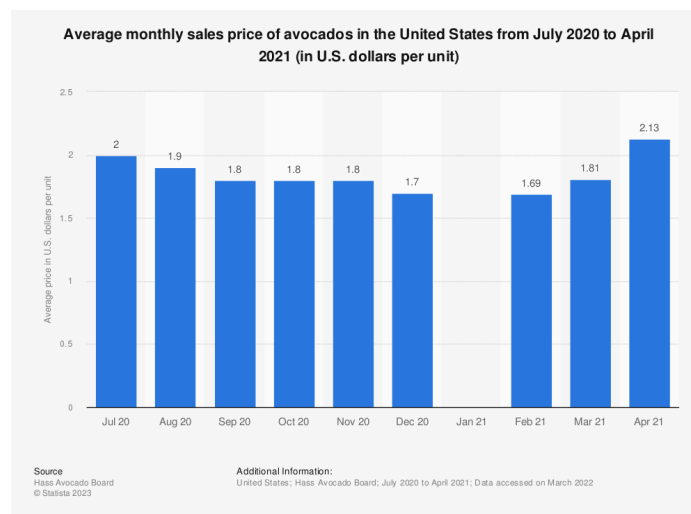
1.1 Latar Belakang

Alpukat (*Persea americana Mill*) adalah tanaman yang berasal dari Amerika Tengah dan Meksiko Selatan. Tanaman ini termasuk dalam suku *Lauraceae* dan menghasilkan buah yang kaya akan sumber energi. Alpukat juga dikenal sebagai salah satu buah yang memiliki kandungan nutrisi yang sangat baik karena mengandung lemak, protein, dan serat dalam jumlah yang signifikan. Selain itu, alpukat juga merupakan sumber yang kaya akan vitamin seperti vitamin C, E, K, B1, B2, B6, dan B9, serta mengandung mineral seperti fosfor, sodium, magnesium, kalium, zat besi, dan zinc.

Alpukat telah menjadi makanan yang sangat populer di Amerika Serikat dalam beberapa tahun terakhir. Permintaan yang tinggi disebabkan oleh popularitas alpukat sebagai makanan yang sehat dan bergizi. Konsumen semakin mengenal manfaat kesehatan alpukat, seperti kandungan lemak sehat dan serat yang tinggi. Permintaan yang meningkat telah mendorong harga alpukat di pasar.

Sebagian besar alpukat yang dikonsumsi di Amerika Serikat berasal dari California, Florida, dan Meksiko. Produksi alpukat memiliki musim yang terbatas, yang dapat menyebabkan fluktuasi harga akibat keterbatasan pasokan. Bencana alam seperti kekeringan, banjir, atau hama juga dapat mempengaruhi pasokan alpukat dan berdampak pada harga. Produksi alpukat memerlukan biaya yang tinggi, termasuk perawatan tanaman yang intensif, pengairan yang memadai, dan perlindungan dari hama dan penyakit. Biaya produksi yang tinggi dapat memengaruhi harga jual alpukat di pasar.

Kemudian, variasi iklim seperti suhu, curah hujan, dan cuaca ekstrem dapat mempengaruhi kualitas dan kuantitas panen alpukat. Perubahan iklim dan kondisi cuaca yang tidak menentu dapat memengaruhi produksi alpukat dan menyebabkan fluktuasi harga. Berikut data mengenai harga rata-rata setiap alpukat per bulan di Amerika Serikat.



Berdasarkan gambar di atas, terlihat bahwa pada April 2021 memiliki rata-rata penjualan alpukat tertinggi sejumlah \$2.13 per unit dan terendah pada Desember 2020 sejumlah \$1.7 per unit.

Project ini menggunakan dataset Price of Avocados yang meneliti harga rata-rata penjualan alpukat. Studi ini dilakukan pada 4 Januari 2015 hingga 25 Maret 2018. Oleh karena itu, kami akan menganalisis data tersebut untuk melakukan *forecasting* dengan menggunakan metode ARIMA dan *Seasonal Naive*.

1.2 Rumusan Masalah

- 1) Apakah terdapat pola dan tren harga alpukat di Amerika Serikat pada periode waktu mingguan, bulanan, dan tahunan?
- 2) Bagaimana penerapan metode ARIMA dalam memprediksi ramalan harga alpukat di masa depan?
- 3) Bagaimana keakuratan ramalan yang didapatkan dengan metode ARIMA?
- 4) Apakah model dan hasil yang dibuat untuk setiap tipe alpukat berbeda?
- 5) Bagaimana penerapan dan keakuratan metode *seasonal naive* dalam memprediksi ramalan harga alpukat di masa depan?

1.3 Tujuan

- 1) Menemukan pola dan tren tertentu dari harga alpukat dengan metode peramalan.
- 2) Memprediksi harga alpukat di masa depan dengan metode ARIMA dan *seasonal naive*.
- 3) Memastikan model yang dibuat cukup akurat dan dapat diandalkan untuk prediksi selanjutnya.
- 4) Menemukan perbedaan dari model dan hasilnya pada tipe alpukat yang berbeda.
- 5) Mengetahui penerapan dan keakuratan metode *seasonal naive* dalam memprediksi ramalan harga alpukat di masa depan

1.4 Manfaat Penelitian

Melalui penelitian ini, penjual alpukat atau bahkan regulator dari distribusi alpukat, bahkan produsen alpukat dapat menggunakan prediksi harga alpukat di masa depan untuk menyesuaikan kebijakan mereka terkait proses penjualan alpukat. Selain itu, dengan adanya penelitian ini, dapat ditemukan beberapa kecenderungan pada pola dan tren harga terhadap waktu dan dihubungkan dengan faktor lainnya pada waktu tersebut untuk mendapat *insight* yang penting dalam mengendalikan harga yang terus berubah.

BAB II METODE PENELITIAN

2.1 Data

Dataset Price of Avocados yang kami gunakan memiliki poin penting dalam penyajian data yang perlu disinggung. Data tersebut memiliki rentang waktu pengambilan data mingguan. Selain itu, data mengambil harga alpukat dari banyak toko yang tersebar pada negara bagian Amerika Serikat yang berbeda. Tidak hanya itu, data tersebut juga memiliki tipe buah alpukat, yaitu *organic* dan *conventional*. Dengan demikian, data yang kami gunakan dapat dibagi menjadi masing-masing tipe buah alpukat dengan karakteristik memiliki banyak baris yang memiliki tanggal yang sama dengan format tanggal Tahun-Bulan-Hari, tetapi diambil mingguan sehingga masih berupa data yang belum bisa dijadikan runtun waktu secara langsung.

Data kami memiliki 18249 pengamatan dengan 14 kolom, dimana kami mengambil hanya kolom “Date”, “AveragePrice”, dan “type” sebagai fokus kami dalam penelitian runtun waktu ini.

2.2 Preprocessing Data

Kami memastikan kolom “Date” menjadi format tanggal agar bisa dibuat runtun waktu, lalu mereduksi dimensi data menjadi kolom yang hanya kami inginkan dan membagi data ke tiga variabel yang berbeda untuk menyimpan setiap tipe alpukat untuk kemudian dianalisis secara terpisah. Ketiga variabel tersebut dipisah sebagai secara keseluruhan (“Overall”), tipe organik saja, dan tipe konvensional saja. Kami juga membuat kolom “type” menjadi faktor untuk kemudian dibuat grafik.

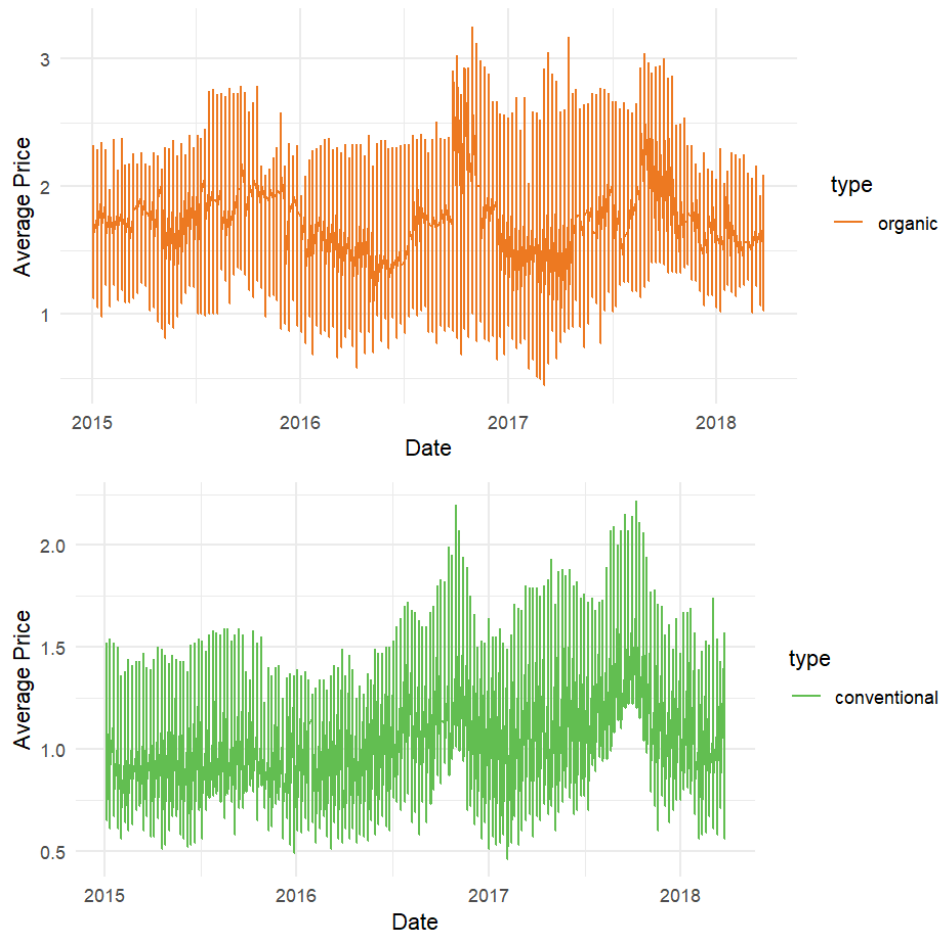
Meski dilakukan setelah membuat grafik, kami membuat data menjadi bentuk runtun waktu dengan rata-rata setiap negara bagian yang berbeda pada tanggal yang sama menjadi satu nilai untuk satu waktu agar tidak terjadi galat pada analisis runtun waktu. Kami juga menyesuaikan untuk setiap tujuan kode kami di R dengan membuat runtun waktu dengan format bulanan dan bulanan yang terpisah dengan setiap tahunnya pada pembuatan grafik.

```
> sum(is.na(df))  
[1] 0
```

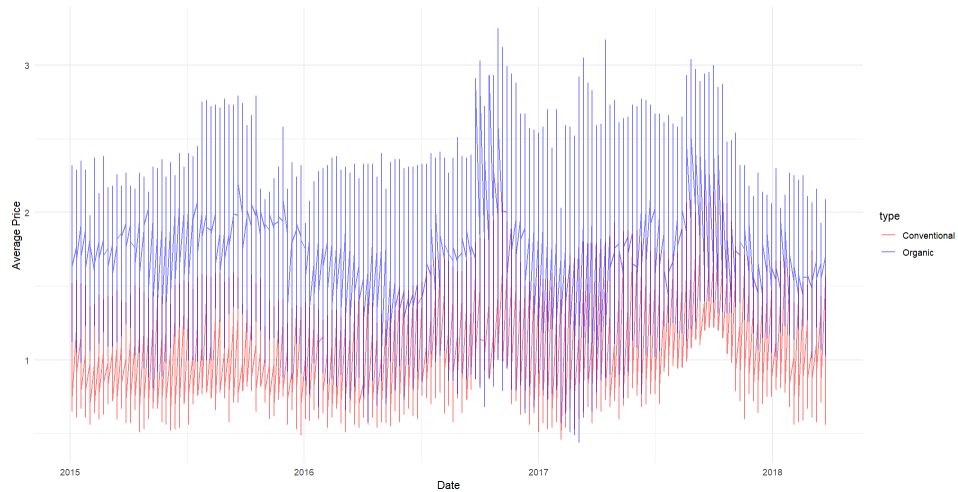
Dapat dilihat dengan R bahwa dataset kami tidak memiliki *missing values* sehingga dapat kami tangani setelah melakukan *pre-processing* sebelumnya untuk menangani karakteristik data yang berasal dari negara bagian yang berbeda dan siklus mingguan untuk setiap pengamatan, serta membagi ke setiap tipe dan keseluruhan buah alpukat.

2.3 Pengecekan Data

Stasioneritas runtun waktu merupakan langkah yang cukup penting dalam analisis runtun waktu sehingga kami membuat grafik terlebih dahulu untuk menentukan apakah dataset yang kami gunakan layak dilakukan analisis runtun waktu dan apakah dataset sudah stasioner. Kemudian, dilakukan hal yang serupa dengan dataset yang sudah dilakukan *differencing* satu kali.



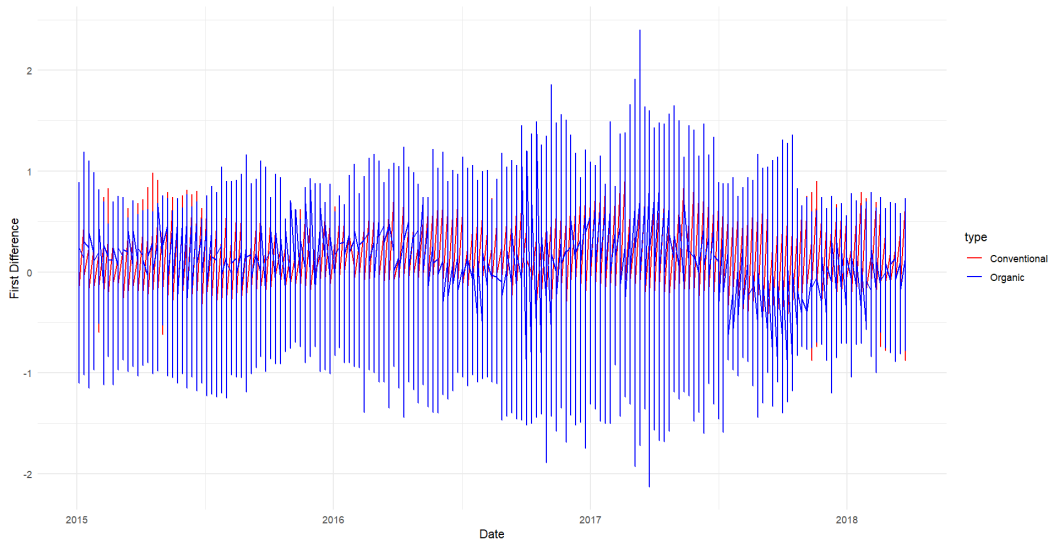
Dengan grafik yang dibuat, terlihat jelas persebaran setiap observasi pada tanggal yang sama dan berbeda dari tipe *organic* dan *conventional*. Persebaran ini terlihat relatif stabil pada tahun 2015 hingga 2016, tetapi memiliki kenaikan cukup besar pada tepat sebelum tahun 2017 dan 2018 yang terlihat lebih jelas pada tipe *conventional*. Untuk mempermudah visualisasi untuk data keseluruhan dan yang terpisah secara sekaligus, dibuat grafik yang menggabungkan keduanya.



Secara keseluruhan masing terlihat kecenderungan naiknya puncak harga pada tepat sebelum tahun 2017 dan 2018, konsisten dengan grafik pada setiap tipe saat dilihat terpisah. Selanjutnya akan dilihat untuk dataset yang sudah dilakukan *differencing* satu kali.



Grafik dari data yang sudah dilakukan *differencing* terlihat lebih stabil, tidak memiliki beberapa puncak yang sangat tinggi, sebagaimana pada plot dari data yang belum dilakukan *differencing*. Akan dilihat juga saat keduanya digabung sebagai grafik data keseluruhan.

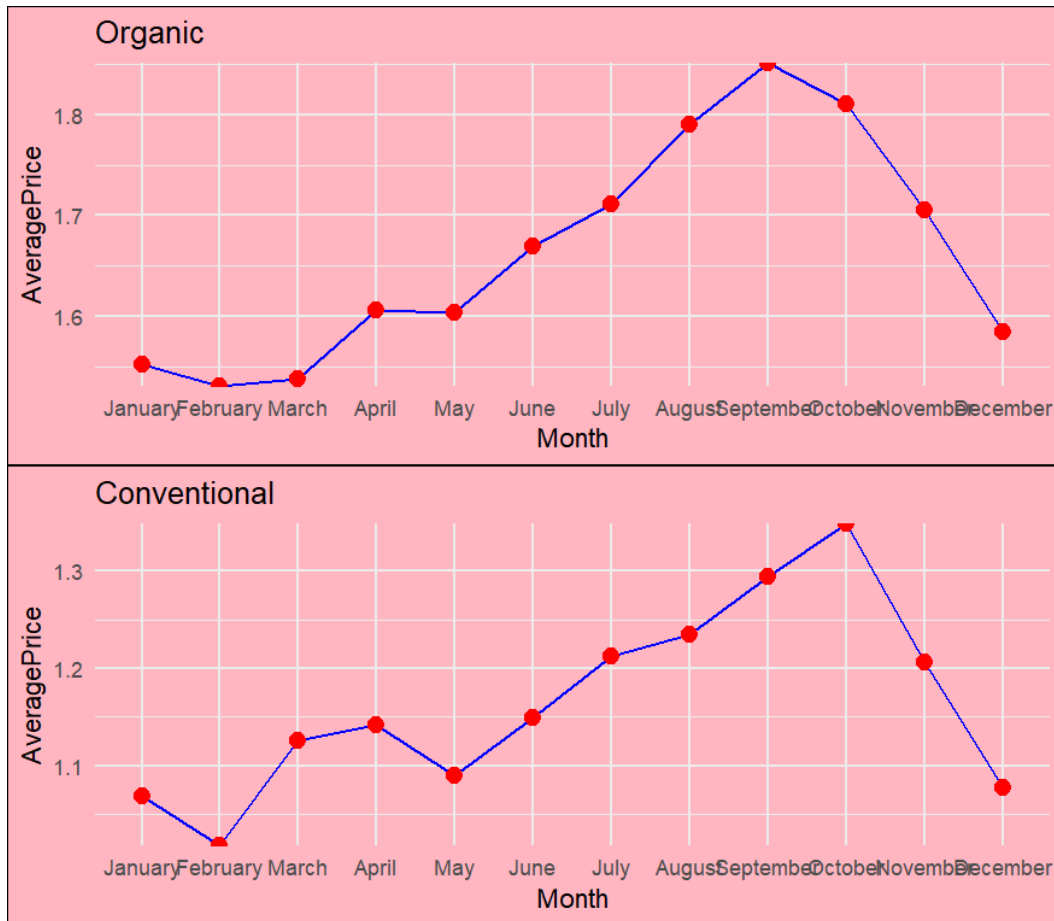


Grafik keseluruhan dari *differencing* juga memberi kesimpulan yang konsisten dengan grafik terpisah dari terlihatnya hanya sedikit puncak yang berbeda jauh dari sisanya.

Selain grafik untuk stasioneritas, akan dibuat juga *line plot* untuk mengecek kecenderungan karakteristik *seasonality* yang mungkin ada pada dataset. Grafik akan dibuat dengan format yang mirip dengan sebelumnya, dibagi menjadi setiap tipe, dibandingkan, dan digabungkan. Untuk mempermudah dan memperjelas visualisasi, dibuat visualisasi bulanan dari mingguan dengan rata-rata setiap minggunya pada bulan yang sama.



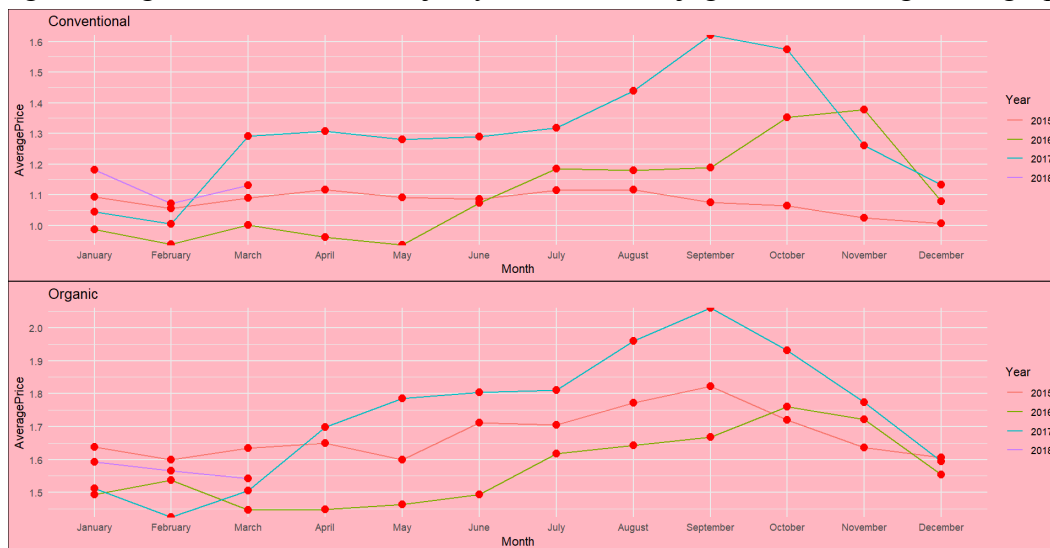
Grafik untuk keseluruhan harga alpukat menunjukkan kenaikan drastis tiap pergantian bulan, mulai saat transisi dari bulan Mei dan berpuncak bulan Oktober, baru kemudian turun lagi. Untuk mendapat gambaran yang lebih jelas lagi, akan dipisah dari setiap tipe alpukat.



Pada kedua jenis, terlihat tren yang sama dengan keseluruhan, tetapi untuk jenis organik memiliki puncak di September, sedangkan konvensional berpuncak di Oktober. Sekilas terlihat terdapat tren *seasonality*, tetapi bisa saja disebabkan *outliers* pada tahun tertentu sehingga kami memastikannya dengan memisahkan plot per tahun juga kali ini.



Terlihat bahwa tahun 2017 memiliki kenaikan paling drastis pada bulan September dan 2016 juga berkontribusi dalam peningkatan pada bulan Oktober, tetapi tidak begitu jauh dibandingkan dengan tahun 2017. Selanjutnya, akan dilihat juga untuk masing-masing tipe.



Semakin jelas bahwa tahun 2017 memberikan rata-rata harga yang sangat tinggi dibanding tahun-tahun sebelumnya pada bulan yang memang cenderung mengalami kenaikan harga. Kami akan menganalisis data runtun waktu ini secara keseluruhan dengan tetap mengikutsertakan tahun 2017, tetapi kami lakukan juga akan menganalisis singkat terkait data yang tidak mengikutsertakan tahun 2017 dengan anggapan *outlier*.

2.4 Metode Analisis Data

Metode analisis data yang digunakan adalah dengan menentukan model terbaik dengan metode ARIMA (Autoregressive Integrated Moving Average) dan *Seasonal Naive* untuk meramalkan harga alpukat pada 3 bulan kedepan (ARIMA) dan setahun kedepan (*Seasonal Naive*). Langkah-langkah analisisnya sebagai berikut :

- Mengidentifikasi kestasioneran data melalui plot dan uji Augmented Dickey Fuller (stasioneritas) pada setiap tipe.
- Apabila data tidak stasioner akan dilakukan differencing/transformasi, hingga data setiap tipe stasioner.
- Melakukan spesifikasi model dengan menduga model ARMA(p,q) dari grafik ACF, PACF, ataupun EACF jika diperlukan. Berikut merupakan acuan pola ACF dan PACF.
 - Jika pola ACF menyusut secara eksponensial atau pola gelombang sinusoidal yang tidak begitu jelas dan pola PACF ada tiang pancang sampai lag p, maka model adalah AR(p).
 - Jika pola ACF ada tiang pancang yang jelas sampai lag q dan pola PACF menyusut secara eksponensial, maka model adalah MA(q).

- Jika pola ACF dan PACF menyusut secara eksponensial maka model adalah ARMA (p,q).
- d. Melakukan overfitting model dengan membandingkan model-model ARIMA yang memenuhi dalam bentuk tabel. Kemudian, diperoleh model terbaik dari setiap tipe dengan AIC dan BIC terkecil (utamakan AIC terkecil).
 - e. Memperoleh model terbaik dari setiap tipe berdasarkan langkah sebelumnya.
 - f. Melakukan estimasi parameter dari model terbaik setiap tipe tersebut.
 - g. Melakukan model diagnostik untuk melihat kecocokan model melalui dua pendekatan yaitu analisis residual (independensi dan normalitas). Apabila, didapat model kurang baik, akan dilakukan penghapusan outlier pada data tahun yang memiliki banyak outlier.
 - h. Melakukan peramalan (forecasting) pada data aktual dan prediksi dengan menggunakan metode ARIMA dan *Seasonal Naive*.
 - i. Melakukan analisis lanjutan, yaitu mengulangi langkah a-h, tetapi dengan menghapus data tahun yang memiliki banyak outlier.

BAB III HASIL DAN PEMBAHASAN

3.1 Uji Stasioneritas

Hal pertama yang perlu dilakukan adalah menjadikan ketiga jenis data alpukat sebelumnya ke bentuk runtun waktu. Setelah itu, dilakukan uji hipotesis untuk stasioneritas dari runtun waktu yang sudah dibuat dengan tes Augmented Dickey-Fuller.

- Hipotesis

$$H_0: \pi = 0 \text{ (data tidak stasioner)}$$

$$H_1: \pi < 0 \text{ (data stasioner)}$$

- Tingkat Signifikansi

$$\alpha = 0.05$$

- Statistik Uji

$$\tau = \frac{\hat{\theta} - 1}{Se(\hat{\theta})} = \frac{\hat{\pi}}{Se(\hat{\pi})}$$

Didapatkan hasil dari masing-masing tipe keseluruhan, organik, dan konvensional secara berurutan seperti demikian:

```
> adf.test(ts1)
```

```
Augmented Dickey-Fuller Test
```

```
data: ts1
Dickey-Fuller = -2.7988, Lag order = 5, p-value = 0.2428
alternative hypothesis: stationary
```

```
> adf.test(ts2)
```

```
Augmented Dickey-Fuller Test
```

```
data: ts2
Dickey-Fuller = -2.2497, Lag order = 5, p-value = 0.4721
alternative hypothesis: stationary
```

```
> adf.test(ts3)
```

```
Augmented Dickey-Fuller Test
```

```
data: ts3
Dickey-Fuller = -3.1499, Lag order = 5, p-value = 0.0985
alternative hypothesis: stationary
```

- Aturan Keputusan

H_0 ditolak jika $p - value < \alpha = 0.05$. Terlihat bahwa ketiganya memiliki $p - value > 0.05$ sehingga dapat disimpulkan H_0 tidak ditolak. Oleh karena itu, dapat dikatakan bahwa data tersebut tidak stasioner.

Selanjutnya, akan dilakukan differencing sebanyak satu kali untuk melihat apakah data dari ketiga tipe stasioner dengan hipotesis, tingkat signifikansi, dan statistik uji yang sama. Dengan R, diperoleh sebagai berikut:

```

> differenced_ts1 <- diff(ts1, differences = 1)
> adf.test(differenced_ts1)

Augmented Dickey-Fuller Test

data: differenced_ts1
Dickey-Fuller = -4.1344, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

> differenced_ts2 <- diff(ts2, differences = 1)
> adf.test(differenced_ts2)

Augmented Dickey-Fuller Test

data: differenced_ts2
Dickey-Fuller = -4.4016, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

> differenced_ts3 <- diff(ts3, differences = 1)
> adf.test(differenced_ts3)

Augmented Dickey-Fuller Test

data: differenced_ts3
Dickey-Fuller = -5.2772, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

```

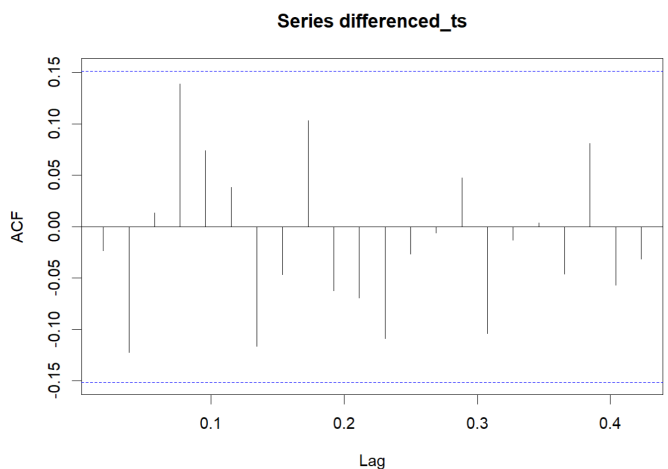
- Aturan Keputusan

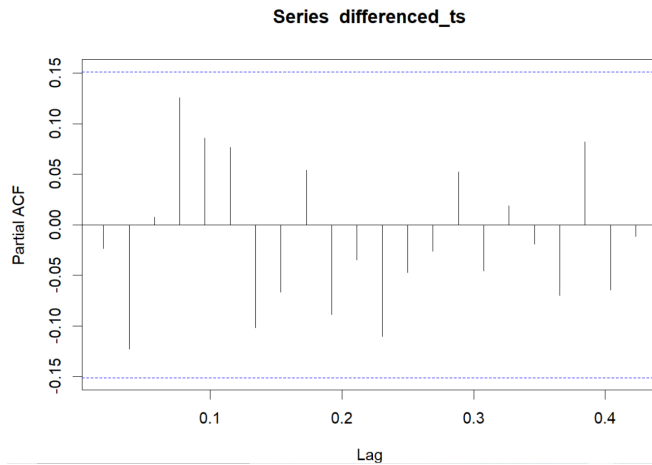
H_0 ditolak jika $p - value < \alpha = 0.05$. Terlihat bahwa ketiganya memiliki $p - value = 0.01 < 0.05$ sehingga dapat disimpulkan H_0 ditolak. Oleh karena itu, dapat dikatakan bahwa data dari tipe keseluruhan, organik, dan konvensional sudah stasioner dengan dilakukan differencing satu kali.

3.2 Spesifikasi Model

Kemudian, akan dilihat dari ACF, PACF, dan EACF untuk menentukan model runtun waktu dari tipe keseluruhan, organik, dan konvensional. Dengan menggunakan R, diperoleh sebagai berikut:

- Tipe Keseluruhan



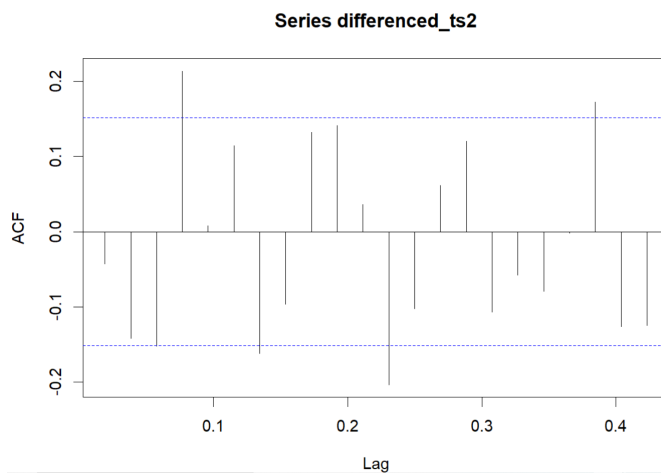


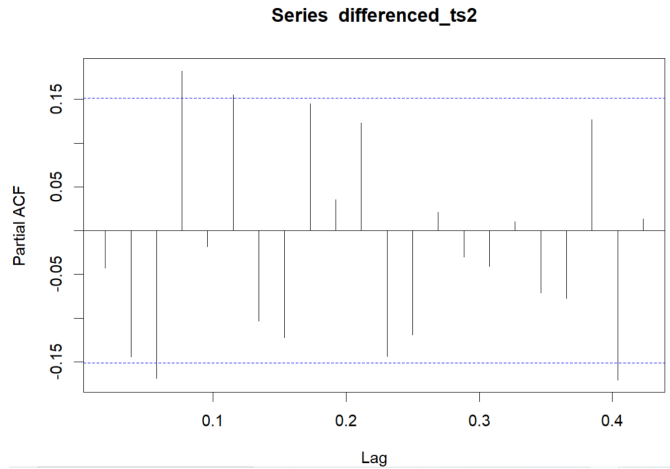
AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	o	o	o	o	o	o	o	o	o	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	o	x	o	o	o	o	o	o	o	o	o	o	o	o
3	o	x	x	o	o	o	o	o	o	o	o	o	o	o
4	x	o	x	x	o	o	o	o	o	o	o	o	o	o
5	x	o	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	x	x	x	x	o	o	o	o	o	o	o	o
7	x	x	x	x	o	x	o	o	o	o	o	o	o	o

Terlihat bahwa dari ACF, PACF, dan EACF bahwa model ARIMA yang tepat untuk tipe keseluruhan adalah ARIMA (0,1,0).

- Tipe Organik

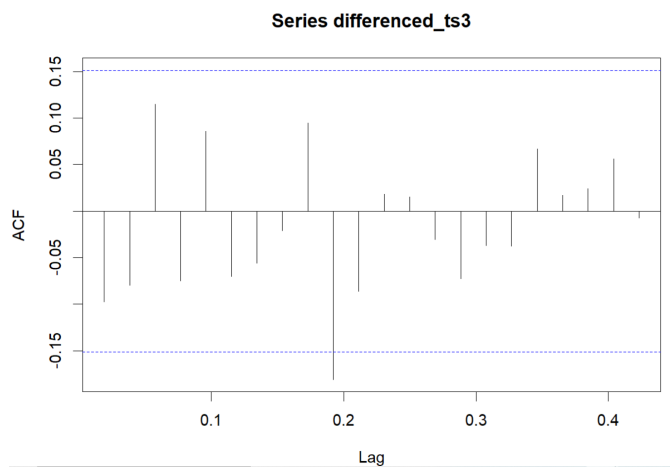


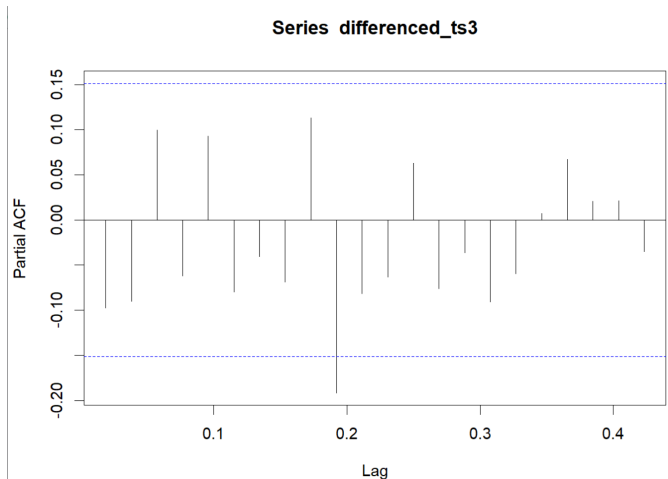


AR/MA															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
0		o	o	o	x	o	o	x	o	o	o	o	x	o	o
1		x	o	o	x	o	o	x	o	o	o	o	x	x	o
2		x	x	o	x	o	o	o	o	o	o	o	o	o	o
3		x	x	x	o	o	o	o	o	o	o	o	o	o	o
4		o	x	x	x	o	o	o	o	o	o	o	o	o	o
5		o	x	x	x	o	o	o	o	o	o	o	o	o	o
6		x	x	o	o	x	o	o	o	o	o	o	o	o	o
7		x	x	x	o	x	x	o	o	o	o	o	o	o	o

Terlihat bahwa dari ACF, PACF, dan EACF bahwa model ARIMA yang tepat untuk tipe organik adalah ARIMA (3,1,3). Akan tetapi, model ini belum terlalu jelas sehingga dibutuhkan pengecekan lebih lanjut.

- Tipe Konvensional





AR/MA		0	1	2	3	4	5	6	7	8	9	10	11	12	13
0		o	o	o	o	o	o	o	o	o	x	o	o	o	o
1		x	o	o	o	o	o	o	o	o	x	o	o	o	o
2		x	x	o	o	o	o	o	o	o	x	o	o	o	o
3		x	x	o	o	o	o	o	o	o	x	x	o	o	o
4		x	o	o	o	o	o	o	o	o	x	o	o	o	o
5		x	o	x	o	o	o	o	o	o	x	o	o	o	o
6		x	x	x	o	x	x	o	o	o	x	o	o	o	o
7		x	o	x	x	o	o	x	o	o	x	o	o	o	o

Terlihat bahwa dari ACF, PACF, dan EACF bahwa model ARIMA yang tepat untuk tipe konvensional adalah ARIMA (0,1,0). Akan tetapi, dibutuhkan pengecekan lebih lanjut untuk mengetahui model yang tepat.

3.3 Overfitting Model

- Tipe Keseluruhan

Model	AIC	BIC
ARIMA(0, 1, 0)	-503.4086	-500.2846
ARIMA(0, 1, 1)	-501.5302	-495.2823
ARIMA(0, 1, 2)	-501.4689	-492.0970
ARIMA(0, 1, 3)	-499.6523	-487.1564
ARIMA(1, 1, 0)	-501.5004	-495.2524
ARIMA(1, 1, 1)	-499.9346	-490.5627
ARIMA(1, 1, 2)	-499.5102	-487.0143
ARIMA(1, 1, 3)	-500.1327	-484.5129
ARIMA(2, 1, 0)	-502.0552	-492.6834
ARIMA(2, 1, 1)	-500.0579	-487.5620
ARIMA(2, 1, 2)	-507.5313	-491.9115
ARIMA(2, 1, 3)	-506.0133	-487.2695
ARIMA(3, 1, 0)	-500.0632	-487.5674

Terlihat bahwa kandidat model yang memenuhi adalah ARIMA (0,1,0) dan ARIMA(2,1,2). Berdasarkan nilai dari AIC dan BIC, maka untuk tipe keseluruhan dipilih model ARIMA (2,1,2) karena memiliki nilai AIC terkecil, yaitu AIC = -507.53 (walaupun BIC nya bukan yang terkecil) dan lebih kompleks dibandingkan dengan ARIMA (0,1,0), serta kami merasa model yang terlalu sederhana tidak bisa menjelaskan data kami.

- Tipe Organik

Model	AIC	BIC
ARIMA(0 , 1 , 0)	-453.1706	-450.0466
ARIMA(0 , 1 , 1)	-451.6099	-445.3620
ARIMA(0 , 1 , 2)	-452.6652	-443.2933
ARIMA(0 , 1 , 3)	-453.2683	-440.7724
ARIMA(1 , 1 , 0)	-451.4768	-445.2289
ARIMA(1 , 1 , 1)	-451.4762	-442.1043
ARIMA(1 , 1 , 2)	-451.1490	-438.6531
ARIMA(1 , 1 , 3)	-461.9118	-446.2920
ARIMA(2 , 1 , 0)	-452.9708	-443.5989
ARIMA(2 , 1 , 1)	-452.3282	-439.8324
ARIMA(2 , 1 , 2)	-465.9256	-450.3057
ARIMA(2 , 1 , 3)	-464.6510	-445.9072
ARIMA(3 , 1 , 0)	-455.8040	-443.3082

Terlihat bahwa kandidat model yang memenuhi adalah ARIMA(2,1,2). Berdasarkan nilai dari AIC dan BIC, maka untuk tipe keseluruhan dipilih model ARIMA (2,1,2) karena memiliki nilai AIC dan BIC terkecil, yaitu AIC = -465.93, BIC = -450.31.

- Tipe Konvensional

Model	AIC	BIC
ARIMA(0 , 1 , 0)	-423.2567	-420.1327
ARIMA(0 , 1 , 1)	-423.1615	-416.9136
ARIMA(0 , 1 , 2)	-421.8261	-412.4542
ARIMA(0 , 1 , 3)	-422.8726	-410.3767
ARIMA(1 , 1 , 0)	-422.8791	-416.6312
ARIMA(1 , 1 , 1)	-421.3699	-411.9980
ARIMA(1 , 1 , 2)	-428.0511	-415.5552
ARIMA(1 , 1 , 3)	-426.7077	-411.0878
ARIMA(2 , 1 , 0)	-422.2635	-412.8916
ARIMA(2 , 1 , 1)	-428.6780	-416.1821
ARIMA(2 , 1 , 2)	-425.6431	-410.0233

ARIMA(2,1,3)	-424.6964	-405.9526
ARIMA(3,1,0)	-421.9418	-409.4459
ARIMA(3,1,1)	-426.8081	-411.1883
ARIMA(3,1,2)	-428.3087	-409.5649
ARIMA(3,1,3)	-429.6997	-407.8319

Terlihat bahwa kandidat model yang memenuhi adalah ARIMA (0,1,0) dan ARIMA(3,1,3). Berdasarkan nilai dari AIC dan BIC, maka untuk tipe keseluruhan dipilih model ARIMA (3,1,3) karena memiliki nilai AIC terkecil, yaitu AIC = -429.70 (walaupun BIC nya bukan yang terkecil) dan lebih kompleks dibandingkan dengan ARIMA (0,1,0).

3.4 Penentuan Model Terbaik

Berdasarkan nilai AIC, BIC, dan kekompleksan model dipilih model terbaik dari ketiga tipe tersebut. Oleh karena itu, diperoleh model terbaik sebagai berikut:

- Tipe Keseluruhan: ARIMA (2,1,2)
- Tipe Organik: ARIMA (2,1,2)
- Tipe Konvensional: ARIMA (3,1,3)

3.5 Estimasi Parameter

Setelah mendapatkan model terbaik, maka akan dilakukan estimasi parameter dengan menggunakan metode *Maximum Likelihood Estimator* (MLE) pada tipe keseluruhan, organik, dan konvensional. Dengan menggunakan R, diperoleh sebagai berikut:

- Tipe Keseluruhan

```
> #overall
> arima(ts1, order=c(2,1,2))

Call:
arima(x = ts1, order = c(2, 1, 2))

Coefficients:
      ar1      ar2      ma1      ma2
  0.6719 -0.9357 -0.7804  0.9997
s.e.  0.0331  0.0348  0.0359  0.0650

sigma^2 estimated as 0.00262:  log likelihood = 258.77,  aic = -509.53
```

Berdasarkan gambar di atas, maka diperoleh estimasi parameter untuk model ARIMA (2,1,2) tipe keseluruhan sebagai berikut:

$$\begin{aligned}\nabla Y_t &= 0.6719\nabla Y_{t-1} - 0.9357\nabla Y_{t-2} + e_t + 0.7804e_{t-1} - 0.9997e_{t-2} \\ Y_t - Y_{t-1} &= 0.6719(Y_{t-1} - Y_{t-2}) - 0.9357(Y_{t-2} - Y_{t-3}) + e_t + 0.7804e_{t-1} - 0.9997e_{t-2} \\ Y_t &= 1.6719Y_{t-1} - 1.6076Y_{t-2} + 0.9357Y_{t-3} + e_t + 0.2562e_{t-1} + 0.3186e_{t-2}\end{aligned}$$

- Tipe Organik

```
> #organic
> arima(ts2, order=c(2,1,2))

call:
arima(x = ts2, order = c(2, 1, 2))

Coefficients:
      ar1      ar2      ma1      ma2
    0.5839 -0.9770 -0.6120  0.9242
s.e.  0.0214  0.0334  0.0526  0.0593

sigma^2 estimated as 0.003423:  log likelihood = 237.96,  aic = -467.93
```

Berdasarkan gambar di atas, maka diperoleh estimasi parameter untuk model ARIMA (2,1,2) tipe organik sebagai berikut:

$$\begin{aligned}\nabla Y_t &= 0.5839\nabla Y_{t-1} - 0.9770\nabla Y_{t-2} + e_t + 0.6120e_{t-1} - 0.9242e_{t-2} \\ Y_t - Y_{t-1} &= 0.5839(Y_{t-1} - Y_{t-2}) - 0.9770(Y_{t-2} - Y_{t-3}) + e_t + 0.6120e_{t-1} - 0.9242e_{t-2} \\ Y_t &= 1.5839Y_{t-1} - 1.5609Y_{t-2} + 0.9770Y_{t-3} + e_t + 0.6120e_{t-1} - 0.9242e_{t-2}\end{aligned}$$

- Tipe Konvensional

```
> #conventional
> arima(ts3, order=c(3, 1, 3))

call:
arima(x = ts3, order = c(3, 1, 3))

Coefficients:
      ar1      ar2      ar3      ma1      ma2      ma3
    -0.9481  0.7693  0.8406  0.9822 -0.9305 -0.9825
s.e.  0.0558  0.0906  0.0570  0.0443  0.0516  0.0441

sigma^2 estimated as 0.004028:  log likelihood = 221.85,  aic = -431.7
```

Berdasarkan gambar di atas, maka diperoleh estimasi parameter untuk model ARIMA (3,1,3) tipe organik sebagai berikut:

$$\begin{aligned}\nabla Y_t &= -0.9481\nabla Y_{t-1} + 0.7693\nabla Y_{t-2} + 0.8406\nabla Y_{t-3} + e_t - 0.9822e_{t-1} + \\ &\quad 0.9305e_{t-2} + 0.9825e_{t-3} \\ Y_t - Y_{t-1} &= -0.9481(Y_{t-1} - Y_{t-2}) + 0.7693(Y_{t-2} - Y_{t-3}) + 0.8406(Y_{t-3} - Y_{t-4}) + \\ &\quad e_t - 0.9822e_{t-1} + 0.9305e_{t-2} + 0.9825e_{t-3} \\ Y_t &= 0.0519Y_{t-1} + 1.714Y_{t-2} + 0.0713 - 0.8406Y_{t-4} + e_t - 0.9822e_{t-1} + \\ &\quad 0.9305e_{t-2} + 0.9825e_{t-3}\end{aligned}$$

3.6 Model Diagnostik (Analisis Residual)

➤ Uji Ljung-Box (Independensi)

- **Hipotesis**

$H_0: \rho_k = 0$ (tidak ada korelasi residual antar lag)

$H_1: \text{minimal ada satu } \rho_k \neq 0$ (terdapat korelasi residual antar lag)

- **Tingkat Signifikansi**

$\alpha = 0.05$

- **Statistik Uji**

$$Q = n(n + 2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right)$$

Dimana:

n : jumlah sampel.

m: jumlah lag.

$\hat{\rho}_k$: nilai koefisien autokorelasi time lag k.

Dengan menggunakan software R, diperoleh output untuk tipe keseluruhan, organik, dan konvensional secara berurutan sebagai berikut:

```
> #overall
> checkresiduals(arima(ts1, order = c(2, 1, 2)))
```

Ljung-Box test

```
data: Residuals from ARIMA(2,1,2)
Q* = 27.772, df = 30, p-value = 0.5825
```

```
Model df: 4. Total lags used: 34
```

```
> #organic
> checkresiduals(arima(ts2, order = c(2, 1, 2)))
```

Ljung-Box test

```
data: Residuals from ARIMA(2,1,2)
Q* = 49.765, df = 30, p-value = 0.01312
```

```
Model df: 4. Total lags used: 34
```

```
> #conventional
> checkresiduals(arima(ts3, order = c(3, 1, 3)))
```

Ljung-Box test

data: Residuals from ARIMA(3,1,3)
Q* = 20.378, df = 28, p-value = 0.8503

Model df: 6. Total lags used: 34

- **Aturan Keputusan**

H_0 ditolak jika $p\text{-value} < \alpha = 0.05$, atau H_0 ditolak jika $Q * > \chi^2_{\alpha, df}$ (table Chi-Square) , sebaliknya H_0 gagal ditolak. Oleh karena itu terlihat bahwa p-value dari ketiganya, yaitu 0.5825, 0.01312, dan 0.8503. Perhatikan bahwa tipe keseluruhan dan konvensional memiliki $p\text{-value} > \alpha = 0.05$, sedangkan tipe organik memiliki $p\text{-value} < \alpha = 0.05$. Dengan demikian, untuk tipe keseluruhan dan konvensional H_0 tidak ditolak sehingga dapat disimpulkan bahwa tidak ada korelasi residual antar lag. Sedangkan, untuk tipe organik H_0 ditolak sehingga dapat disimpulkan bahwa terdapat korelasi residual antar lag.

➤ **Uji Jarque Bera (Normalitas)**

- **Hipotesis**

H_0 : Residual berdistribusi normal

H_1 : Residual tidak berdistribusi normal

- **Tingkat Signifikansi**

$\alpha = 0.05$

- **Statistik Uji**

$$JB = n \left[\frac{s^2}{6} + \frac{(K-3)^2}{24} \right]$$

Keterangan:

n : ukuran sampel.

S : koefisien *skewness*.

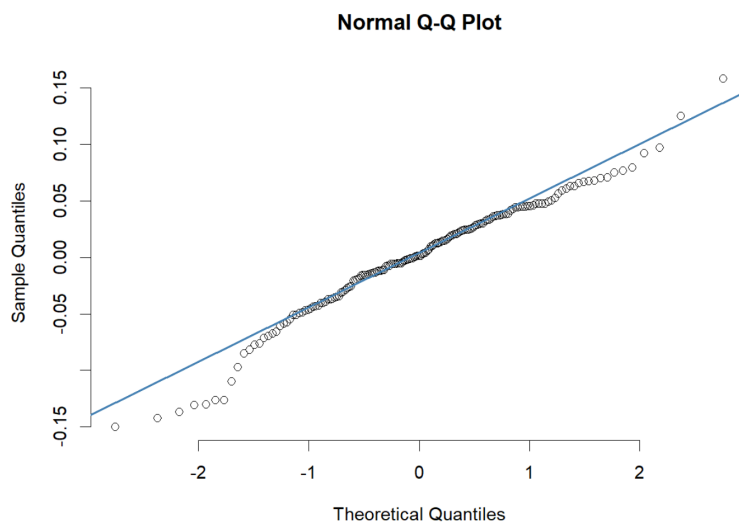
K : koefisien kurtosis.

Dengan menggunakan software R, diperoleh output untuk tipe keseluruhan, organik, dan konvensional secara berurutan sebagai berikut:

```
> #overall  
> jarque.bera.test(residuals(arima(ts1, order = c(2, 1, 2))))
```

Jarque Bera Test

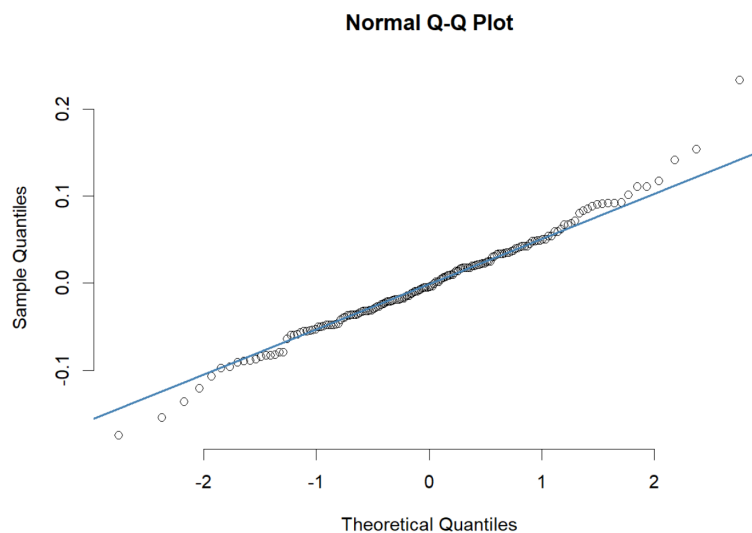
data: residuals(arima(ts1, order = c(2, 1, 2)))
X-squared = 11.457, df = 2, p-value = 0.003252



```
> #organic  
> jarque.bera.test(residuals(arima(ts2, order = c(2, 1, 2))))
```

Jarque Bera Test

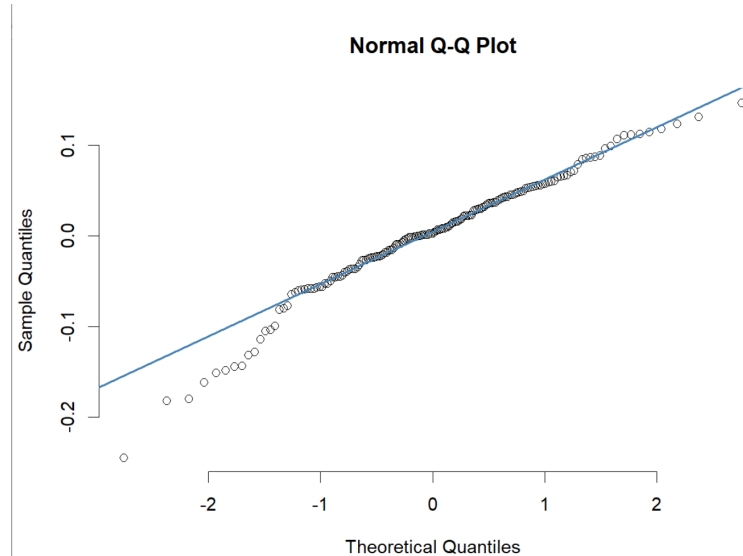
data: residuals(arima(ts2, order = c(2, 1, 2)))
X-squared = 14.107, df = 2, p-value = 0.0008644



```
> #conventional
> jarque.bera.test(residuals(arima(ts3, order = c(3, 1, 3))))
```

Jarque Bera Test

data: residuals(arima(ts3, order = c(3, 1, 3)))
X-squared = 24.197, df = 2, p-value = 5.568e-06



• Aturan Keputusan

H_0 ditolak jika $p\text{-value} < \alpha = 0.05$, atau H_0 ditolak jika $JB > \chi^2_{\alpha, df}$ (table Chi-Square), sebaliknya H_0 gagal ditolak. Oleh karena itu terlihat bahwa p-value dari ketiganya, yaitu 0.003252, 0.0008644, dan 5.568e-06. Perhatikan bahwa ketiga tipe memiliki $p\text{-value} < \alpha = 0.05$. Dengan demikian, untuk tipe keseluruhan, organik, dan konvensional H_0 ditolak sehingga dapat disimpulkan bahwa residual tidak berdistribusi normal. Kemudian, dari plot ketiganya juga terlihat bahwa ketiganya memiliki residual yang tidak berdistribusi normal.

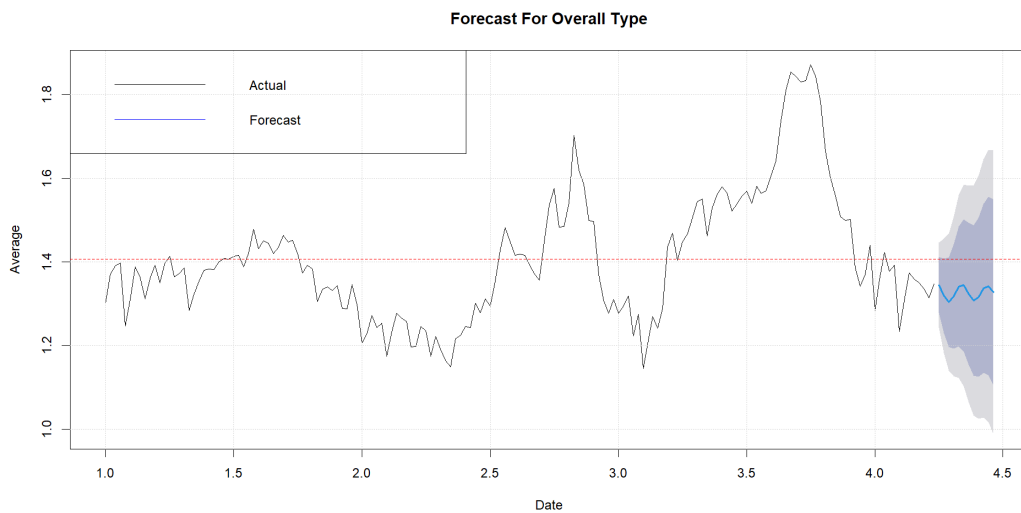
Berdasarkan uji Jarque Bera, dapat disimpulkan bahwa tipe keseluruhan dan konvensional tidak memenuhi asumsi normalitas (residual tidak berdistribusi normal). Namun, hal ini tidak perlu dimasalahkan karena *outlier* yang ada, terutama pada tahun 2017, memengaruhi kedua tipe tersebut sehingga tidak memenuhi asumsi normalitas. Kemudian, berdasarkan uji Ljung-Box dan Jarque Bera, tipe organik tidak memenuhi asumsi independensi dan normalitas. Hal ini dikarenakan kami mendeteksi bahwa terdapat banyak *outlier* di tahun 2017 sehingga tipe organik tidak memenuhi asumsi independensi dan normalitas. Selanjutnya, akan dianalisis lebih lanjut dengan menghilangkan seluruh data pada tahun 2017 untuk melihat khususnya tipe organik agar memenuhi asumsi independensi dan normalitas.

3.7 Forecasting

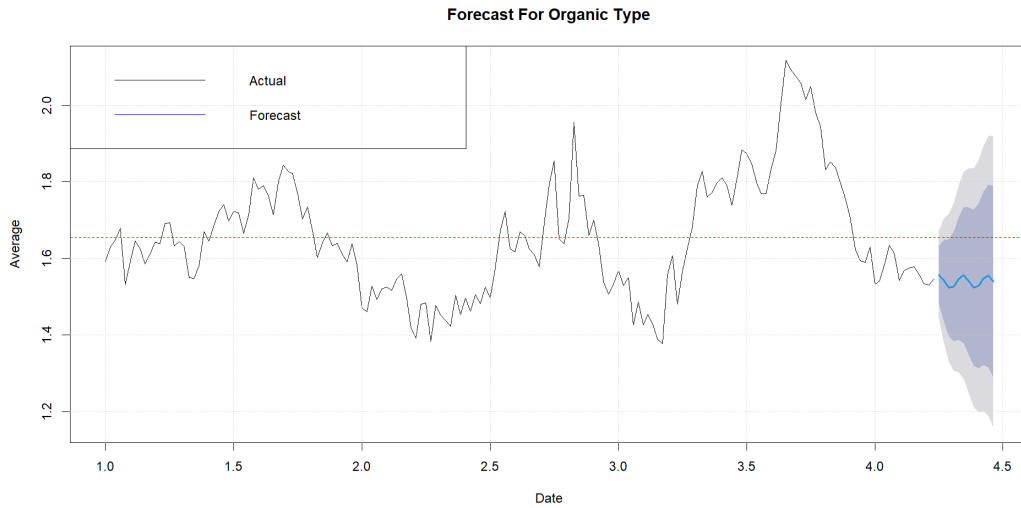
Peramalan adalah langkah yang esensial untuk mendapatkan informasi utama terkait prediksi harga alpukat dan merupakan tujuan utama dan paling berguna dari pembuatan penelitian ini. Sebagaimana model sudah ditentukan yang terbaik, akan dibuat prediksi yang didasarkan model-model yang sudah ditentukan.

Dibuat grafik peramalan untuk melihat prediksi 3 bulan ke depan atau 12 minggu ke depan dengan kode dan output yang diberikan di bawah untuk tipe keseluruhan:

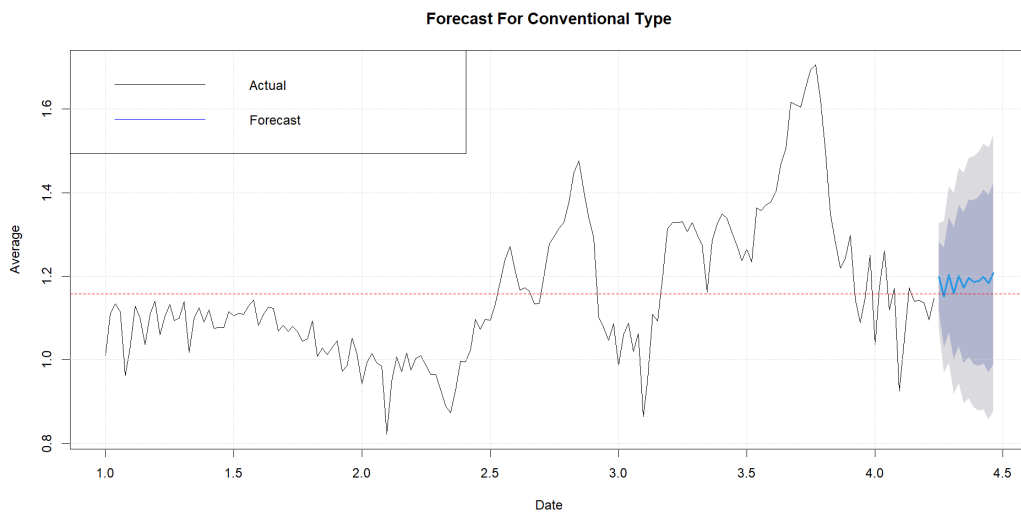
```
> forecast2<-forecast::forecast(fit1, h=12)
> plot(forecast2, main="Forecast For Overall Type", xlab="Date", ylab="Average")
> # Add a legend
> legend("topleft", legend = c("Actual", "Forecast"), col = c("black", "blue"), lty = 1)
> # Add gridlines
> grid(lty = 3)
> # Add a horizontal line at the mean value
> abline(h = mean(df$AveragePrice), col = "red", lty = 2)
```



Dapat dilihat grafik bahwa data yang ada memiliki fluktuasi yang cukup tinggi, meskipun pada awal pengamatan masih relatif stabil. Model memberikan fluktuasi yang tidak begitu besar dibanding data-data sebelumnya. Hasil ini dapat mengundang keraguan karena tidak begitu menangkap tren fluktuasi yang cukup besar, tetapi merupakan model yang sederhana dalam memberikan prediksi harga alpukat pada rentang waktu lima minggu. Berikutnya juga akan dilihat untuk tipe organik.



Pada jenis organik, digunakan model juga ARIMA(2,1,2) dengan pola yang dibuat dan hasil peramalan mirip dengan jenis keseluruhan sehingga dapat ditarik kesimpulan yang sama karena model yang digunakan pun juga sama. Akan tetapi, pada uji asumsi normalitas galat, didapat bahwa model ini tidak memenuhi asumsi tersebut sehingga hasil prediksi ini tidak begitu kredibel, meskipun tren dari data terbaru terlihat konsisten dengan hasil prediksi. Selanjutnya, akan dilihat grafik untuk tipe konvensional:



Model untuk tipe konvensional menggunakan ARIMA (3, 1, 3) dan dapat dilihat secara tren, tetap terdapat kemiripan dengan grafik-grafik sebelumnya. Dapat disimpulkan bahwa prediksi untuk 3 bulan ke depan tidak memiliki perbedaan yang begitu jauh dengan jenis lainnya yang cenderung dekat dengan garis rata-rata, serta tidak memiliki puncak ekstrem sebagaimana ada pada data asli. Hasil dari peramalan dapat dilihat sebagai berikut:

```

> forecast4
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
4.250000      1.112798  1.0277091  1.197887  0.9826659  1.242930
4.269231      1.132654  1.0176075  1.247700  0.9567057  1.308602
4.288462      1.171367  1.0325264  1.310209  0.9590283  1.383707
4.307692      1.163344  1.0004647  1.326224  0.9142415  1.412447
4.326923      1.123886  0.9392269  1.308546  0.8414741  1.406299
4.346154      1.120337  0.9189404  1.321734  0.8123275  1.428346
4.365385      1.157209  0.9416893  1.372728  0.8276002  1.486817
4.384615      1.171132  0.9405242  1.401739  0.8184480  1.523816
4.403846      1.139701  0.8932047  1.386198  0.7627174  1.516685
4.423077      1.117370  0.8571481  1.377592  0.7193949  1.515345
4.442308      1.141179  0.8696522  1.412705  0.7259148  1.556442
4.461538      1.169409  0.8865173  1.452300  0.7367635  1.602054

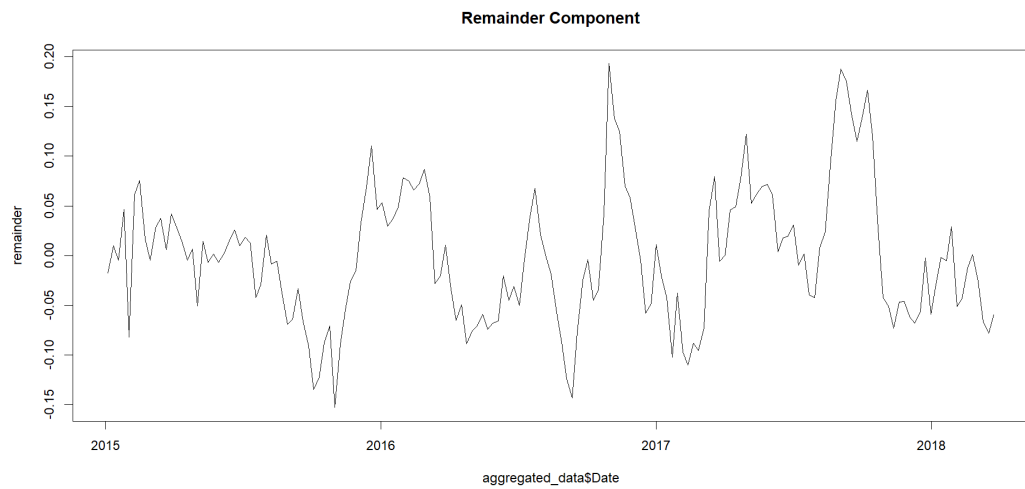
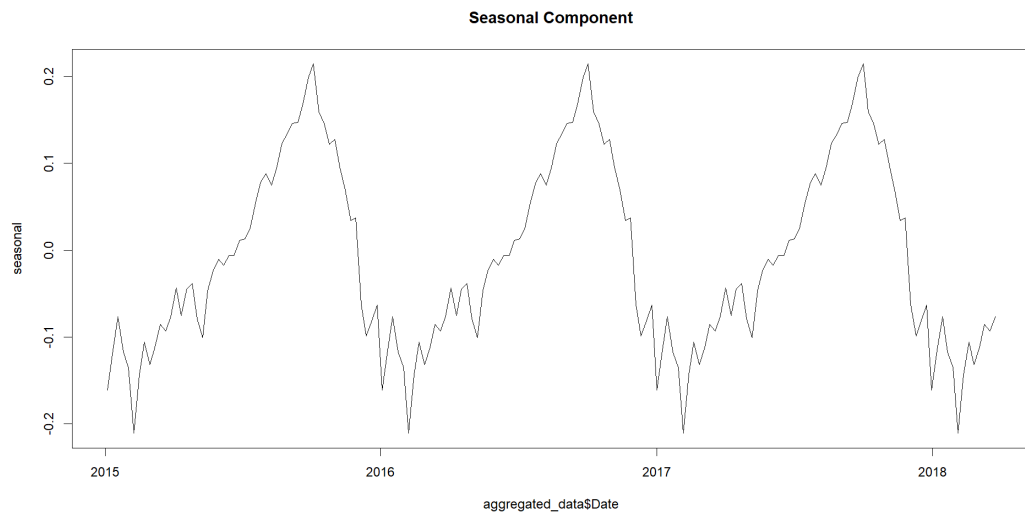
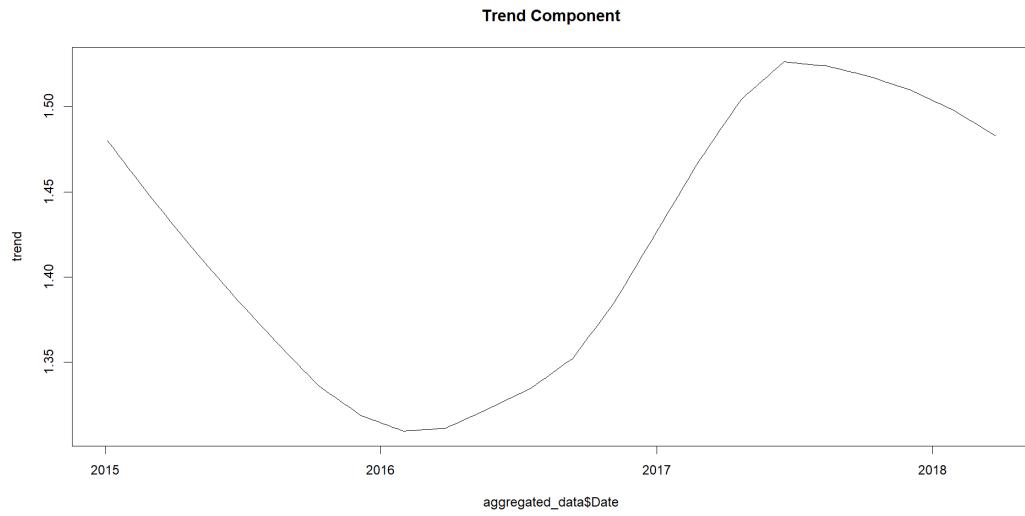
> forecast2
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
4.250000      1.344667  1.277931  1.411403  1.2426037  1.446730
4.269231      1.318936  1.229356  1.408517  1.1819345  1.455938
4.288462      1.303693  1.196615  1.410771  1.1399311  1.467454
4.307692      1.317528  1.192552  1.442503  1.1263945  1.508661
4.326923      1.341087  1.198153  1.484021  1.1224888  1.559686
4.346154      1.343971  1.186005  1.501937  1.1023824  1.585560
4.365385      1.323863  1.154444  1.493282  1.0647587  1.582967
4.384615      1.307654  1.128202  1.487107  1.0332051  1.582103
4.403846      1.315580  1.125255  1.505904  1.0245030  1.606656
4.423077      1.336072  1.133899  1.538244  1.0268754  1.645268
4.442308      1.342424  1.129299  1.555549  1.0164780  1.668370
4.461538      1.327517  1.105430  1.549604  0.9878639  1.667170

> forecast3
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
4.250000      1.557324  1.481440  1.633209  1.441269  1.673380
4.269231      1.542451  1.436629  1.648274  1.380610  1.704293
4.288462      1.523173  1.397110  1.649236  1.330376  1.715970
4.307692      1.526446  1.383437  1.669456  1.307732  1.745160
4.326923      1.547192  1.387046  1.707338  1.302270  1.792114
4.346154      1.556109  1.378974  1.733243  1.285205  1.827012
4.365385      1.541047  1.349322  1.732773  1.247829  1.834266
4.384615      1.523541  1.319958  1.727124  1.212188  1.834894
4.403846      1.528033  1.313430  1.742635  1.199826  1.856239
4.423077      1.547759  1.321295  1.774223  1.201413  1.894106
4.442308      1.554890  1.316175  1.793606  1.189806  1.919974
4.461538      1.539782  1.290155  1.789409  1.158010  1.921553

```

Selain penggunaan model ARIMA, digunakan juga model *seasonal naive*. Metode tersebut digunakan juga karena terlihat pada *line plot* adanya kecenderungan naik di mulai di bulan Mei hingga Oktober dengan puncak di September dan Oktober. Dengan demikian, kami menduga adanya tren *seasonality* juga pada model kami.

Untuk membuat model *seasonal*, dilakukan dekomposisi STL terlebih dahulu untuk mendapat gambaran kasar terkait *seasonality* dari data, lalu dapat dilihat plot komponen tren, *seasonal*, dan *remainder* yang secara berturut sebagai berikut:



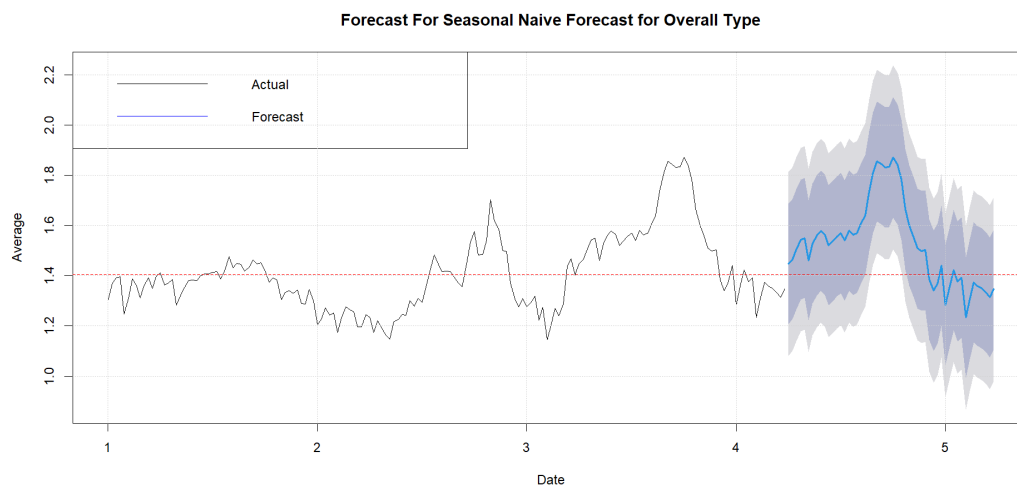
Pada grafik komponen tren yang menghiraukan fluktuasi dan *noise* pada rentang waktu pendek terlihat secara sederhana bahwa kenaikan harga drastis terjadi pada 2017 setelah penurunan pada 2016, lalu mulai menurun perlahan menuju 2018.

Pada grafik komponen *seasonal* dapat dilihat bahwa setiap tahun memiliki tren yang sangat konsisten, yaitu kenaikan drastis mulai pertengahan tahun lalu mulai turun drastis mendekati beberapa saat setelah awal tahun baru.

Pada grafik komponen *remainder* yang memuat besar galat, dapat dilihat bahwa pada sekitar puncak kenaikan di tahun 2017 dan tiga perempat menuju akhir tahun 2018 terdapat fluktuasi galat yang besar. Ini dapat menandakan perubahan yang tidak dapat disangka dari data tahun sebelumnya.

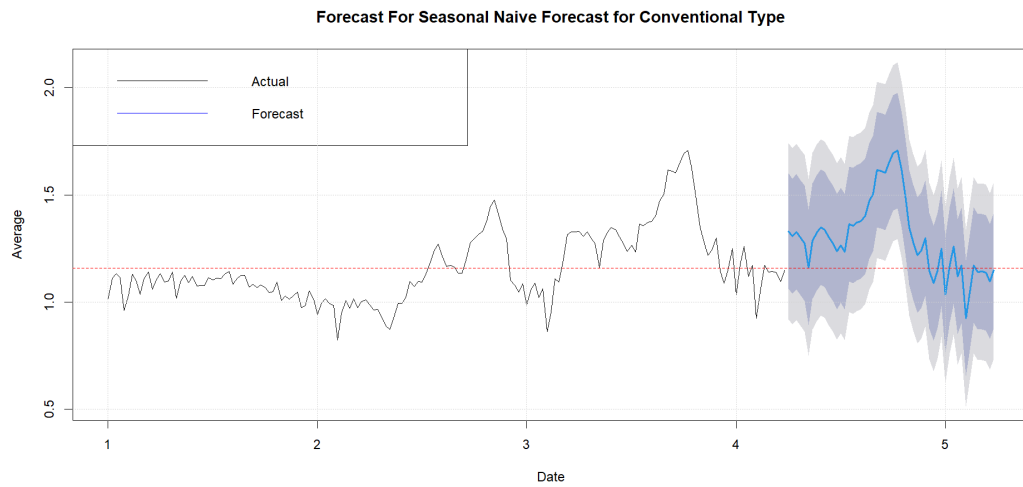
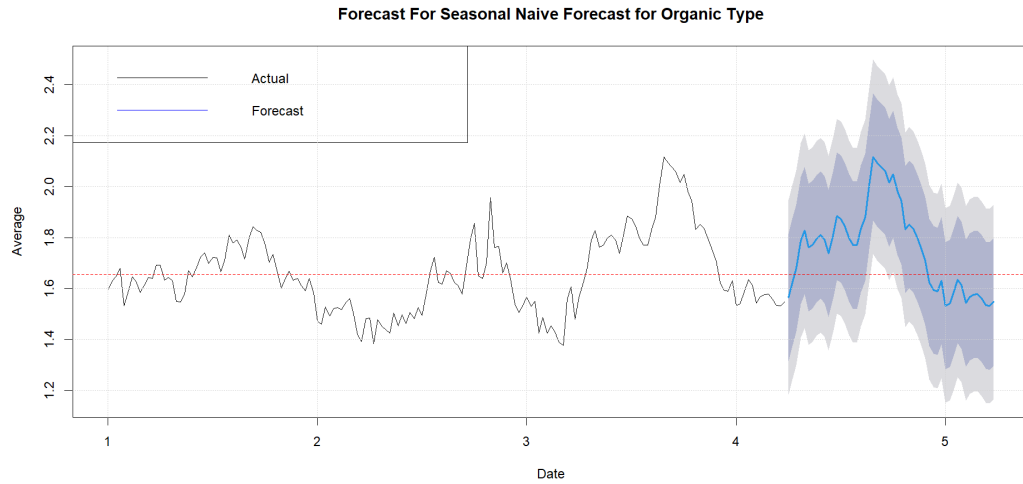
Secara keseluruhan dari dekomposisi STL, terlihat bahwa mungkin ada tren *seasonality* pada data sebagaimana terlihat fluktuasi yang konsisten dari setiap tahunnya pada grafik komponen *seasonal*. Hal ini juga konsisten bahkan saat dibagi ke setiap tipe alpukat.

Berikutnya, mulai dilakukan peramalan dengan metode *naive forecasting*. Peramalan dilakukan untuk satu tahun ke depan atau 52 minggu karena berupa prediksi *seasonality* sehingga perlu mencakup keseluruhan tahun yang cenderung berfluktuasi pada bulan-bulan tertentu. Berikut adalah hasil peramalan untuk jenis keseluruhan:



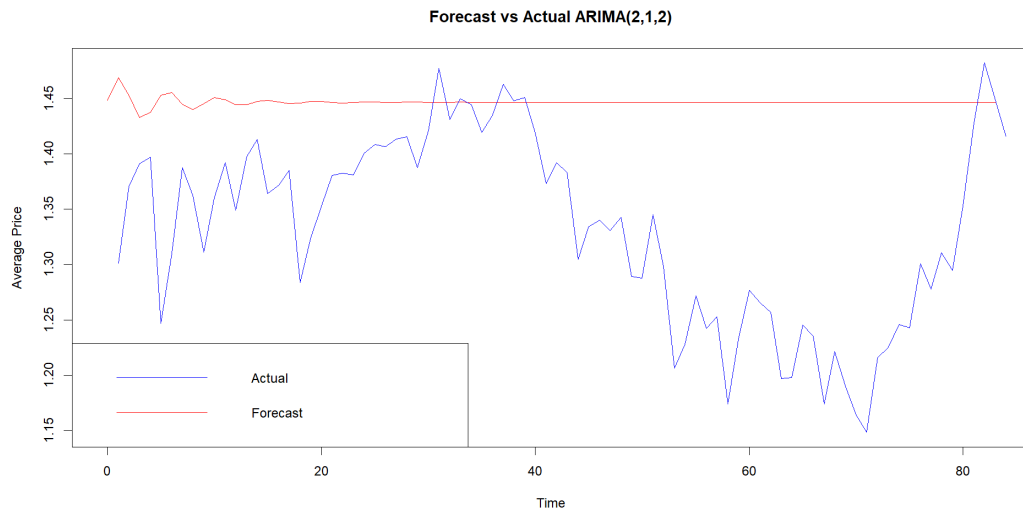
Terlihat hasil prediksi *naive forecasting* hanya mengikuti persis observasi 52 minggu sebelumnya. Model ini bisa cukup bagus apabila memang faktor yang menyebabkan kenaikan dan penurunan harga alpukat tipe keseluruhan konsisten untuk satu tahun lagi.

Berikut akan ditunjukkan juga grafik peramalan dari tipe organik dan konvensional berurutan dengan metode *seasonal naive*:



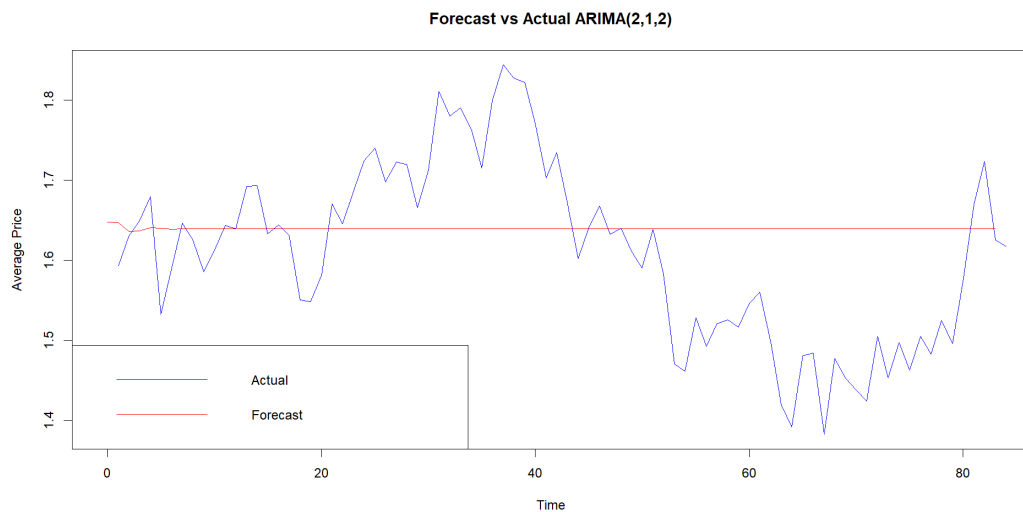
Keduanya memiliki karakteristik dan kesimpulan yang sama dengan tipe keseluruhan.

Berikutnya akan dilihat keakuratan dari model dengan menggunakan grafik dan memisahkan data menjadi data *train* dan data *test*. Data runtun waktu yang kami miliki berjumlah ganjil sehingga kami menghapus pengamatan terakhir agar mendapat data *train* dan *test* agar mempermudah perbandingan. Berikut hasil untuk tipe keseluruhan



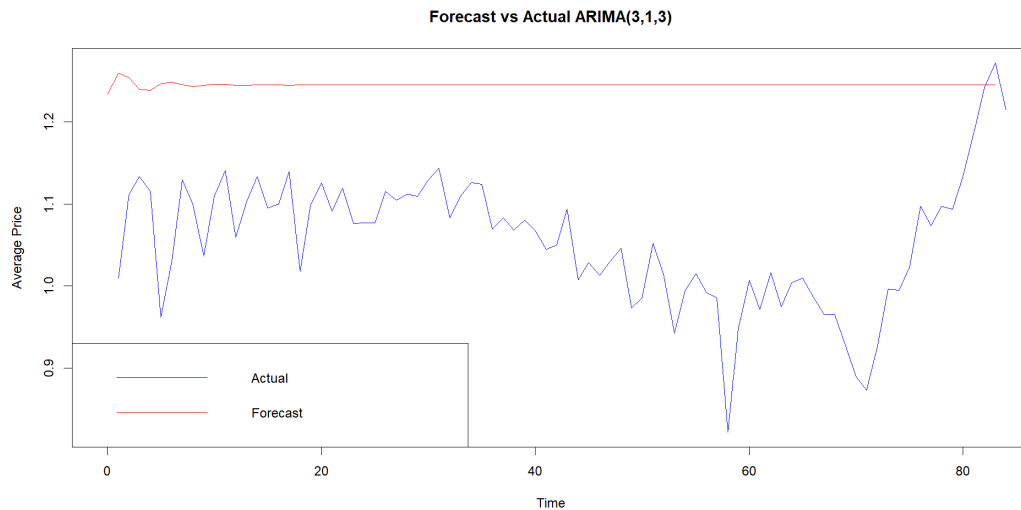
Terlihat bahwa prediksi cenderung jauh dari data sesungguhnya yang terjadi. Ini bisa terjadi karena model ARIMA gagal menangkap tren *seasonality* yang mungkin ada atau terdapat kejanggalan berupa *outliers* pada data *train*, maupun *test*.

Berikut hasil untuk organik:



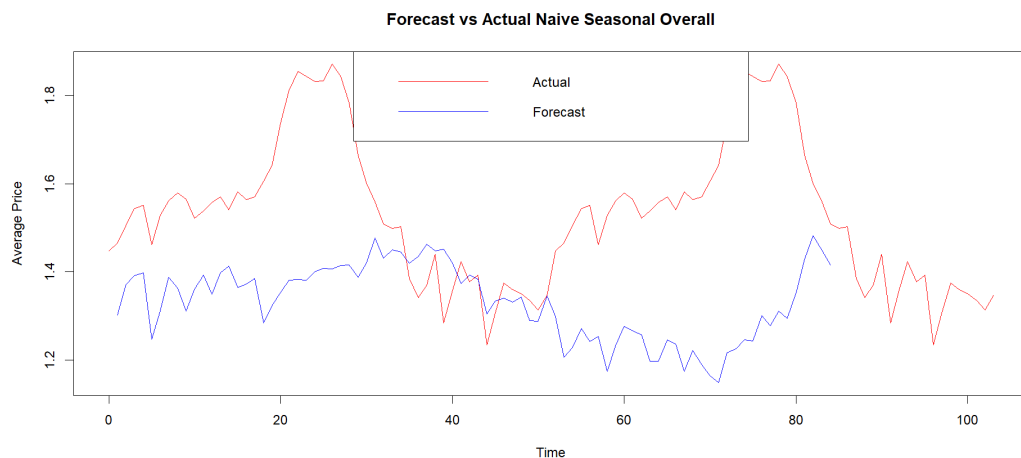
Terlihat bahwa peramalan lebih terlihat stabil dan akurat dibanding pada grafik sebelumnya. Memang ada fluktuasi, terutama saat di tengah, tetapi secara keseluruhan masih lebih baik dibanding grafik tipe keseluruhan, meski kesimpulan dari dari tipe keseluruhan masih cukup relevan untuk grafik ini.

Berikutnya akan dilihat juga untuk tipe konvensional.



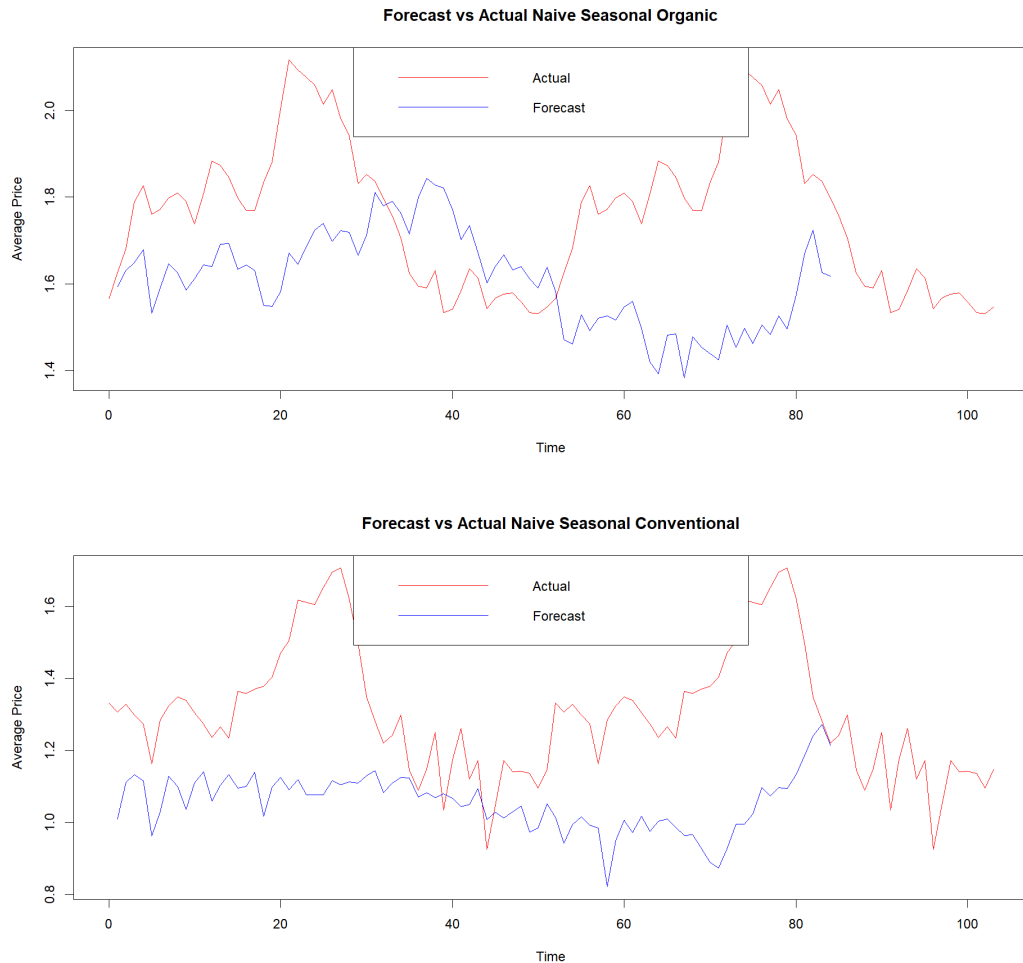
Jelas bahwa peramalan untuk tipe konvensional tidak bagus. Dengan basis alasan yang mirip dengan model-model sebelumnya, kami menduga hal ini berhubungan perbedaan drastis kenaikan puncak harga alpukat pada tahun 2017 sebagaimana ada pada *line plot* sebelumnya.

Selain untuk ARIMA, kami juga mencoba mengukur akurasi dengan *seasonal naive* dimulai dengan tipe keseluruhan.



Berhubung plot ini hanyalah *seasonal naive*, bisa dikatakan ini hanya menumpuk grafik sebelumnya yang sudah dibuat setelah memisahkan keduanya. Di sini semakin jelas bahwa dua tahun pertama dengan berikutnya memiliki kenaikan yang cukup drastis dalam harga alpukat. Inilah alasan mengapa peramalan tidak terlihat begitu bagus.

Selain tipe keseluruhan, akan dilihat juga untuk tipe organik dan konvensional secara berurutan.



Dapat ditarik kesimpulan yang sama dengan tipe buah alpukat keseluruhan bahwa tahun 2017 dan 2018 memiliki kenaikan yang jauh dari prediksi dan kemungkinan disebabkan oleh faktor-faktor lain yang terlepas dari waktu saja. Akan tetapi, tipe organik terlihat cenderung lebih konsisten dengan hasil prediksi dibanding konvensional yang berbeda sangat jauh. Ini berarti peningkatan harga alpukat lebih drastis pada tipe konvensional dibanding organik.

3.8 Analisis dan Modeling Lanjutan Tanpa Outlier

Mengingat pada *line plot* sebelumnya terlihat bahwa pada tahun 2017 memiliki kenaikan pada bulan yang memiliki tren naik, akan dilakukan *handling outliers* dengan menghapus tahun 2017 sepenuhnya dan membandingkan hasil model terbaik sebelum dan sesudah dihapus, sekaligus menyimpulkan apakah model tanpa tahun 2017 akan mengalami perubahan.

```
> df <- df[format(df$Date, "%Y") != "2017", ]
> organic <- df %>% select(Date, AveragePrice, type) %>% filter(type == "organic")
> conventional <- df %>% select(Date, AveragePrice, type) %>% filter(type == "conventional")
> head(organic)
```

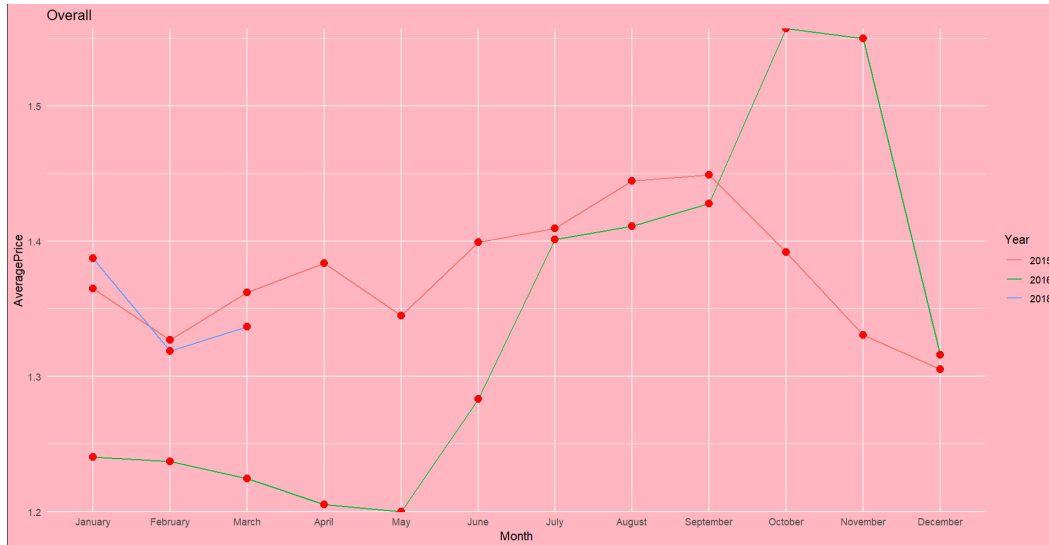

Mirip dengan kode-kode sebelumnya yang sudah dijalankan, perbedaan kali ini hanyalah tahun 2017 sudah dihilangkan. Dilakukan ulang beberapa kode penting untuk data yang sudah disesuaikan ini.



Dengan data yang diperbarui, terlihat tren yang mirip, tetapi kenaikan jauh lebih tidak ekstrem dibanding sebelumnya. Poin lainnya yang bisa disimpulkan dari grafik ini bahwa September dan Oktober masih menjadi 2 titik tertinggi, tetapi jarak September dengan Oktober menjauh, konsisten dengan seberapa ekstrem kenaikan di bulan September pada *line plot* tahun 2017 sebelumnya. Akan dilihat lagi secara lebih rinci untuk setiap tipe alpukat.



Tren untuk kedua tipe sekarang sudah lebih jelas tanpa lompatan ekstrem. Akan dilanjutkan analisis grafik dengan *line plot* untuk melihat kembali tren masing-masing tahun.



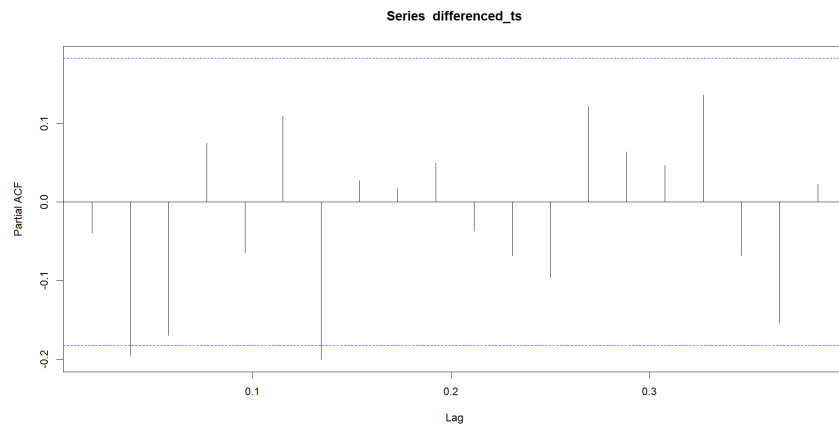
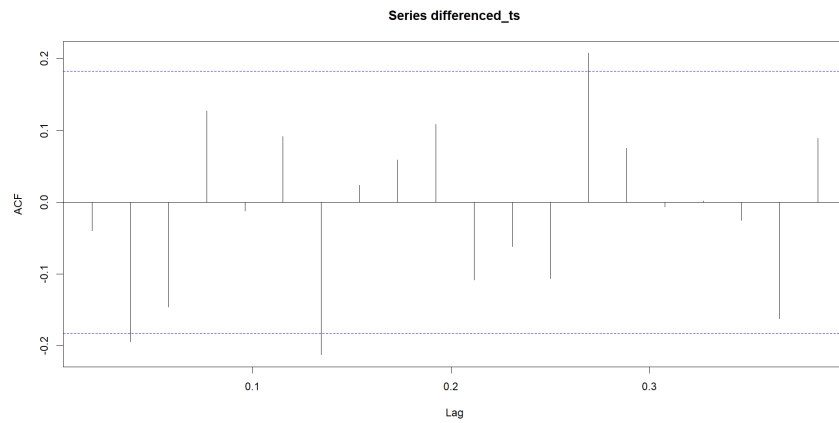
Jelas bahwa data tahun 2017 sudah tidak ada. Tahun 2016 memiliki titik penjualan yang cukup tinggi dibanding tahun 2015 yang menyebabkan grafik keseluruhan memuncak di bulan Oktober, tetapi perbedaan ini tidak securam tahun 2017.



Dari grafik tersebut jelas bahwa kenaikan yang lebih signifikan adalah tahun 2016 pada tipe konvensional dibanding tipe organik yang relatif lebih stabil dibanding tahun sebelumnya.

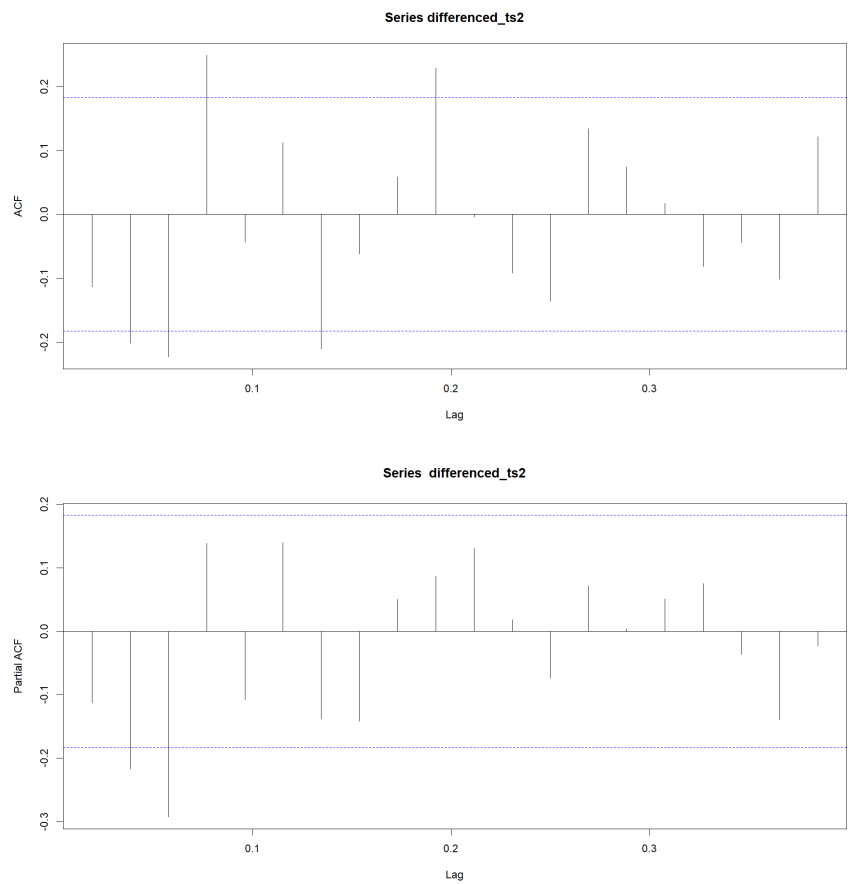
Berikutnya dilakukan modeling lagi, tetapi hanya akan dilakukan modeling ARIMA dengan pendekatan yang sama dari sebelumnya. Setelah menjadikan ketiga variabel data menjadi runtun waktu, dilakukan tes ADF untuk ketiganya yang memberikan kesimpulan yang sama dengan sebelumnya, yakni masih belum stasioner. Oleh karena itu, dilakukan juga *differencing* satu kali dan menyebabkan tes adf sekarang sudah stasioner. Berikut hasil ACF, PACF, dan EACF dari runtun waktu yang sudah *differencing* satu kali, berurutan untuk keseluruhan, organik, dan konvensional:

1) Tipe Keseluruhan



AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	o	o	o	o	x	o	o	o	o	o	o	x
1	x	o	o	o	o	o	x	o	o	o	o	o	o	x
2	x	x	x	o	o	o	x	o	o	o	o	o	o	x
3	x	o	o	o	o	o	o	o	o	o	o	o	o	x
4	x	o	o	o	o	o	o	o	o	o	o	o	o	x
5	x	o	o	o	x	o	o	o	o	o	o	o	o	o
6	x	o	o	x	x	o	o	o	o	o	o	o	o	o
7	o	o	x	o	x	o	o	o	o	o	o	o	o	o

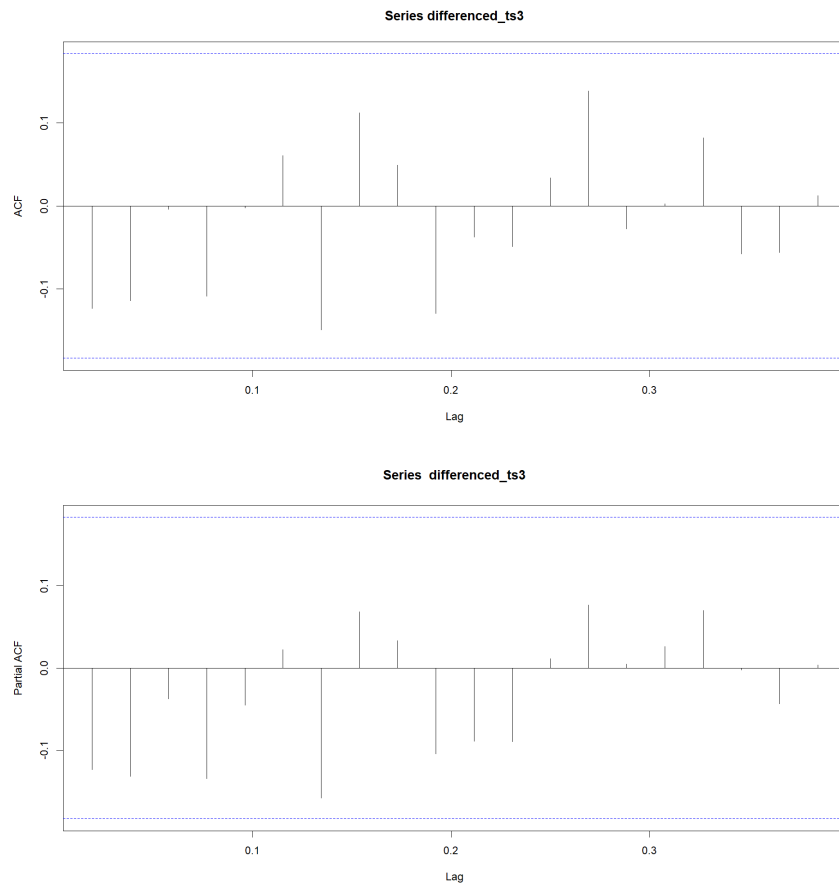
2) Tipe Organik



AR/MA

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	x	x	o	o	x	o	o	x	o	o	o	o
1	x	o	x	x	o	o	x	o	o	x	o	o	o	o
2	x	o	x	x	o	o	x	o	o	o	o	o	o	o
3	x	o	o	o	x	o	o	o	o	o	o	o	o	o
4	x	o	o	o	x	o	o	o	o	o	o	o	o	o
5	x	x	o	o	x	o	o	o	o	o	o	o	o	o
6	x	o	o	o	x	o	o	o	o	o	o	o	o	o
7	x	x	o	x	o	o	o	o	o	o	o	o	o	o

3) Tipe Konvensional



AR/MA															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	
0	o	o	o	o	o	o	o	o	o	o	o	o	o	o	
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o	
2	x	x	o	o	o	o	o	o	o	o	o	o	o	o	
3	x	x	x	o	o	o	o	o	o	o	o	o	o	o	
4	x	x	x	o	o	o	o	o	o	o	o	o	o	o	
5	x	o	x	x	o	o	o	o	o	o	o	o	o	o	
6	o	x	x	x	o	o	o	o	o	o	o	o	o	o	
7	x	x	x	o	o	o	o	o	o	o	o	o	o	o	

Hanya dengan sekilas melihat EACF, terlihat perbedaan yang cukup jauh dari perkiraan model yang cocok. Untuk tipe alpukat keseluruhan, terlihat model yang cukup cocok adalah ARIMA (0, 1, 2), untuk organik adalah ARIMA (3, 1, 1), dan untuk konvensional adalah ARIMA (0, 1, 0).

Selanjutnya, akan dilakukan *overfitting model* dari dugaan awal orde ARIMA dimulai dari tipe keseluruhan.

Model	AIC	BIC
ARIMA(0 , 1 , 2)	-358.3988	-350.1640
ARIMA(0 , 1 , 3)	-358.6167	-347.6369
ARIMA(1 , 1 , 2)	-357.3462	-346.3665
ARIMA(1 , 1 , 3)	-362.9223	-349.1976
ARIMA(2 , 1 , 2)	-362.5514	-348.8268
ARIMA(2 , 1 , 3)	-360.5530	-344.0834
ARIMA(3 , 1 , 2)	-361.0820	-344.6124
ARIMA(3 , 1 , 3)	-361.9168	-342.7023

Terlihat bahwa AIC terkecil benar ada pada ARIMA (0, 1, 2), sedangkan BIC terkecil oleh ARIMA (1, 1, 3). Berikutnya dilihat hasil overfitting dari tipe organik.

ARIMA(3 , 1 , 1)	-325.6357	-311.9111
ARIMA(3 , 1 , 2)	-323.6384	-307.1688
ARIMA(3 , 1 , 3)	-323.0502	-303.8357
ARIMA(3 , 1 , 4)	-321.5884	-299.6289
ARIMA(3 , 1 , 5)	-324.2126	-299.5082
ARIMA(4 , 1 , 1)	-323.6374	-307.1679
ARIMA(4 , 1 , 2)	-321.6613	-302.4467
ARIMA(4 , 1 , 3)	-321.1416	-299.1822
ARIMA(4 , 1 , 4)	-321.0936	-296.3892
ARIMA(4 , 1 , 5)	-322.7142	-295.2648

Jelas bahwa model ARIMA (3, 1, 1) tetap model terbaik untuk tipe organik. Selanjutnya juga dilihat untuk tipe konvensional.

ARIMA(0,1,0)	-314.3816	-311.6367
ARIMA(0,1,1)	-314.8324	-309.3425
ARIMA(0,1,2)	-315.2914	-307.0566
ARIMA(0,1,3)	-313.4983	-302.5186
ARIMA(1,1,0)	-314.1839	-308.6941
ARIMA(1,1,1)	-316.3916	-308.1568
ARIMA(1,1,2)	-314.3996	-303.4198
ARIMA(1,1,3)	-314.0204	-300.2957
ARIMA(2,1,0)	-314.2400	-306.0052
ARIMA(2,1,1)	-314.3982	-303.4185
ARIMA(2,1,2)	-312.5569	-298.8322
ARIMA(2,1,3)	-312.0242	-295.5546
ARIMA(3,1,0)	-312.4087	-301.4289
ARIMA(3,1,1)	-313.8272	-300.1025
ARIMA(3,1,2)	-311.9063	-295.4367

Model ARIMA (0, 1, 0) dan (1, 1, 1) merupakan model terbaik untuk tipe konvensional.

Berdasarkan seluruh tabel di atas dan alasan kami yang sama sebelumnya, yaitu kompleksitas data kami menjadi basis kami memilih model yang lebih kompleks dengan AIC atau BIC yang bisa lebih baik dibanding model terbaik lainnya. Dengan demikian, untuk keseluruhan kami mengambil model ARIMA (1, 1, 3), untuk organik adalah ARIMA (3, 1, 1), dan untuk konvensional adalah ARIMA(1, 1, 1).

Didapatkan parameter dari model setiap tipe:

- Keseluruhan:

```
> arima(ts1, order=c(1, 1, 3))
```

Call:

```
arima(x = ts1, order = c(1, 1, 3))
```

Coefficients:

```
      ar1      ma1      ma2      ma3
-0.9181  0.8934 -0.2562 -0.3186
s.e.    0.0991  0.1223  0.1187  0.0958
```

Berdasarkan gambar di atas, maka estimasi parameter untuk model ARIMA (1,1,3) tipe keseluruhan adalah sebagai berikut:

$$\nabla Y_t = -0.9181 \nabla Y_{t-1} + e_t + 0.2562 e_{t-1} + 0.3186 e_{t-2}$$

$$Y_t - Y_{t-1} = -0.9181(Y_{t-1} - Y_{t-2}) + e_t + 0.2562 e_{t-1} + 0.3186 e_{t-2}$$

$$Y_t = 0.0819 Y_{t-1} - 0.9181 Y_{t-2} + e_t + 0.2562 e_{t-1} + 0.3186 e_{t-2}$$

- Organik:

```
> arima(ts2, order=c(3, 1, 1))
```

Call:

```
arima(x = ts2, order = c(3, 1, 1))
```

Coefficients:

	ar1	ar2	ar3	ma1
	-0.7512	-0.3322	-0.4063	0.6290
s.e.	0.1477	0.1058	0.0849	0.1485

Berdasarkan gambar di atas, maka estimasi parameter untuk model ARIMA (3,1,1) tipe organik adalah sebagai berikut:

$$\nabla Y_t = -0.7512 \nabla Y_{t-1} - 0.3322 \nabla Y_{t-2} - 0.4063 \nabla Y_{t-3} + e_t - 0.6290 e_{t-1}$$

$$Y_t - Y_{t-1} = -0.7512(Y_{t-1} - Y_{t-2}) - 0.3322(Y_{t-2} - Y_{t-3}) - 0.4063(Y_{t-3} - Y_{t-4}) + e_t - 0.6290 e_{t-1}$$

$$Y_t = 0.2488 Y_{t-1} + 0.419 Y_{t-2} - 0.0741 Y_{t-3} + 0.4063 Y_{t-4} + e_t - 0.6290 e_{t-1}$$

- Konvensional:

```
> arima(ts3, order=c(1, 1, 1))
```

Call:

```
arima(x = ts3, order = c(1, 1, 1))
```

Coefficients:

	ar1	ma1
	0.6668	-0.8405
s.e.	0.2236	0.1686

Berdasarkan gambar di atas, maka estimasi parameter untuk model ARIMA (1,1,1) tipe organik adalah sebagai berikut:

$$\nabla Y_t = 0.6668 \nabla Y_{t-1} + e_t + 0.8405 e_{t-1}$$

$$Y_t - Y_{t-1} = 0.6668(Y_{t-1} - Y_{t-2}) + e_t + 0.8405 e_{t-1}$$

$$Y_t = 1.6668 Y_{t-1} - 0.6668 Y_{t-2} + e_t + 0.8405 e_{t-1}$$

Setelah mendapat model, sekarang akan dilakukan juga uji asumsi untuk independensi dan normalitas. Hasil dari uji independensi sebagai berikut:

```
> checkresiduals(arima(ts1, order = c(1, 1, 3)))
```

Ljung-Box test

data: Residuals from ARIMA(1,1,3)

Q* = 15.384, df = 19, p-value = 0.6979

Model df: 4. Total lags used: 23


```
> checkresiduals(arima(ts2, order = c(3, 1, 1)))
```

Ljung-Box test

```
data: Residuals from ARIMA(3,1,1)
Q* = 17.801, df = 19, p-value = 0.5358
```

```
Model df: 4. Total lags used: 23
```

```
> #conventional
```

```
> checkresiduals(arima(ts3, order = c(1, 1, 1)))
```

Ljung-Box test

```
data: Residuals from ARIMA(1,1,1)
Q* = 11.808, df = 21, p-value = 0.9447
```

```
Model df: 2. Total lags used: 23
```

Terlihat bahwa ketiganya memenuhi asumsi independensi, berbeda dengan sebelumnya dimana tipe organik tidak memenuhi asumsi. Selanjutnya, juga akan dilihat untuk uji normalitas.

```
> jarque.bera.test(residuals(arima(ts1, order = c(1, 1, 3))))
```

Jarque Bera Test

```
data: residuals(arima(ts1, order = c(1, 1, 3)))
X-squared = 8.0225, df = 2, p-value = 0.01811
```

```
> jarque.bera.test(residuals(arima(ts2, order = c(3, 1, 1))))
```

Jarque Bera Test

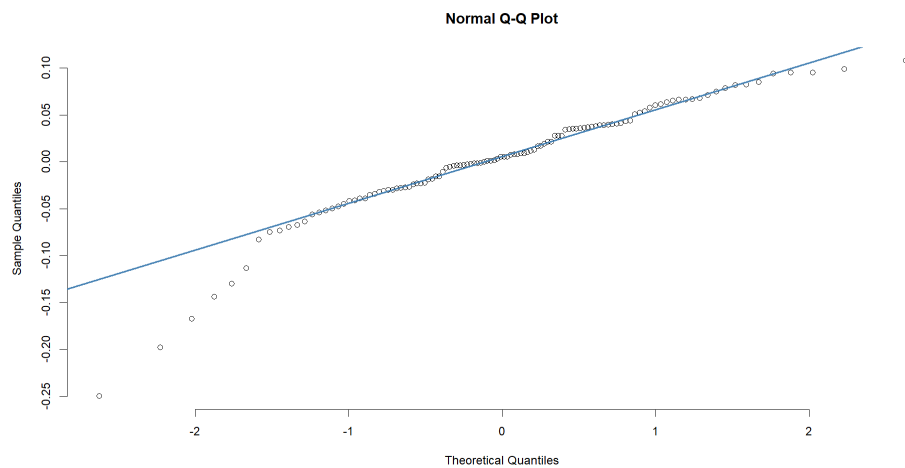
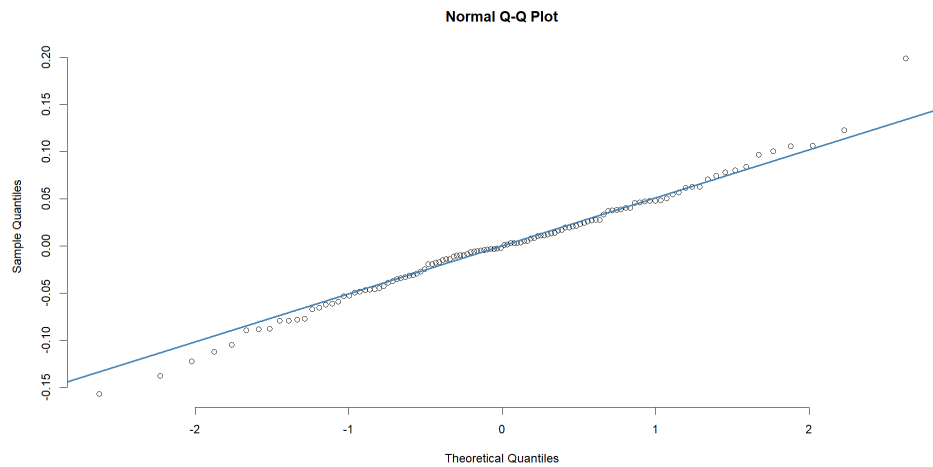
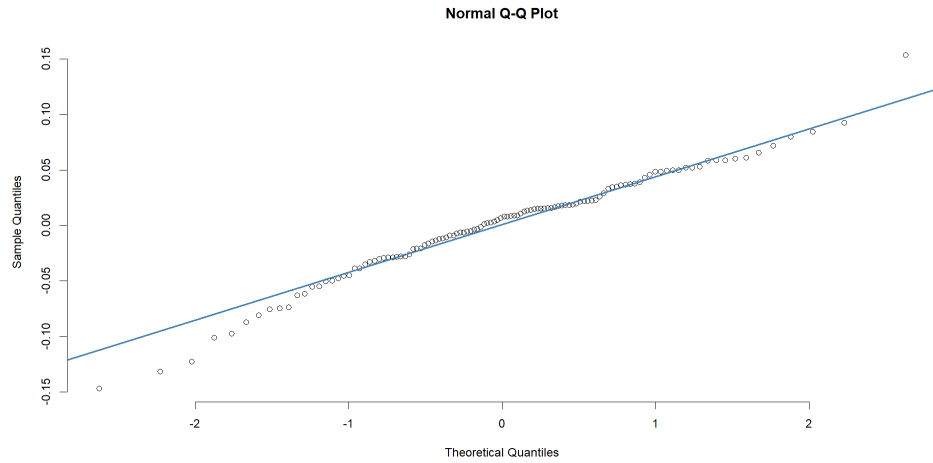
```
data: residuals(arima(ts2, order = c(3, 1, 1)))
X-squared = 4.7533, df = 2, p-value = 0.09286
```

```
> jarque.bera.test(residuals(arima(ts3, order = c(1, 1, 1))))
```

Jarque Bera Test

```
data: residuals(arima(ts3, order = c(1, 1, 1)))
X-squared = 69.252, df = 2, p-value = 8.882e-16
```

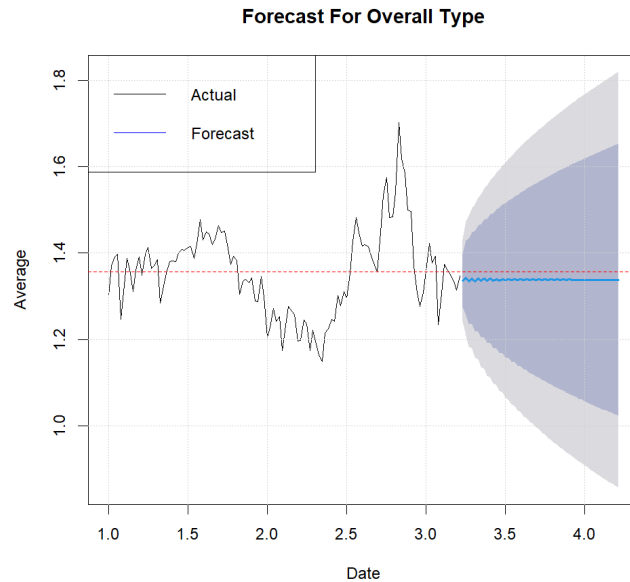
Berbeda dengan sebelumnya, kali ini tipe organik memenuhi asumsi normalitas. Dapat dilihat bahwa tipe konvensional semakin jelas bahwa persebaran pada data runtun waktu tersebut tidak normal sehingga tipe keseluruhan pun terkena dampak dari tipe konvensional. Ini berarti dengan penghilangan outlier, data runtun waktu untuk tipe organik lebih dapat diprediksi dengan model ARIMA. Apabila dilihat lagi pada qq-plot,



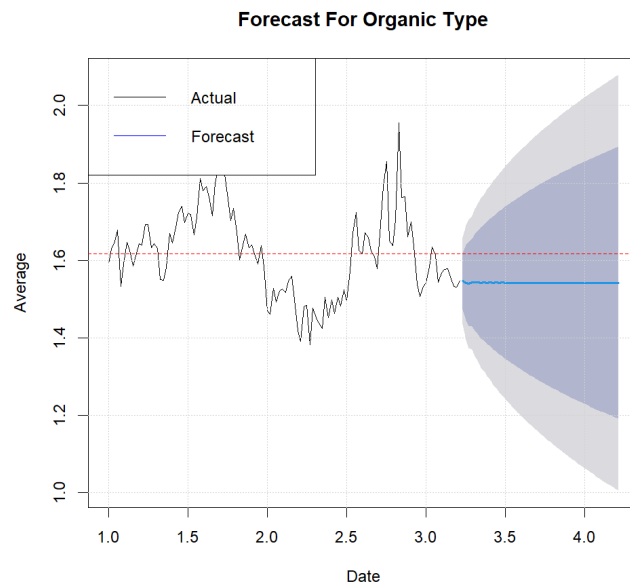
Semakin jelas bahwa tipe konvensional memiliki banyak titik observasi yang melenceng jauh dari garis normal. Ini memberikan p-value yang sangat kecil pada uji hipotesis sehingga tipe konvensional tidak berdistribusi normal.

3.9 Forecasting Lanjutan

Menindaklanjuti data yang sudah diperbarui dan mengikuti jejak langkah analisis data sebelumnya, akan dilakukan peramalan dari runtun waktu yang baru. Berikut hasil ramalan untuk tipe keseluruhan:

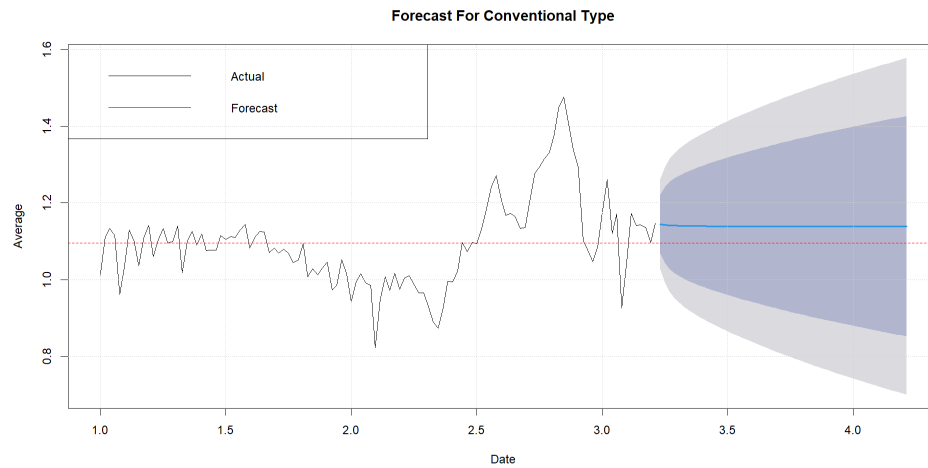


Terlihat bahwa data aslinya jauh lebih stabil sehingga ramalannya pun terlihat lebih masuk akal dibanding sebelumnya. Selanjutnya, akan dilihat juga untuk tipe organik.



Model ini lebih stabil dari sebelumnya, tetapi dapat dilihat tidak sebagus untuk tipe keseluruhan.

Terakhir, akan dilihat juga untuk tipe konvensional.



Model ini terlihat paling stabil dibanding model lainnya, dengan peramalan yang terlihat paling masuk akal dibanding sebelumnya. Akan tetapi, mengingat runtun waktu ini tidak memenuhi asumsi normalitas error, hasil peramalan pun tidak akan secredibel model lainnya.

BAB IV PENUTUP

4.1 Kesimpulan

Melalui model yang kami buat dengan metode peramalan, dapat disimpulkan bahwa pola dan tren harga alpukat di Amerika Serikat cukup layak kami buat dengan model ARIMA dan *seasonal naive*. Secara umum, kami melihat tren mingguan, tetapi lebih jelas pada tren yang kami bagi secara bulanan. Jelas bahwa data kami menunjukkan peningkatan yang drastis mulai dari transisi dari bulan Mei hingga berpuncak di bulan September dan Oktober. Selain itu, kami juga menemukan bahwa pada tahun 2015 dan 2016, harga alpukat untuk seluruh tipe masih relatif stabil, sedangkan pada tahun 2017 mulai terlihat peningkatan dan terutama puncak yang sangat tinggi dibanding tahun-tahun sebelumnya pada bulan September dan Oktober. Tren ini mungkin bisa menunjukkan adanya faktor eksternal atau awal mula kenaikan harga alpukat secara permanen. Berhubung kurangnya informasi kami terkait faktor-faktor yang mungkin menyebabkan tren ini pada tahun 2017, kami dapat menganggap tahun ini hanya sebagai *outlier* yang bisa kami tangani dalam model.

Dalam melakukan analisis runtun waktu, kami menggunakan metode ARIMA juga untuk seluruh tipe alpukat. Secara keseluruhan, dapat disimpulkan bahwa metode ARIMA bisa menjadi peramalan yang bagus dalam jangka pendek memprediksi beberapa minggu saja. Akan tetapi, model ini tidak bisa menangkap kecenderungan tren naik harga dari transisi bulan Mei hingga puncaknya pada bulan September dan Oktober. Saat diuji dengan membagi data runtun waktu kami menjadi data *test* dan *train*, jelas bahwa model tidak begitu mampu memprediksi secara akurat dari data *test*. Kami menyimpulkan bahwa ini disebabkan data *test* mengandung data di kisaran tahun 2017 sehingga pada tahun yang mengalami fluktuasi tinggi, model pun tidak akan bisa memprediksi secara tepat. Setelah kami menghapus tahun 2017 pada runtun waktu kami, kami mendapat peramalan yang jauh lebih baik dibanding sebelumnya. Meski demikian, masalah dasar bahwa model ini tidak bisa menangkap tren kenaikan periodik masih menjadi masalah utama dalam prediksi jangka panjang.

Di sisi lain, *seasonal naive* hanya mengulang data *train* sebagai prediksi. Model ini bisa saja bagus apabila faktor-faktor lainnya yang tidak dianalisis bersamaan dengan metode peramalan pada penelitian ini cenderung konsisten pada masa depan. Terkait keakuratan dari model ini, dengan alasan yang sama pada model ARIMA, bahwa tahun 2017 menunjukkan fluktuasi yang tinggi, model terlihat tidak begitu bagus dalam memprediksi data *test* dari data *train*, tetapi tidak bisa disimpulkan langsung bahwa model kurang cocok. Setelah data runtun waktu 2017 dihilangkan, model tidak menunjukkan titik-titik ekstrem pada tren. Akan tetapi, model ini masih bergantung pada konsistensi seluruh variabel yang tidak diukur pada penelitian ini untuk menjadi model yang baik.

Selain itu, kami menganalisis secara sekaligus pembuatan tiga model untuk tipe keseluruhan jenis buah alpukat, tipe organik, dan tipe konvensional. Dari pemodelan kami, jelas bahwa harga tipe organik dan tipe konvensional berada pada rentang yang berbeda. Tidak hanya

itu, model yang dibuat dari kedua tipe secara terpisah menghasilkan model yang berbeda. Hanya saja, tipe alpukat memiliki kecenderungan kurang bisa diprediksi menggunakan metode ARIMA karena tidak memenuhi asumsi. Ini berarti untuk memprediksi harga alpukat, akan lebih baik jika mengetahui jenisnya terlebih dahulu sebelum memprediksi harga tipe tersebut apabila ingin lebih spesifik dalam prediksi, lalu dapat digunakan metode ARIMA untuk organik dan model lainnya, seperti regresi yang bisa lebih cocok untuk tipe konvensional. Secara keseluruhan, model untuk memprediksi keseluruhan tipe masih cukup signifikan dalam memprediksi harga alpukat selain tipe konvensional secara spesifik setelah penghapusan *outlier*.

Dengan demikian, kami menyimpulkan bahwa pola dan tren harga alpukat di Amerika Serikat memiliki pola dan tren yang nampak meningkat drastis pada transisi dari bulan Mei hingga Oktober secara periodik. Model ARIMA dan *seasonal naive* memiliki kekuatan dan kelemahan masing-masing dimana ARIMA mampu memberikan prediksi yang lebih *up to date* dan cukup bagus dalam prediksi jangka pendek, tetapi tidak dapat menangkap tren periodik dari data runtun waktu. Di sisi lain, model *seasonal naive* menangkap tren periodik dari data runtun waktu, tetapi sangat bergantung pada variabel yang tidak diteliti untuk menjadi model yang bagus. Setiap tipe buah alpukat pun bisa diprediksi secara terpisah untuk mendapat informasi yang lebih spesifik dengan model yang berbeda.

DAFTAR PUSTAKA

- ❖ Centers for Disease Control and Prevention. (2022). Get Your Fats Straight! How to Know Which Fats to Choose. Di akses dari <https://www.cdc.gov/nutrition/data-statistics/know-your-fats/index.html>
- ❖ U.S. Department of Agriculture. (2021). Hass Avocado Production Forecast. National Agricultural Statistics Service. Diakses dari https://www.nass.usda.gov/Statistics_by_State/California/Publications/Fruits_and_Nuts/2021-avocado-hass-production.pdf
- ❖ Ranjan, R., & Ghosh, D. (2020). Optimal Irrigation and Fertilizer Application Strategies for Commercial Avocado Cultivation in California. Sustainability, 12(11), 4616. <https://doi.org/10.3390/su12114616>
- ❖ Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., & Wiltshire, A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 2973-2989. <https://doi.org/10.1098/rstb.2010.0158>

LAMPIRAN

Link drive untuk data dan code R:

https://drive.google.com/drive/folders/1g5mRe7uRzLLXLFnuWFO6PgITr_iETVKK

Persentase kontribusi anggota:

NO	Nama	NPM	Kontribusi	Persentase
1	Abdul Muhaimin	2106703052	Mengerjakan project	100%
2	Andri Naufal Al'afif	2106725085	Mengerjakan project	100%
3	Justin Martinus	2106724832	Mencari data dan mengerjakan project	100%