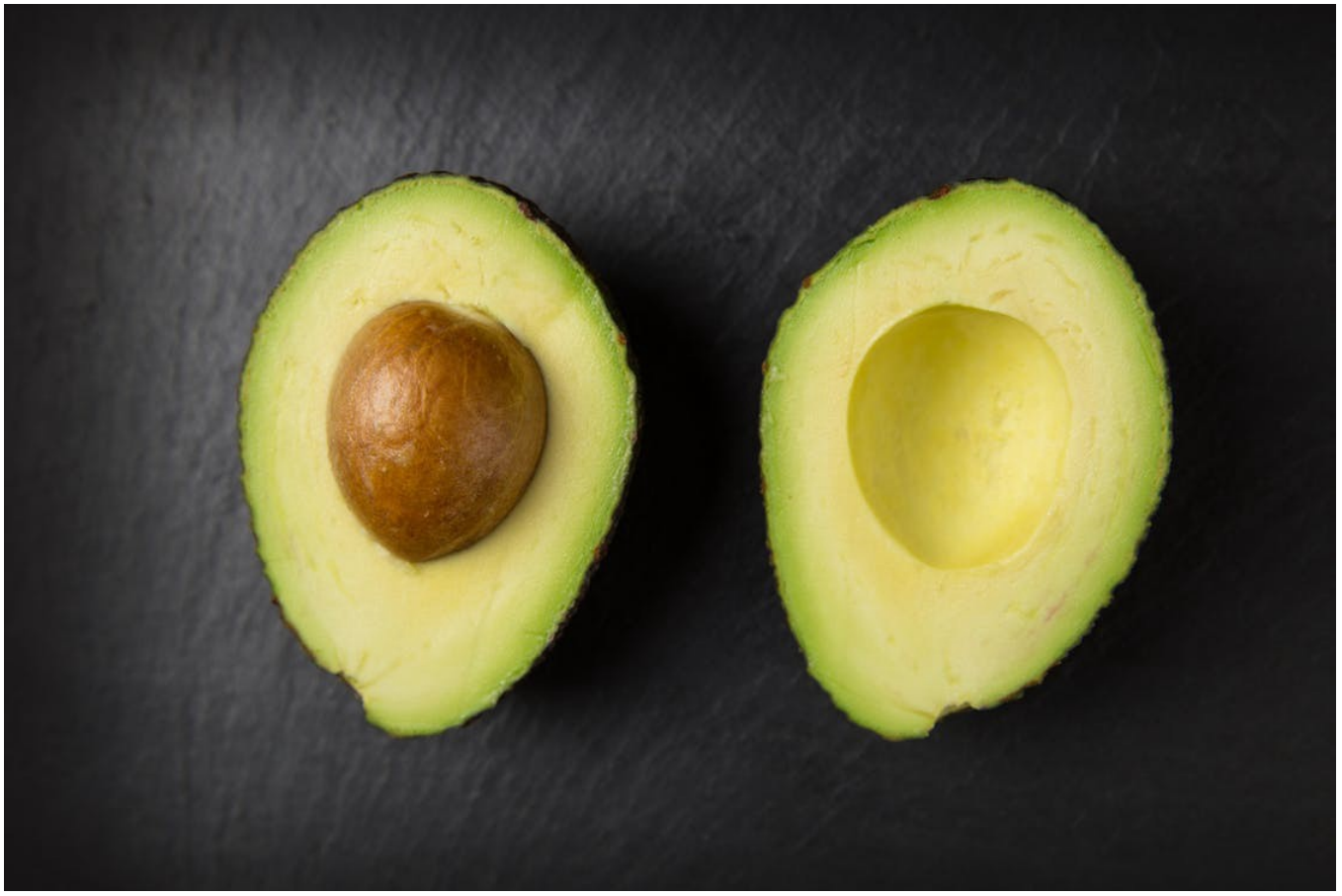
Millennials' Favorite Fruit: Forecasting Avocado Prices with ARIMA Models

Sean Holland Follow

Jul 25, 2019 · 8 min read ★



It's a fruit.

If you know the first thing about finance, you know that the single obstacle standing between Millenials and the white picket fence is avocado toast. If the kids of the Information Age would just get off their butts, walk away from the brunch table, and leave their smashed avocado and poached egg on whole grain behind then maybe they could afford their dream two-bedroom in Ames, Iowa. *Obviously*.

In solidarity with Millenials' avocado toast-induced financial woes, I wanted to take a deeper look at America's favorite fruit. Yes, the avocado is a fruit — a berry even. My research also suggests that the avocado is not, in fact, a conspiracy by toast manufacturers to sell more toast.

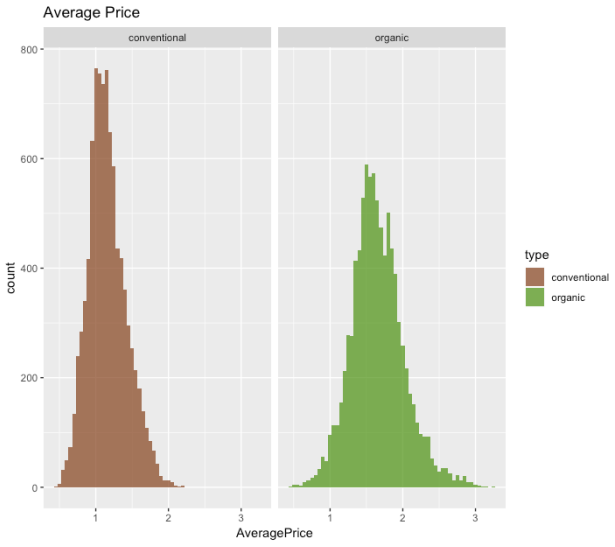
So, in this post, I will be analyzing data from the Hass Avocado Board (HAB) — a trade organization that provides industry professionals with research and data on avocado prices around the world. If you are looking to get a taste of avocado-related research, you can find the data here.

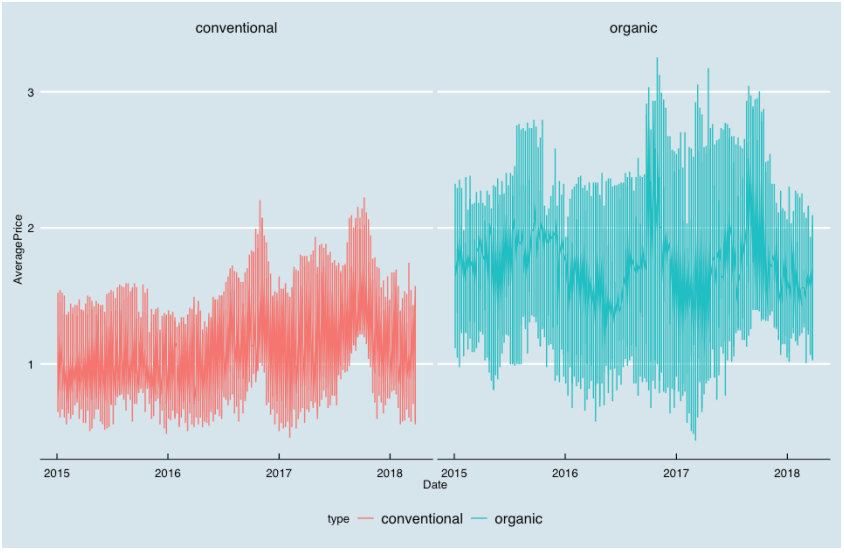
. . .

Data

The HAB dataset is comprised of weekly time series data spanning 168 weeks from January 4, 2015, to March 25, 2018. For this period, the dataset provides average price and total volume data for conventional and organic avocados in 54 regions of the United States, totaling over 18,000 observations.

So let's get into it — what do our avocados look like?



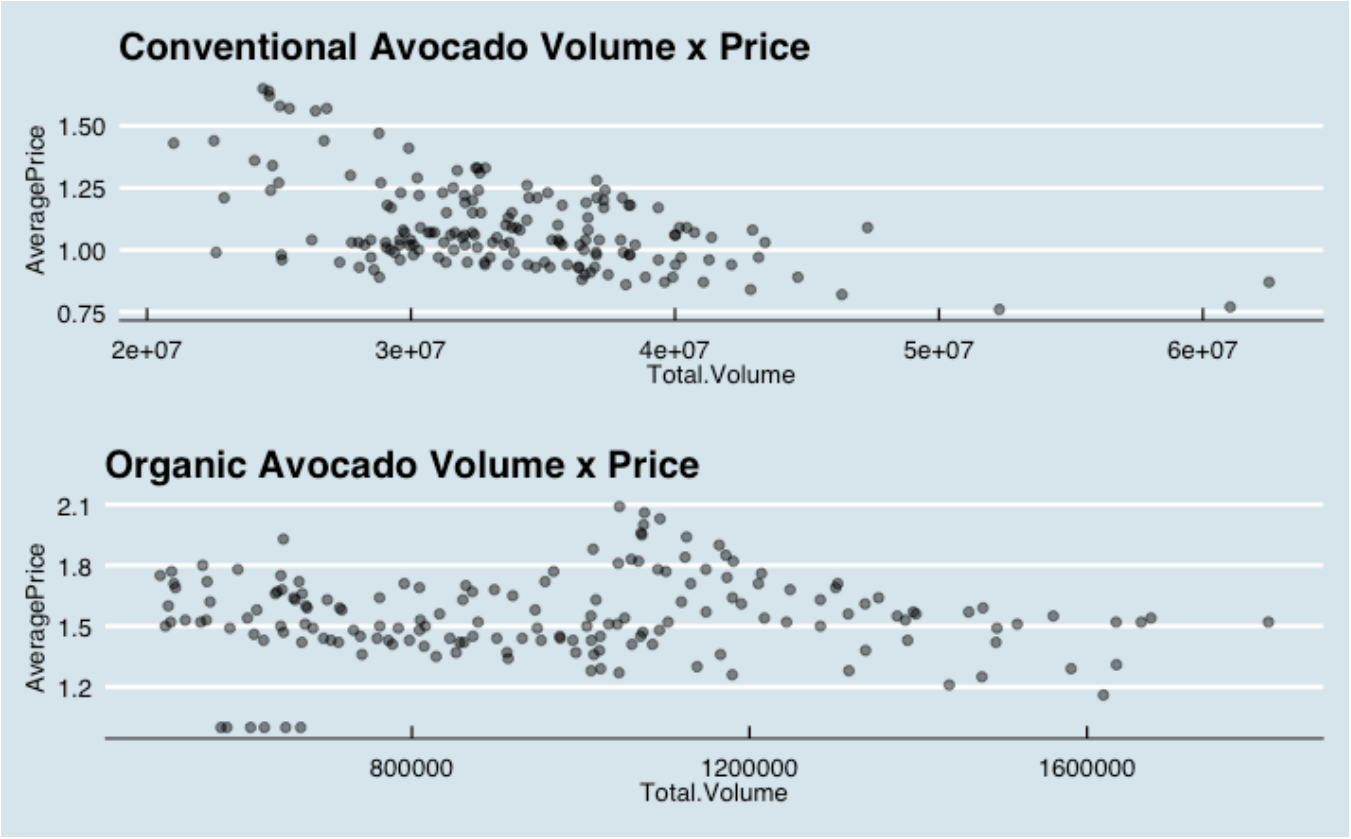


The eye test tells us that organic avocado prices are both higher on average and more volatile than conventional avocados. Digging into the numbers, conventional avocados

cost on average $1.16 across all regions with a standard deviation of $0.26. Meanwhile, organic avocados cost on $1.65 on average with a standard deviation of $0.36.

Looking at avocados sold, however, reveals that organic avocados only constitute about 3 percent of all avocados sold in the US over this period. Between January 2015 and March 2018, consumers bought over 33 million conventional avocados and under one million organic avocados.

We know that conventional avocados are more expensive on average and vastly outpace organic avocados in terms of sales. So, this raises the question — what is the nature of the relationship between price and volume sold?



(Conventional) Pearson's Correlation = -0.51 p-value<0.001 | (Organic) Pearson's Correlation = 0.02 p-value=0.75

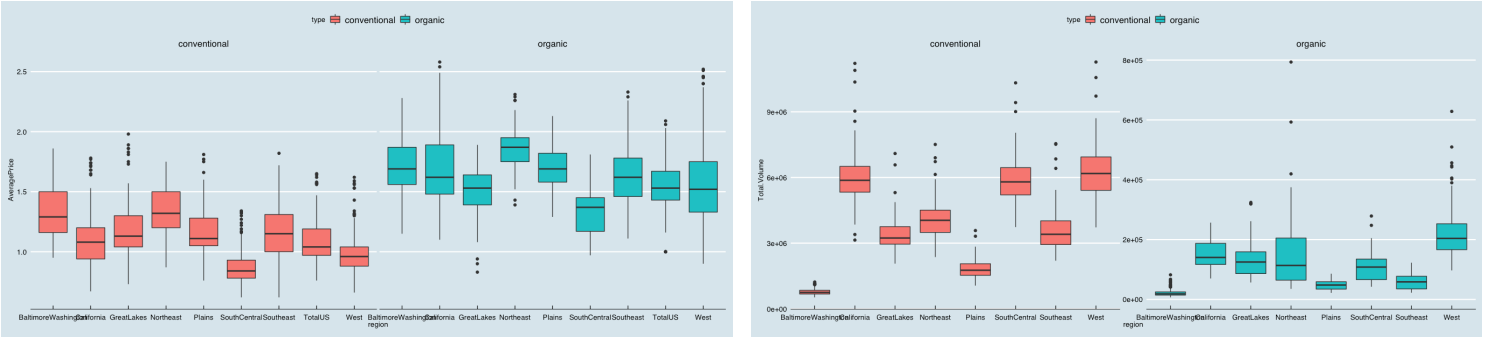
For the 168 weeks studied, conventional avocados demonstrated a moderate negative correlation between average prices and units sold. Nothing out of the ordinary here, rather basic principles of economics — quantity demanded increases as price decreases.

Organic avocados tell a different story. There is no statistically significant relationship between the price tag and units sold for organic avocados. Demand is flat, meaning that

regardless of price, consumers are unlikely to significantly change the quantity they demand. Context is important in understanding this relationship.

Research has shown that consumers with higher incomes, young children, and/or college degrees are more likely to be “devoted” organic consumers than their peers. In this context, “devoted” indicates that the consumer spends a higher-than-average portion of their household income on organic fruits and vegetables. Therefore, it is reasonable to think that regular organic avocado consumers belong to a non-typical segment of the population. Unfortunately, I don’t have the resources to find the causal factors behind this flat relationship, but I can certainly (and will gladly) venture a guess. Consumers who regularly buy organic goods are (probably) more likely to do so out of some principle. Perhaps the consumer believes organic foods to be healthier or more ethical than non-organic foods. Therefore, the organic consumer would not likely consume non-organic foods due to simply shifts in organic food prices, giving us the flat demand curve for organic avocados.

. . .



Right: Average Price (USD) | Left: Total Volume (Avocados Sold)

Next, I wanted to take a quick look at prices and volume by region. Northeasterners (my people) and Washingtonians (also my people) pay amongst the highest rates for organic and conventional avocados in the country. Northeasterners pay $1.34 on average for conventional and $1.86 for organic avocados. Meanwhile, the national average for avocado prices stands at $1.09 for conventional and $1.55 for organic.

In terms of volume, Westerners consume the most avocados, both conventional and organic, per week on average. Westerners consumed over 6,180,000 conventional and

220,000 organic avocados per week. Compared to national averages, Westerners consume about 18 percent of all conventional and 23% of all organic avocados sold in the US.

That's a lot of avocado toast.

. . .

Forecasts

I think it's pretty safe to say that the ship has sailed for Millenials in terms of being able to afford that two-bedroom I mentioned earlier. I mean Americans are spending around $7,000,000 a week on avocados so all hopes of the white picket fence should be dashed for Millenials.

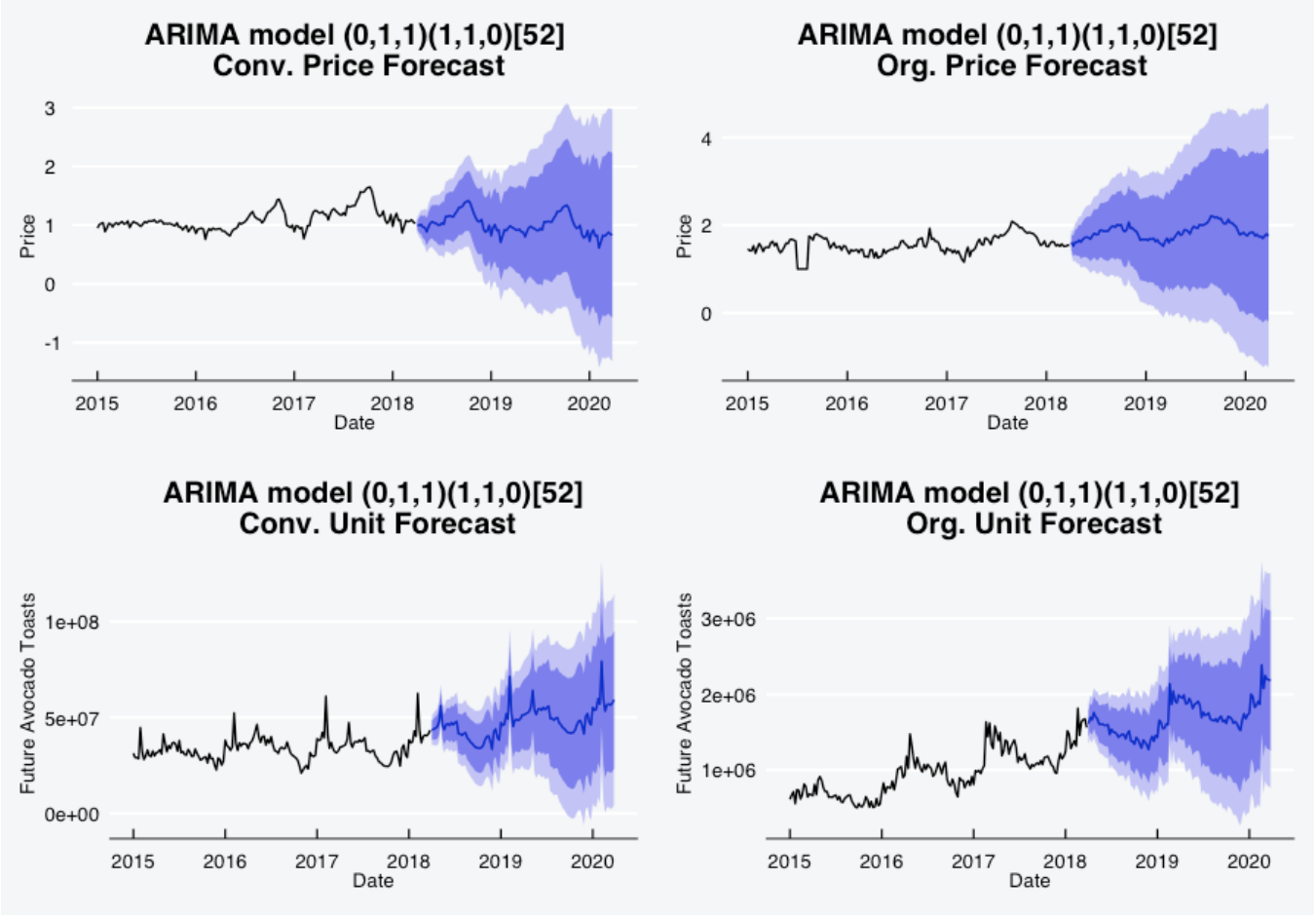
Thankfully, according to the Pew Research Center, I belong to the Gen Z cohort. Of course, this means that I need to have some idea of how the avocado economy is going to affect *my* chances at picket-fence related dreams.

To do so, I turned to ARIMA forecasting. **A**uto-**R**egressive **I**ntegrated **M**oving **A**verage models are a general class of models that allow for forecasting models to be made stationary through "differencing." Stationarity is an important condition in forecasting, as nonstationarity can undermine the reliability of traditional confidence intervals, hypothesis tests, and forecasts.

ARIMA models can be broken down into three components, AR, I, and MA. First, the AR component describes a linear multiple regression with a given number of previous values of *Y* as the predictors for *Y(t)*. The I component describes the order of differencing necessary to make the model stationary. The MA component is similar to the AR component in that it introduces a lagged value into the model. However, the lagged value is the previous error terms, not the previous values.

Thankfully for an ARIMA novice such as myself, Rob Hyndman's "Forecast" package in R comes equipped with the auto.arima function. Auto.arima tests several ARIMA models and selects the best one based on Akaike Information Criteria, providing an excellent starting point from which to select a final model.

Using Augmented Dickey-Fuller tests, which tests for unit roots that can bias the OLS estimators, I concluded that one order of differencing was necessary for each of my four models — conventional prices, conventional volume, organic prices, and organic volume. Additionally, I elected to use seasonal ARIMA models, which contain another set of AR, I, and MA parameters for a given period (in this case 52 weeks or one year).




```
Model Information:

Call:
arima(x = cp_ts, order = c(0, 1, 1), seasonal = c(1, 1, 0))

Coefficients:
         ma1     sar1
      0.1332  -0.2508
s.e.  0.0930   0.1349

sigma^2 estimated as 0.004435:  log likelihood = 147.96,  aic = -289.91

Error measures:
                      ME       RMSE        MAE       MPE     MAPE      MASE       ACF1
Training set -0.0001794556 0.05518001 0.03557149 -0.112153 3.233187 0.2158643 0.0145852
```

```
Model Information:

Call:
arima(x = op_ts, order = c(0, 1, 1), seasonal = c(1, 1, 0))

Coefficients:
          ma1     sar1
      -0.1429  -0.5056
s.e.   0.1064   0.0950

sigma^2 estimated as 0.0189:  log likelihood = 57.89,  aic = -109.78

Error measures:
                      ME      RMSE        MAE         MPE     MAPE      MASE       ACF1
Training set 0.001428955 0.1139049 0.06455178 -0.01954315 4.172968 0.322621 0.01597786
```

```
Model Information:

Call:
arima(x = cv_ts, order = c(0, 1, 1), seasonal = c(1, 1, 0))

Coefficients:
          ma1     sar1
      -0.5369  -0.1126
s.e.   0.0894   0.1387

sigma^2 estimated as 1.488e+13:  log likelihood = -1924.3,  aic = 3854.61

Error measures:
```

```
Model Information:

Call:
arima(x = ov_ts, order = c(0, 1, 1), seasonal = c(1, 1, 0))

Coefficients:
          ma1     sar1
      -0.5902  -0.5275
s.e.   0.0969   0.0839

sigma^2 estimated as 1.8e+10:  log likelihood = -1542.88,  aic = 3091.76
```

```
Error measures:
                     ME    RMSE     MAE       MPE     MAPE      MASE        ACF1
Training set 20729.28 3195997 1777060 -0.3072458 5.194414 0.4605576 -0.02555861
```

```
Error measures:
                     ME    RMSE     MAE       MPE     MAPE      MASE      ACF1
Training set 3815.841 111151.6 67305.11 -0.1530083 5.995722 0.2404425 0.0847484
```

So, what can we learn from our models? Well, first of all, our average price models seem to show us that avocado prices reach their peak each year in the Autumn. In October 2017, conventional avocados peaked at $1.65 and our model forecasts that Autumn spikes will continue to occur through 2018 and 2019. However, the model predicts that conventional avocado prices will trend downwards over the following two years. Between 2015 and 2018, the average conventional avocado price was $1.09. Between 2018 and 2020, the forecasted average price is $1.01, a decrease of about 7 percent.

In stark contrast, the model forecasts that organic avocado prices will experience moderate growth, bolstered by less volatility. The forecasted average price between 2018 and 2020 is $1.84, compared to $1.55 between 2015 and 2018. The model certainly demonstrated a moderate upward trend for organic avocado prices.

However, the growth in average prices is also due in part to decreased volatility, with the standard deviation of prices down to $0.16 for the forecasted period from the $0.36 standard deviation experienced between 2015 and 2018.

Both conventional and organic avocados are forecasted to experience strong growth in terms of units sold over the 2018–2020 period. Conventional avocado production is forecasted to grow over 40 percent from 33.7 million to 47.6 million avocados sold each week on average.

Organic avocados are forecasted to experience even more meteoric growth. A consistent growth trend between 2015 and 2018 is predicted to continue through 2019 with a whopping 74 percent growth between 2018 and 2020. Between 2015 and 2018, suppliers sold about 970,000 organic avocados weekly. Between 2018 and 2020, the number is forecasted to jump to *1.68 million* organic avocados.

Discussion

I think these models are really exciting. First, in terms of pattern recognition, it is interesting to see the models picking up on seasonal trends. The price models captured the Autumn price hikes and the volume models captured the spikes that occur in avocado volume each February.

Additionally, the diagnostics look pretty good. Mean Absolute Percentage Error, or MAPE, measures the error between the forecast and observed values, so a lower value indicates a better fitting model. The highest MAPE amongst the four models is for the organic volume model, at around 6 percent, indicating the model is about 94 percent accurate. I am very happy with this level of fit and I am confident in all of the models. At least, I am as confident as an forecasting newbie could be.

If I had to editorialize (which I will gladly do) I would say that the growth forecast for organic avocado volume is somewhat bullish. I am not sure how realistic 74 percent growth is over only two years, but testing several models yielded similar results.

Alternatively, I think that the conventional avocado price model is somewhat bearish. With the Autumn price hike increasing each year since 2015, I was surprised that the model forecasted overall price decreases. However, if it ends up being the case through 2020, you will not see me complaining at brunch.

. . .

Conclusion

So there it is folks — my little introduction to ARIMA forecasting and with a worthy test subject like avocados. And although conventional avocado prices might be decreasing, it looks like the increased volume of avocados sold is going to keep yet another generation from attaining suburban, property-owning paradise.

I hope you enjoyed my foray into forecasting and I hope to return to this subject soon with more improved ARIMA models and other more complex models. And hey, maybe invest in organic avocado growers — you just might see a 74 percent return.

Well, probably not.

Food Millennials Statistics Forecasting Avocado