Predicting Sales of Hass Avocados

STA440 Individual Project

Outline

- Background
- Research Questions
- Existing Literature
- Methodology
- Results
- Conclusion
- Limitations and Future Steps
- Q & A

Background

Research Questions

From the perspective of a grower/marketer,

Can we predict the **sales volume** of avocados?

From the perspective of a buyer/consumer,

Can we predict the **price** of avocados?

Data

- Kaggle Avocado Prices dataset ← Hass Avocado Board
- Hass avocados, a cultivar of avocados
- Per-unit prices, total volumes sold, regions, and types of Hass avocados
- Weekly data from the January 2015 to March 2018

Research Questions

From the perspective of a grower/marketer,

Can we predict the **sales volume** of avocados?

From the perspective of a buyer/consumer,

Can we predict the **price** of avocados?

Research Questions

From the perspective of a grower/marketer,

Using the Kaggle dataset, can we reasonably predict the total number of conventional Hass avocados sold in the U.S.?

From the perspective of a buyer/consumer,

Can we predict the **price** of avocados?

Research Questions

From the perspective of a grower/marketer,

Using the Kaggle dataset, can we reasonably predict the total number of conventional Hass avocados sold in the U.S.?

From the perspective of a buyer/consumer,

Can we reasonably predict the average per-unit price of conventional Hass avocados in the U.S.?

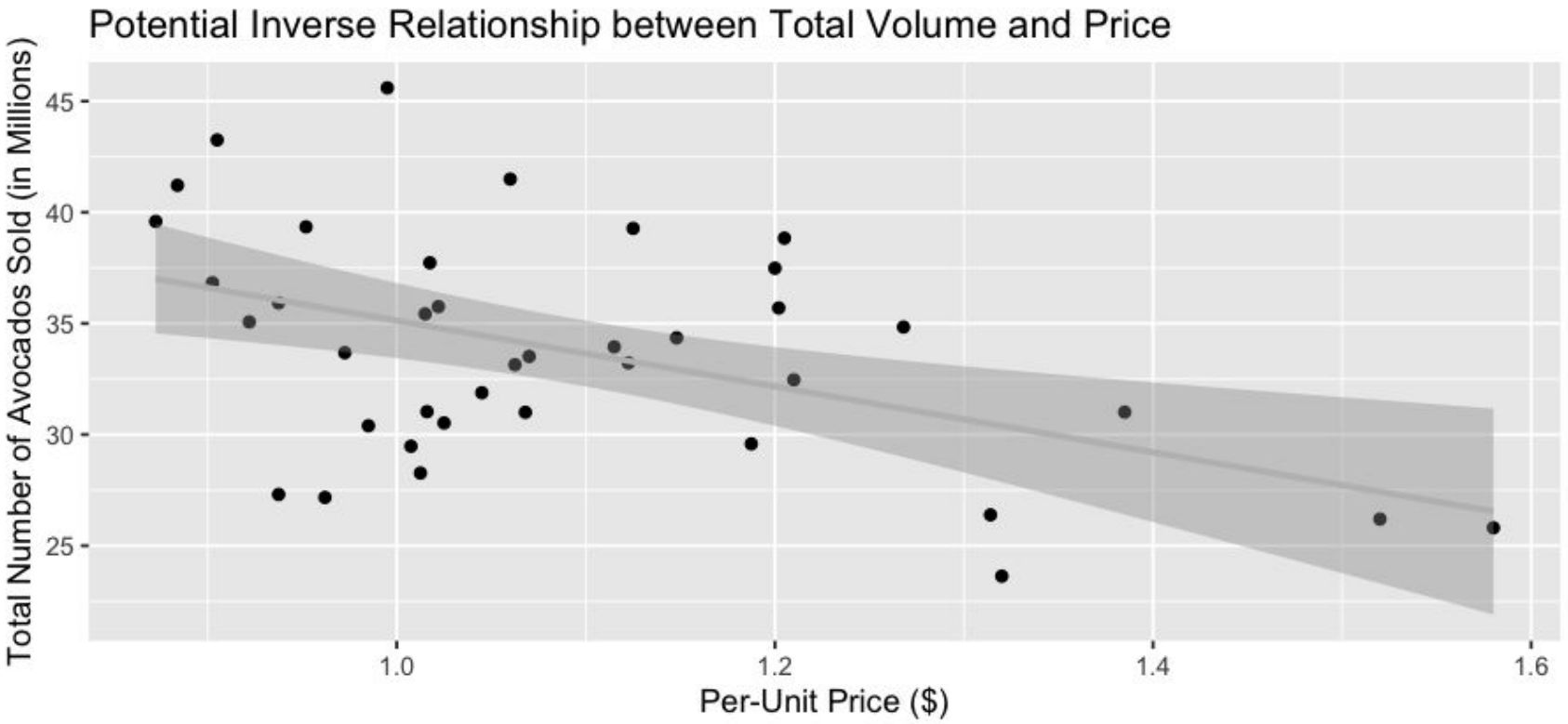
Prediction Range?

Research Question #1

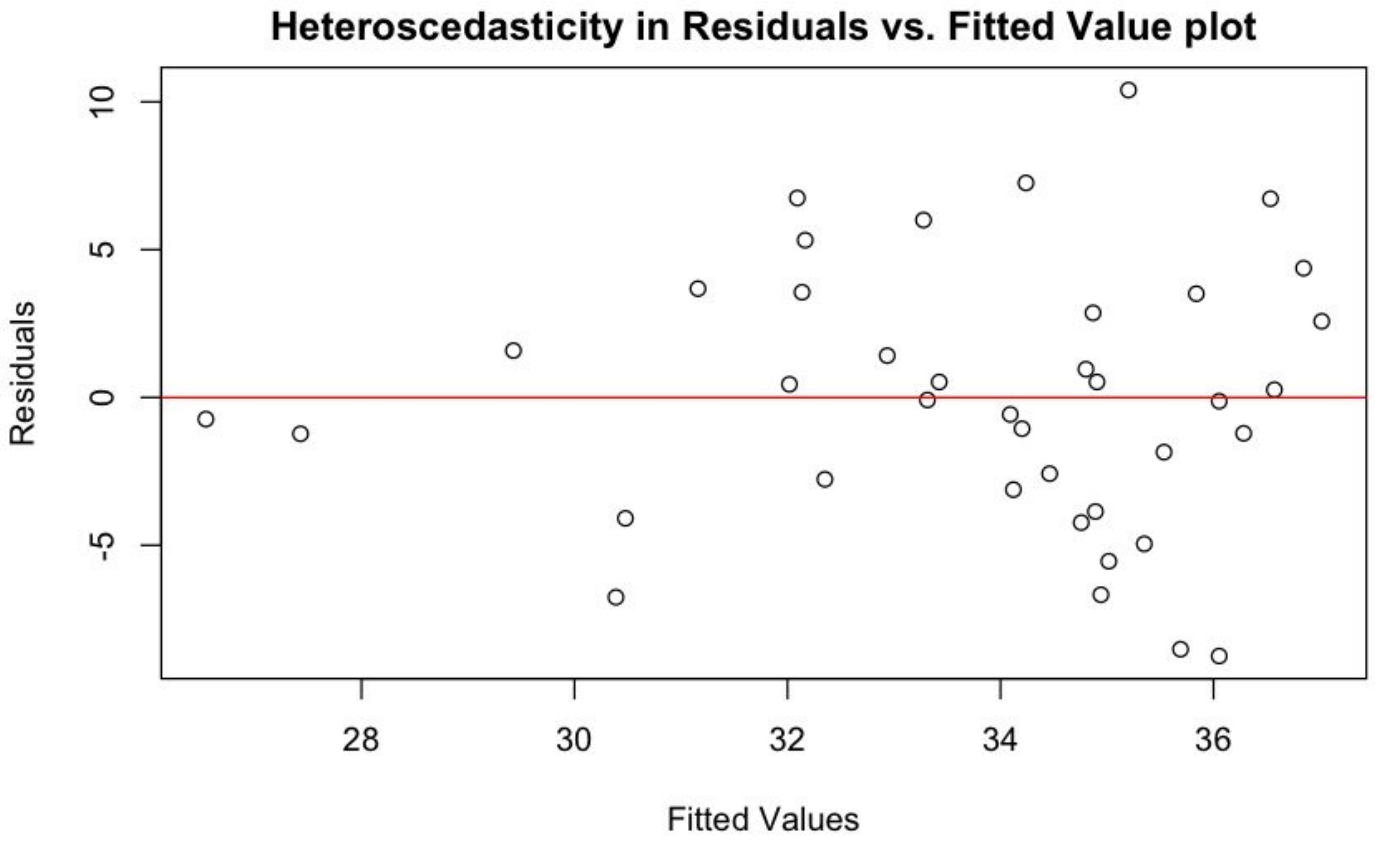
From the perspective of a grower/marketer,

Using the Kaggle dataset, can we reasonably predict the total number of conventional Hass avocados sold in the U.S. from March 2017 to March 2018?

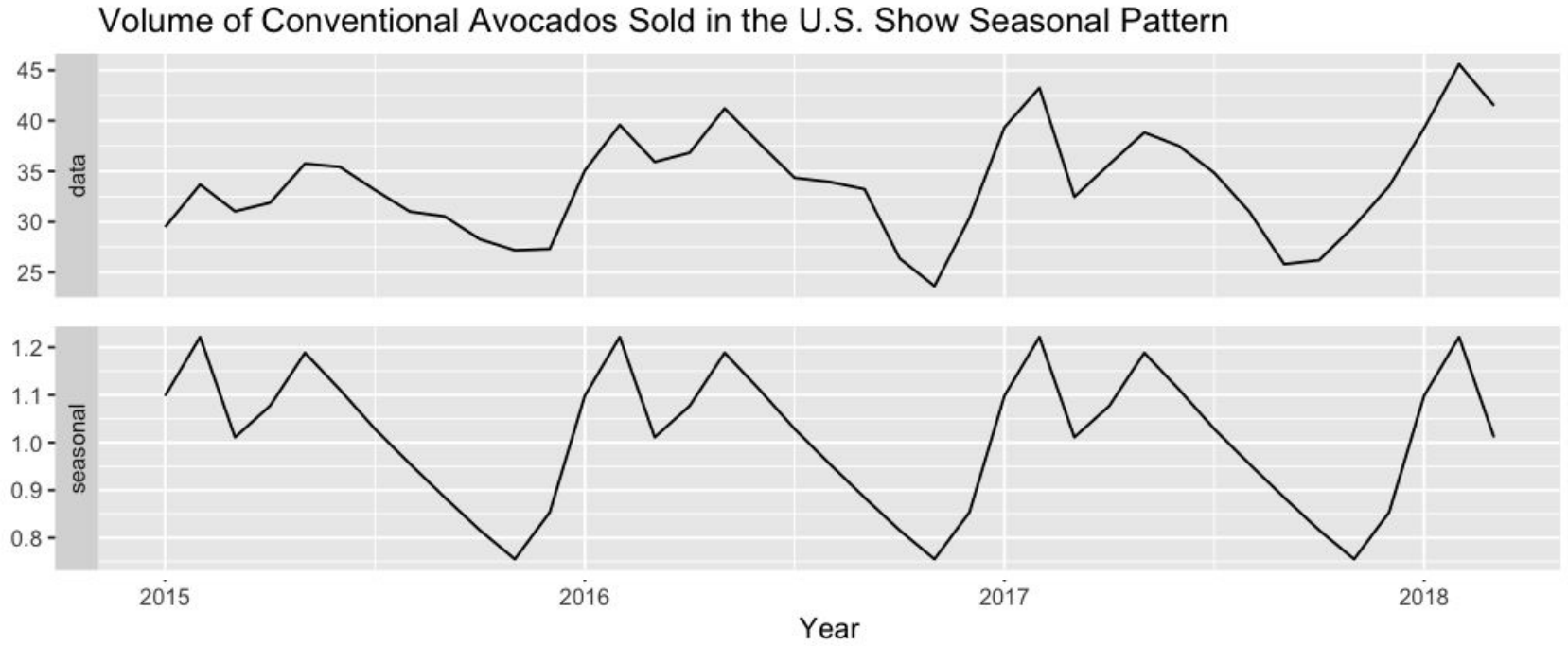
Volume vs. Price?



OLS…Maybe Not?



A Closer Look at Sales Volume



Alternative Model

SARIMA, or

Seasonal AutoRegressive Integrated Moving Average model

SARIMA

$$\text{Seasonal ARIMA } (p, d, q) \times (P, D, Q)_s$$

SARIMA

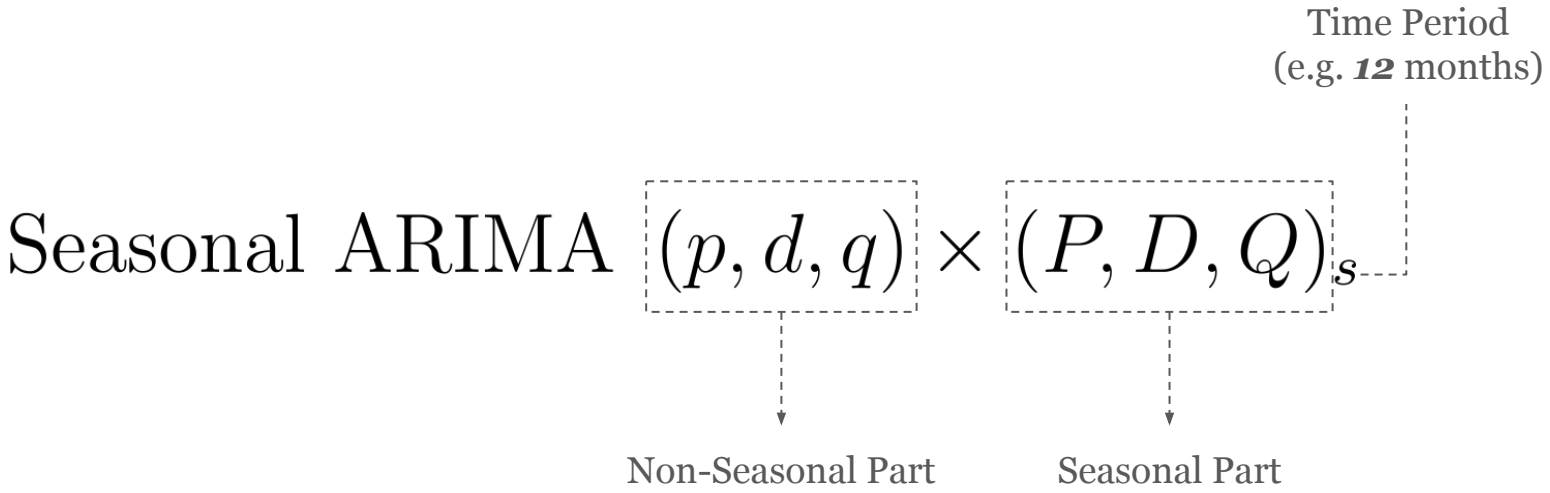
$$\text{Seasonal ARIMA } \boxed{(p, d, q)} \times (P, D, Q)_s$$

Non-Seasonal Part

SARIMA

$$\text{Seasonal ARIMA } \underbrace{(p, d, q)}_{\text{Non-Seasonal Part}} \times \underbrace{(P, D, Q)_s}_{\text{Seasonal Part}}$$

SARIMA


Time Period
(e.g. 12 months)
Seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$
Non-Seasonal Part
Seasonal Part

SARIMA

$$\text{Seasonal ARIMA } (p, d, q) \times (P, D, Q)_s$$

Order of
AutoRegression
e.g. AR(1)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

SARIMA

Order of Differencing

$$\text{Seasonal ARIMA } (p, d, q) \times (P, D, Q)_s$$

Order of AutoRegression e.g. AR(1)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

SARIMA

Order of
Differencing

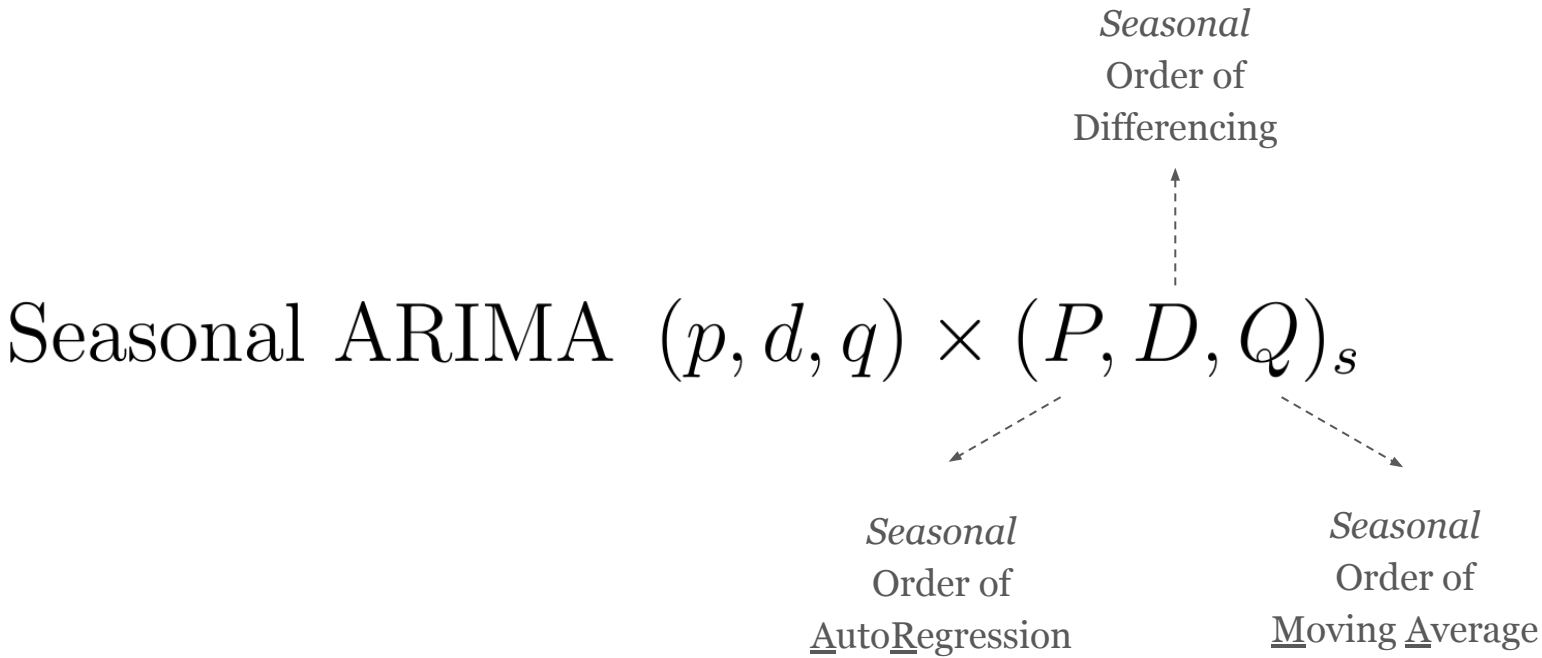
$$\text{Seasonal ARIMA } (p, d, q) \times (P, D, Q)_s$$

Order of
AutoRegression
e.g. AR(1)

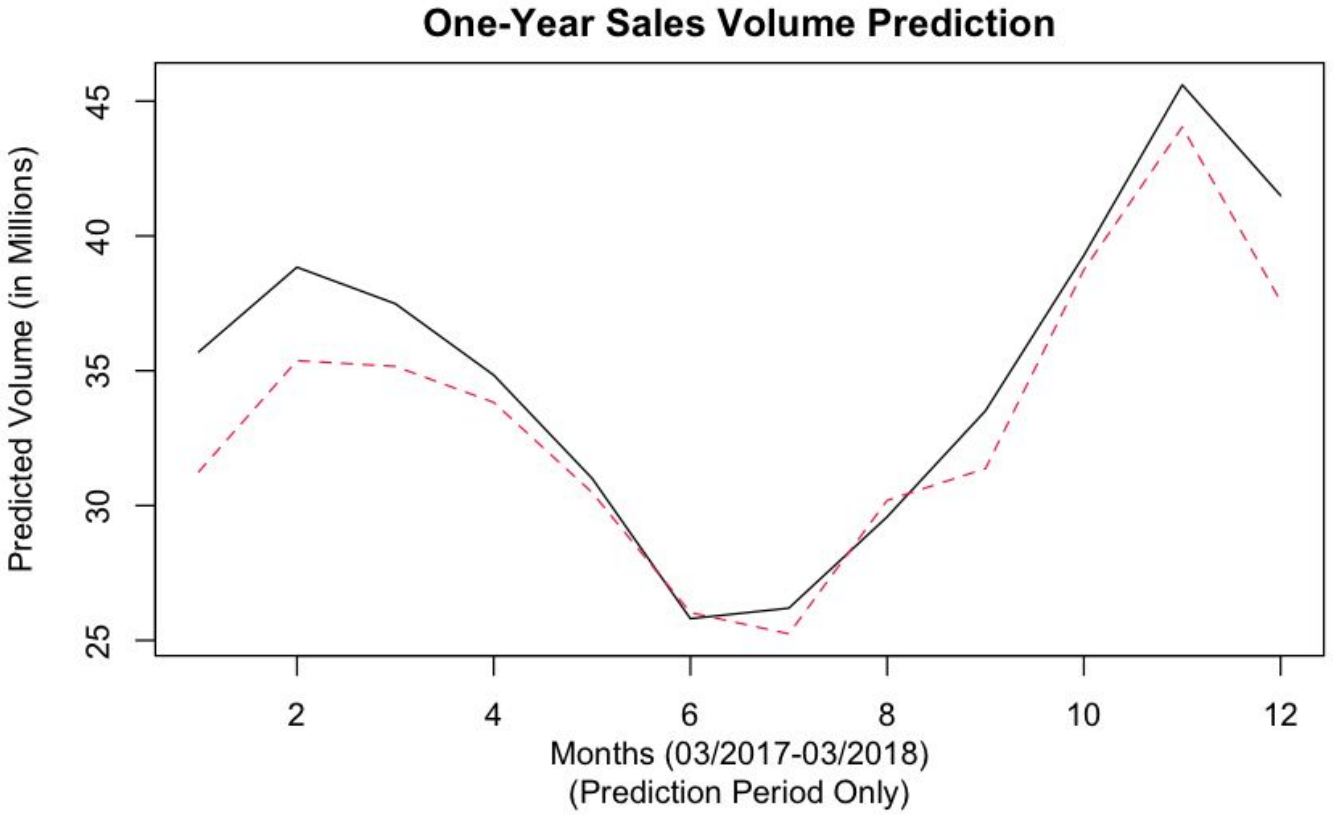
Order of
Moving Average
e.g. MA(1)

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1}$$

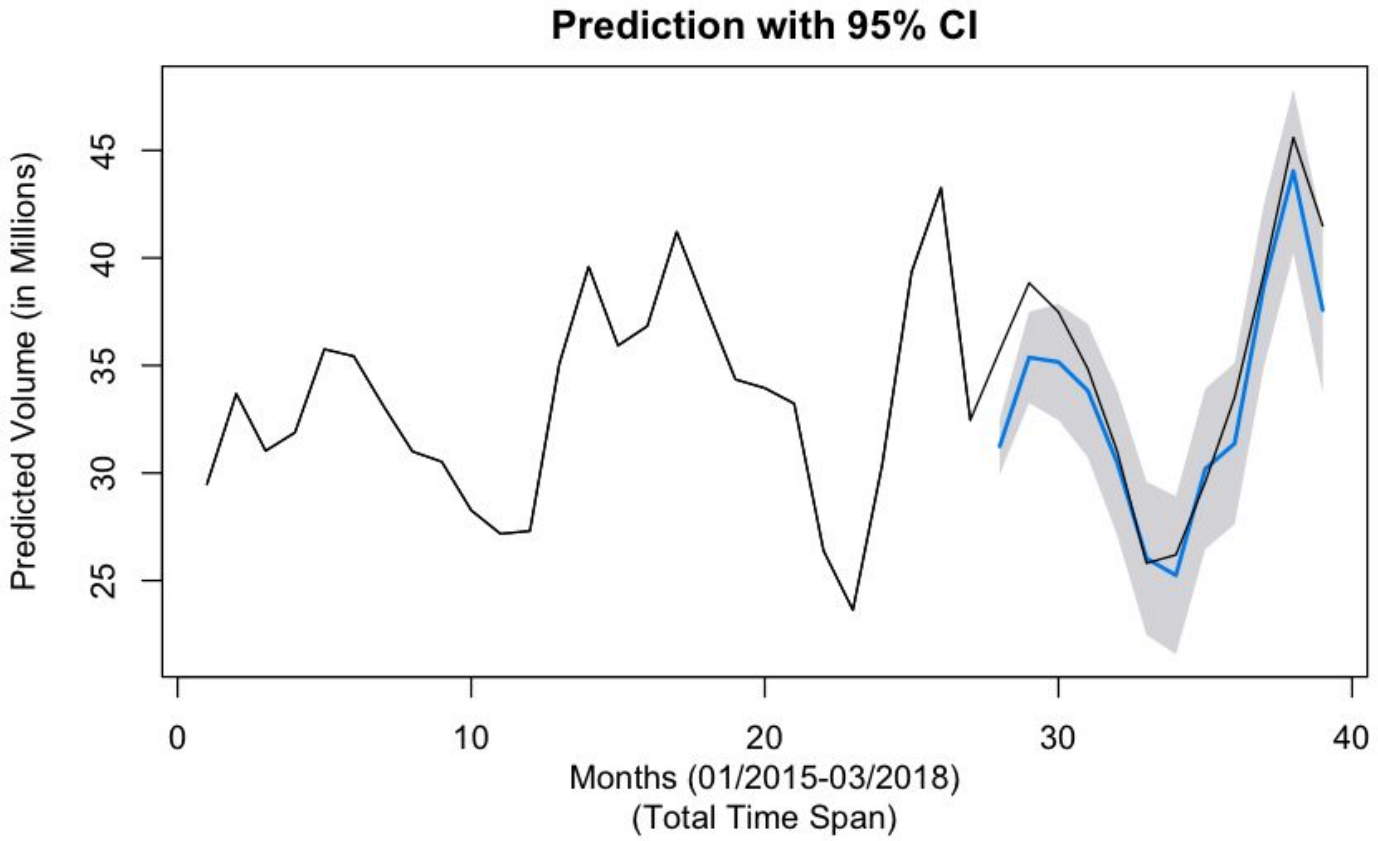
SARIMA


Seasonal
Order of
Differencing
Seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$
Seasonal
Order of
AutoRegression
Seasonal
Order of
Moving Average

Prediction of Sales Volume



Prediction of Sales Volume (With Uncertainty)

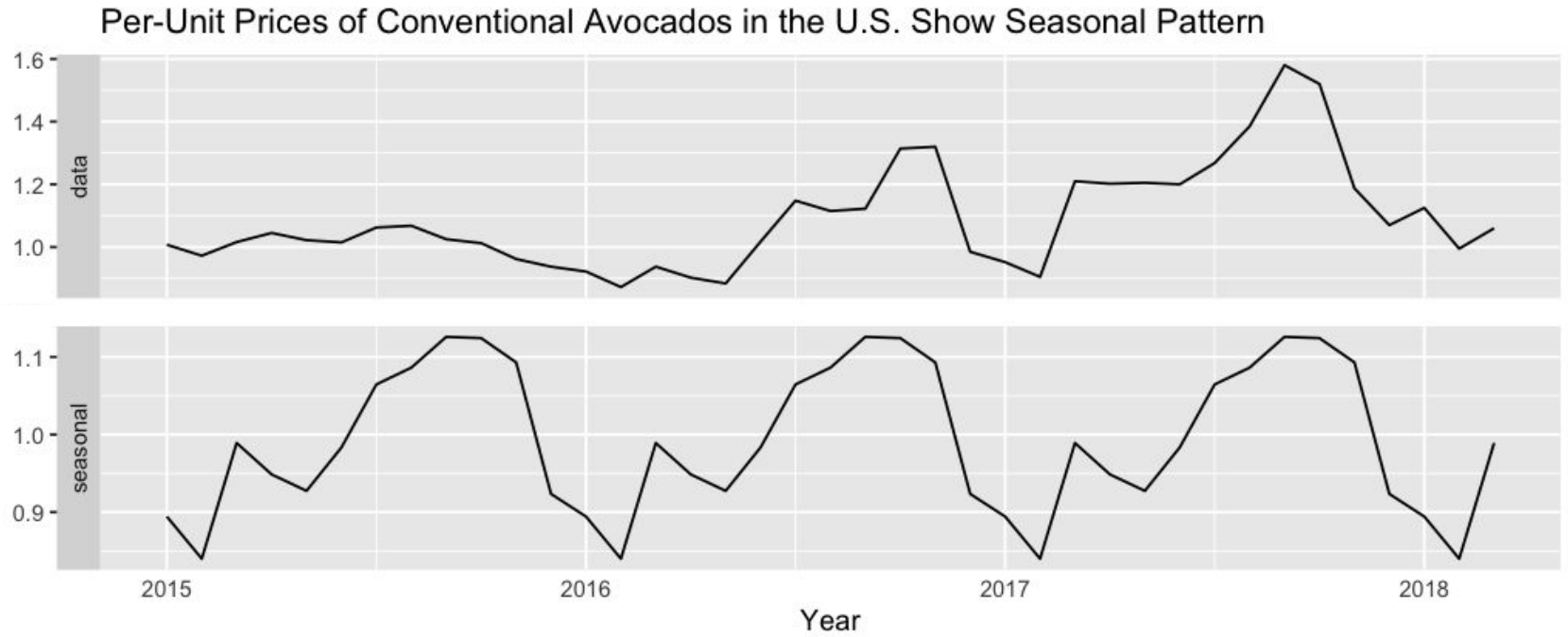


Research Question #2

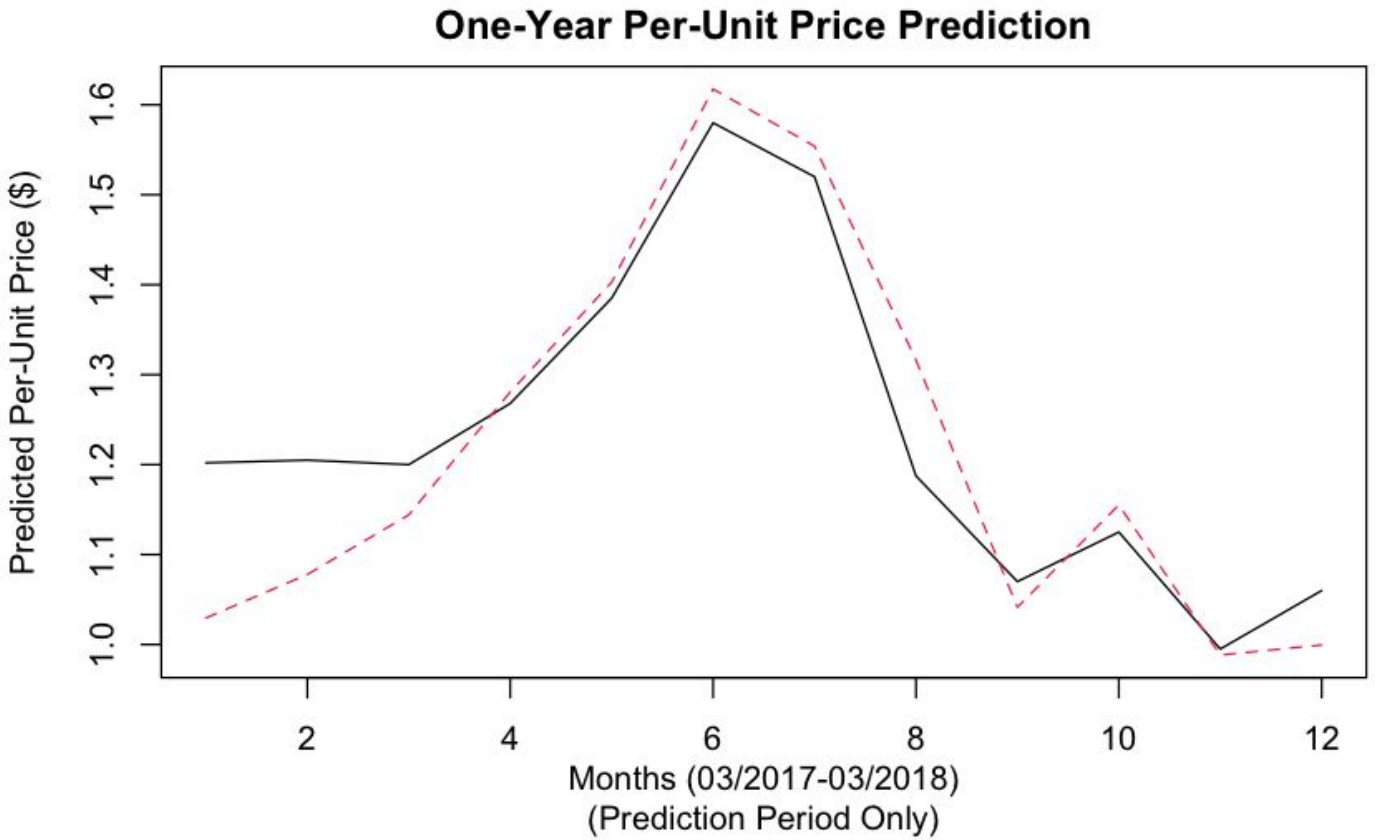
From the perspective of a buyer/consumer,

Can we reasonably predict the average per-unit price of conventional Hass avocados in the U.S. from March 2017 to March 2018?

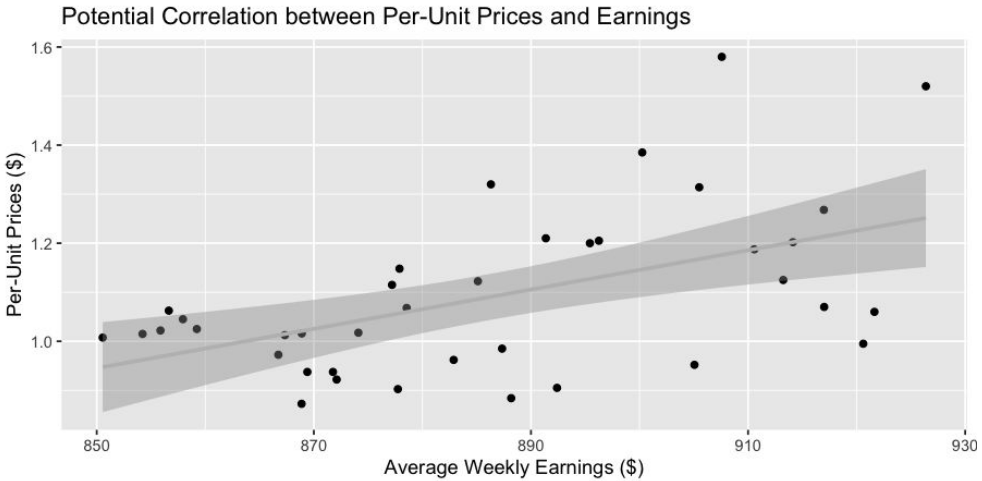
A Closer Look at Per-Unit Prices

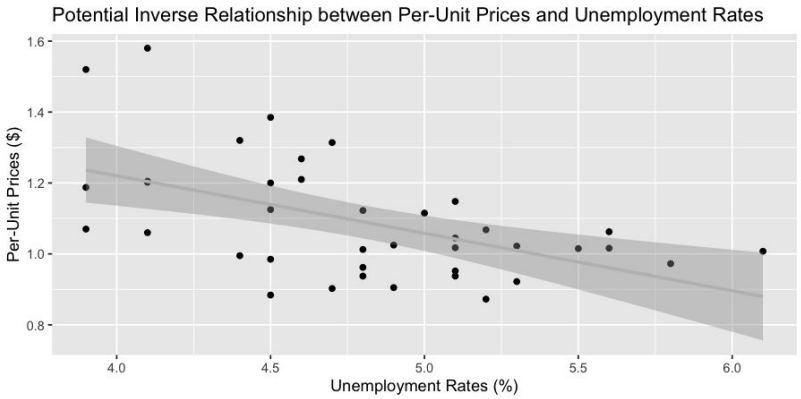


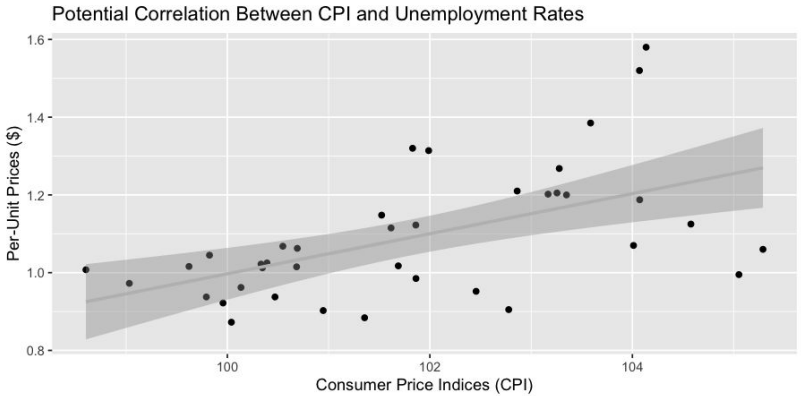
Price vs. Volume Sold…or Not?



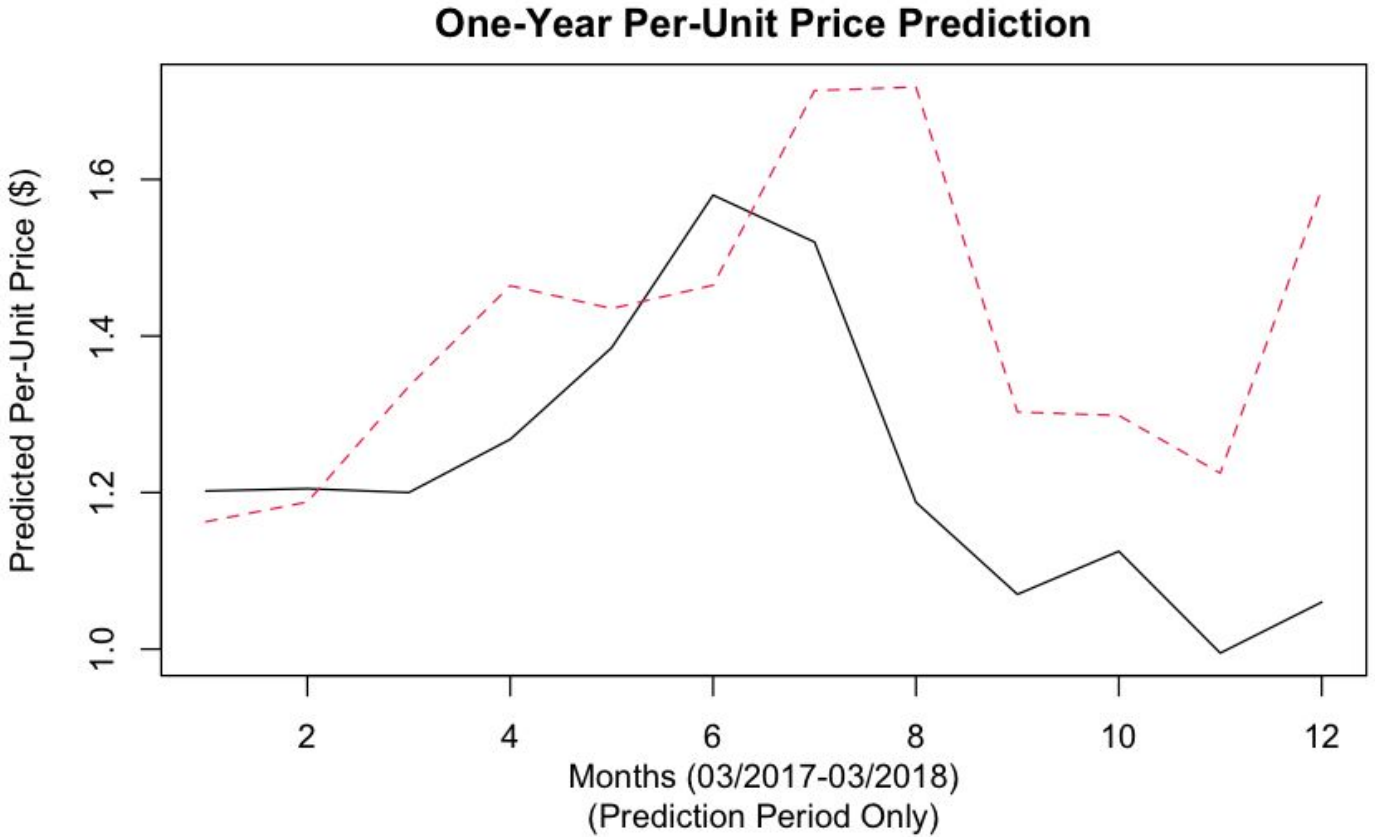
Additional Predictors?







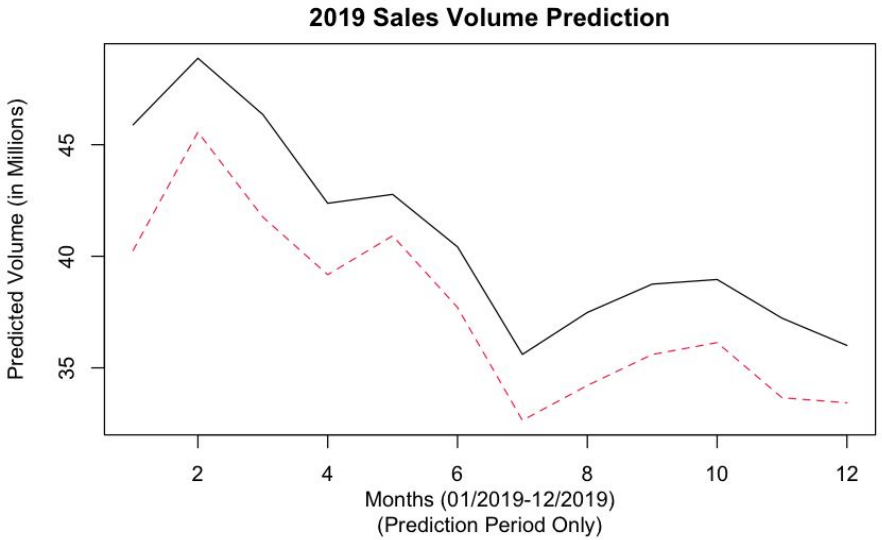
Per-Unit Price Prediction

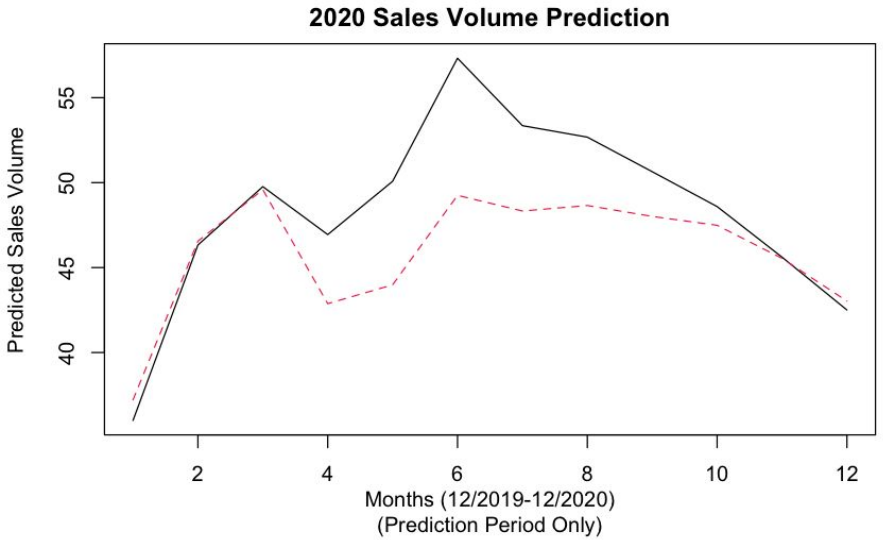


New Data?

- 2015-2020 Data from the Hass Avocado Board
- Same format, same variables
- Existing literature:
    - Study by Evans et al. (2009) predicted avocado prices from 2009-2010

Further Analysis?





Conclusion and Discussion

From the perspective of a grower/marketer,

Using the Kaggle dataset, can we reasonably predict the total number of conventional Hass avocados sold in the U.S.?

From the perspective of a buyer/consumer,

Can we reasonably predict the average per-unit price of conventional Hass avocados in the U.S.?

Limitations and Future Steps

- Additional predictors
  - e.g. supply volume, import/export, tariff, weather, etc.

- Spatial analysis

Q & A

Model Formulation

Seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$:

$$\Phi_P(L^s)\,\phi_p(L)\,\Delta_s^D\,\Delta^d\,y_t = \delta + \Theta_Q(L^s)\,\theta_q(L)\,w_t$$

where

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$$
$$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^q$$
$$\Delta^d = (1 - L)^d$$

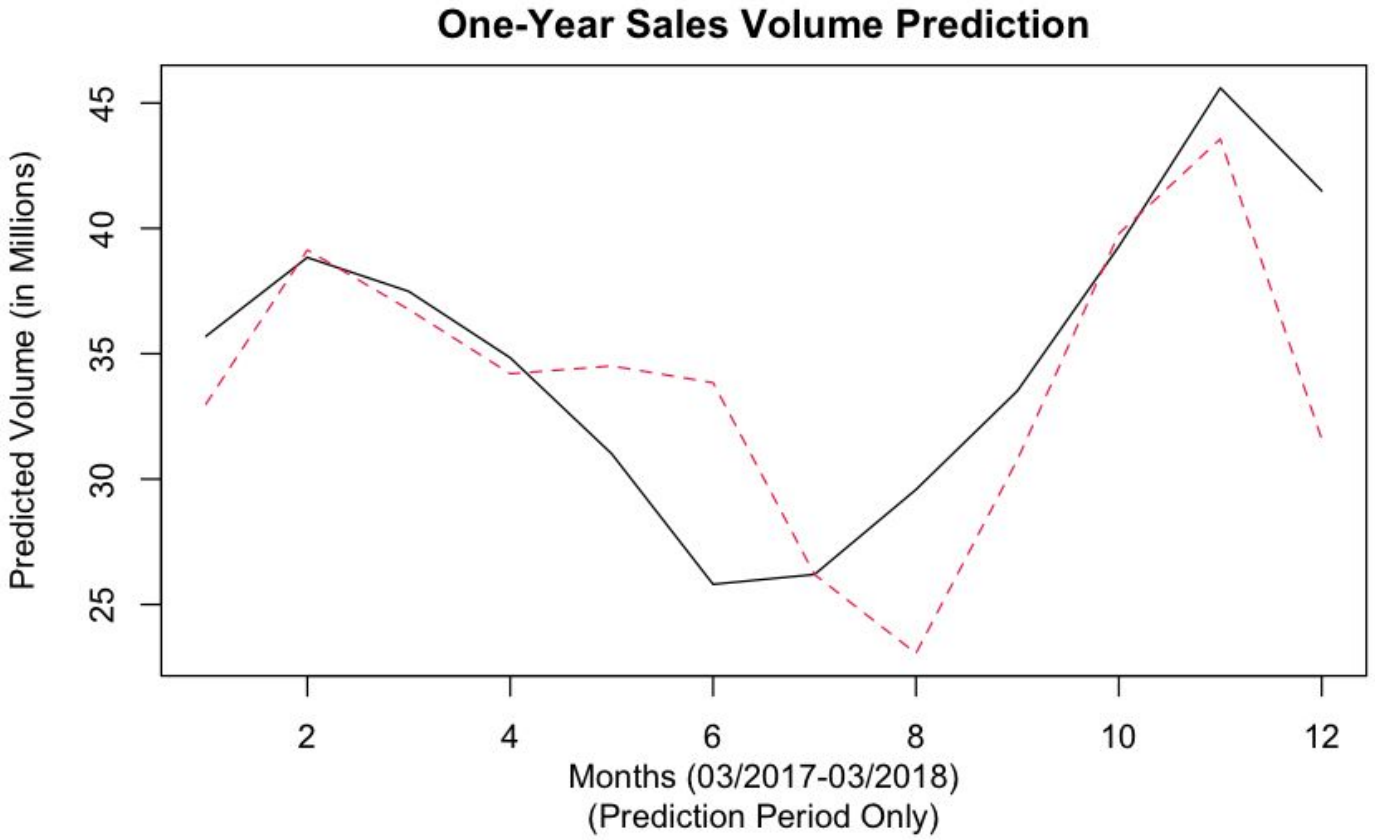
$$\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \ldots - \Phi_P L^{Ps}$$
$$\Theta_Q(L^s) = 1 + \Theta_1 L + \Theta_2 L^{2s} + \ldots + \theta_p L^{Qs}$$
$$\Delta_s^D = (1 - L^s)^D$$

Model Formulation: Volume Prediction

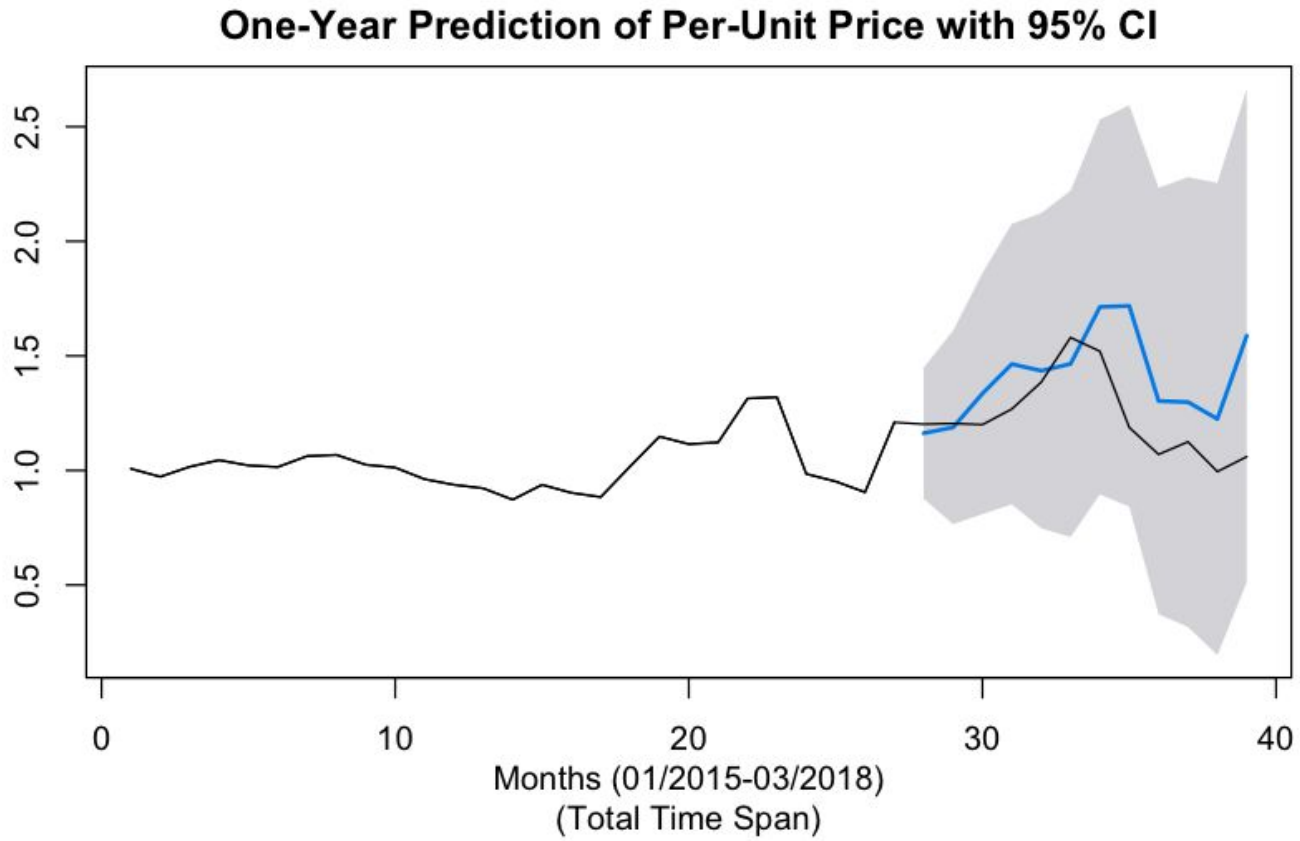
$$\text{Seasonal ARIMA } (2,0,0) \times (0,1,2)_{12}$$

$$(1 - \phi_1 L - \phi_2 L^2)(1 - L^{12}) y_t = \delta + (1 + \Theta_1 L^{12} + \Theta_2 L^{24}) w_t$$

Prediction of Sales Volume: No Exogenous Predictor



Per-Unit Price Prediction with Uncertainty



SARIMA: Full Expression

Seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$:

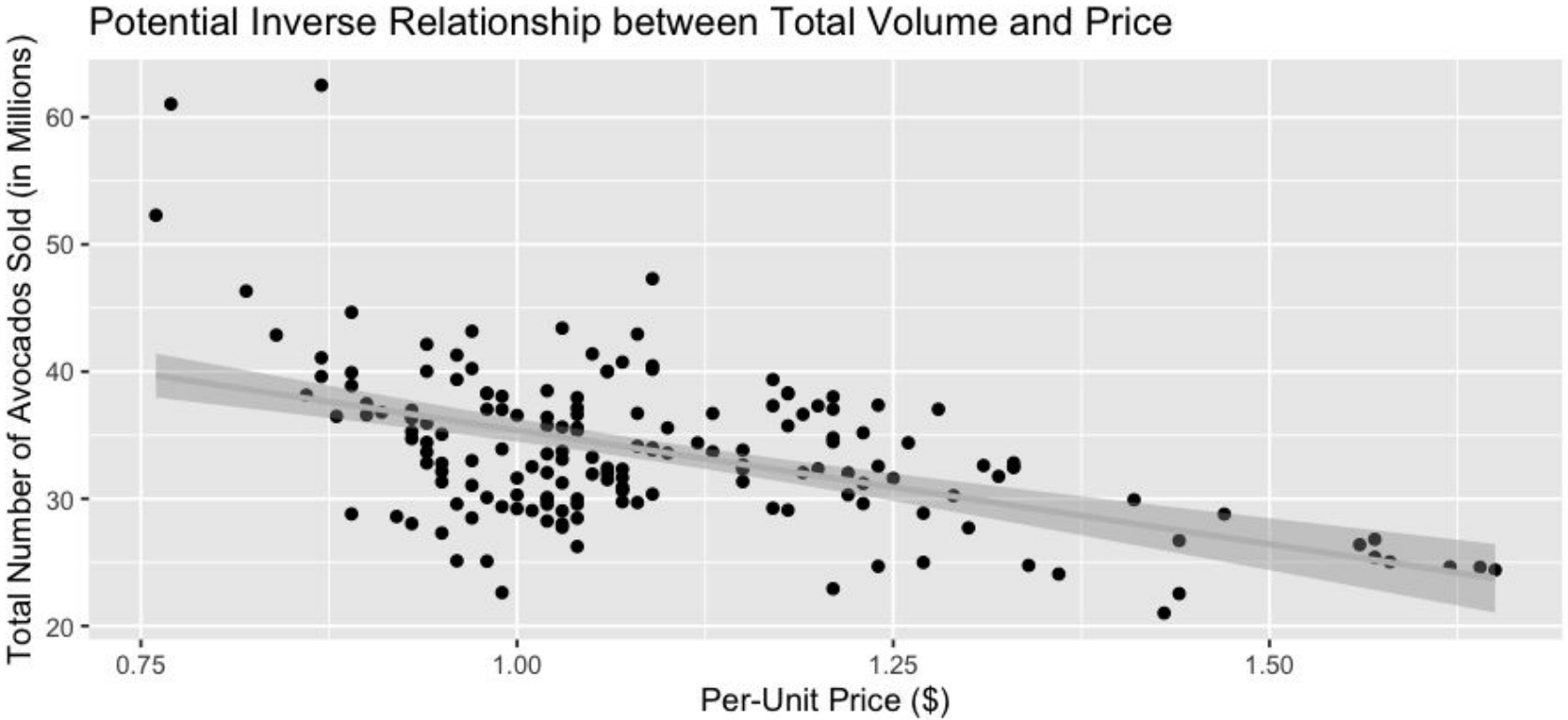
$$\Phi_P(L^s)\,\phi_p(L)\,\Delta_s^D\,\Delta^d\,y_t = \delta + \Theta_Q(L^s)\,\theta_q(L)\,w_t$$

where

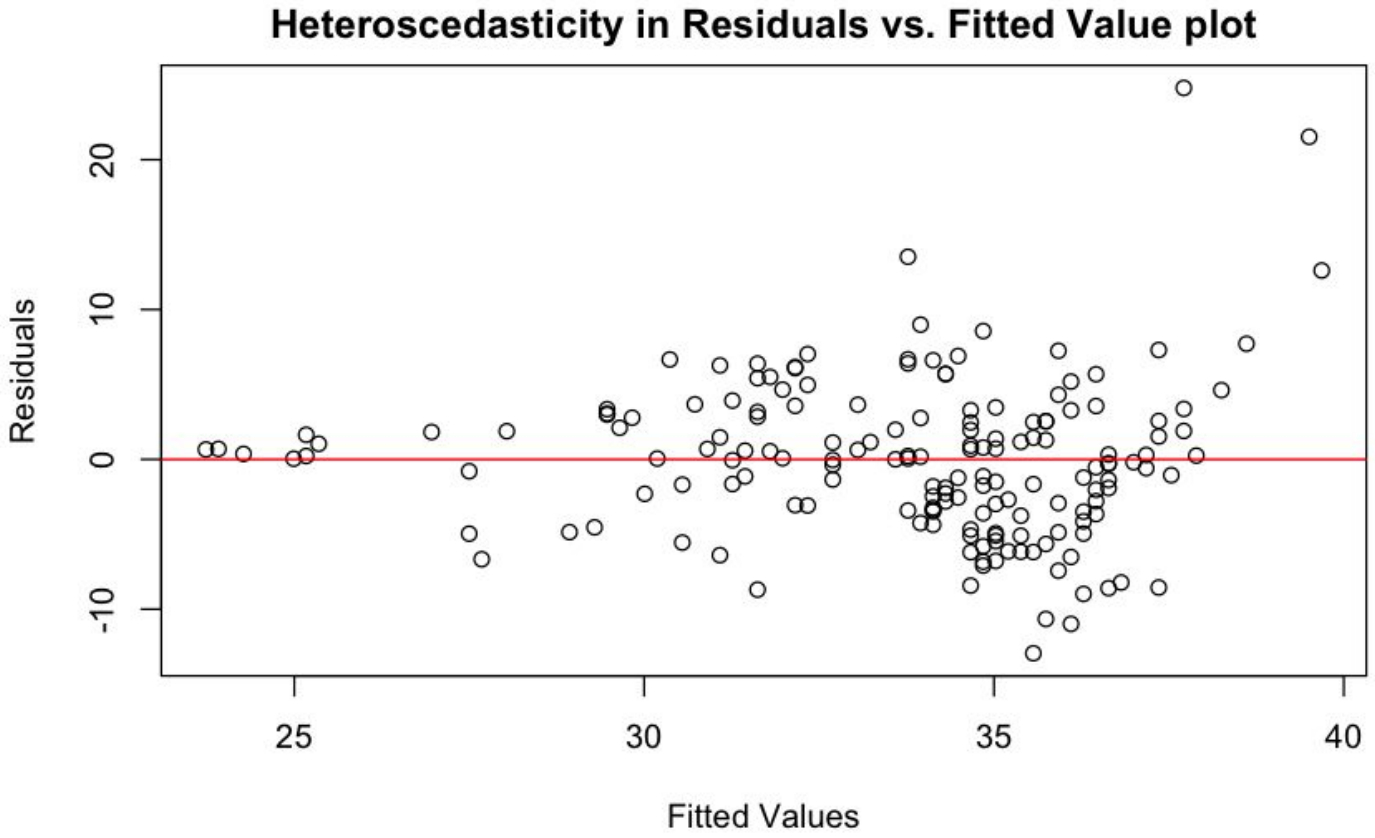
$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$$
$$\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^q$$
$$\Delta^d = (1 - L)^d$$

$$\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \ldots - \Phi_P L^{Ps}$$
$$\Theta_Q(L^s) = 1 + \Theta_1 L + \Theta_2 L^{2s} + \ldots + \theta_p L^{Qs}$$
$$\Delta_s^D = (1 - L^s)^D$$

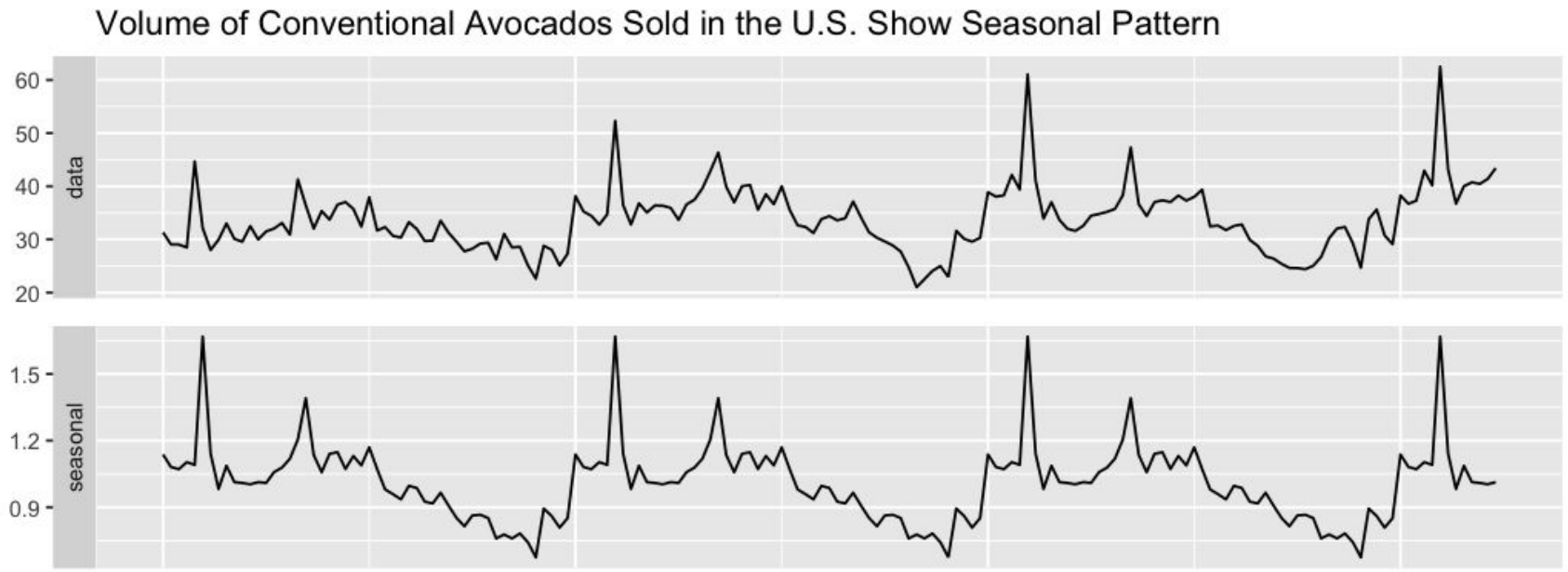
Weekly Data: Volume vs. Price?



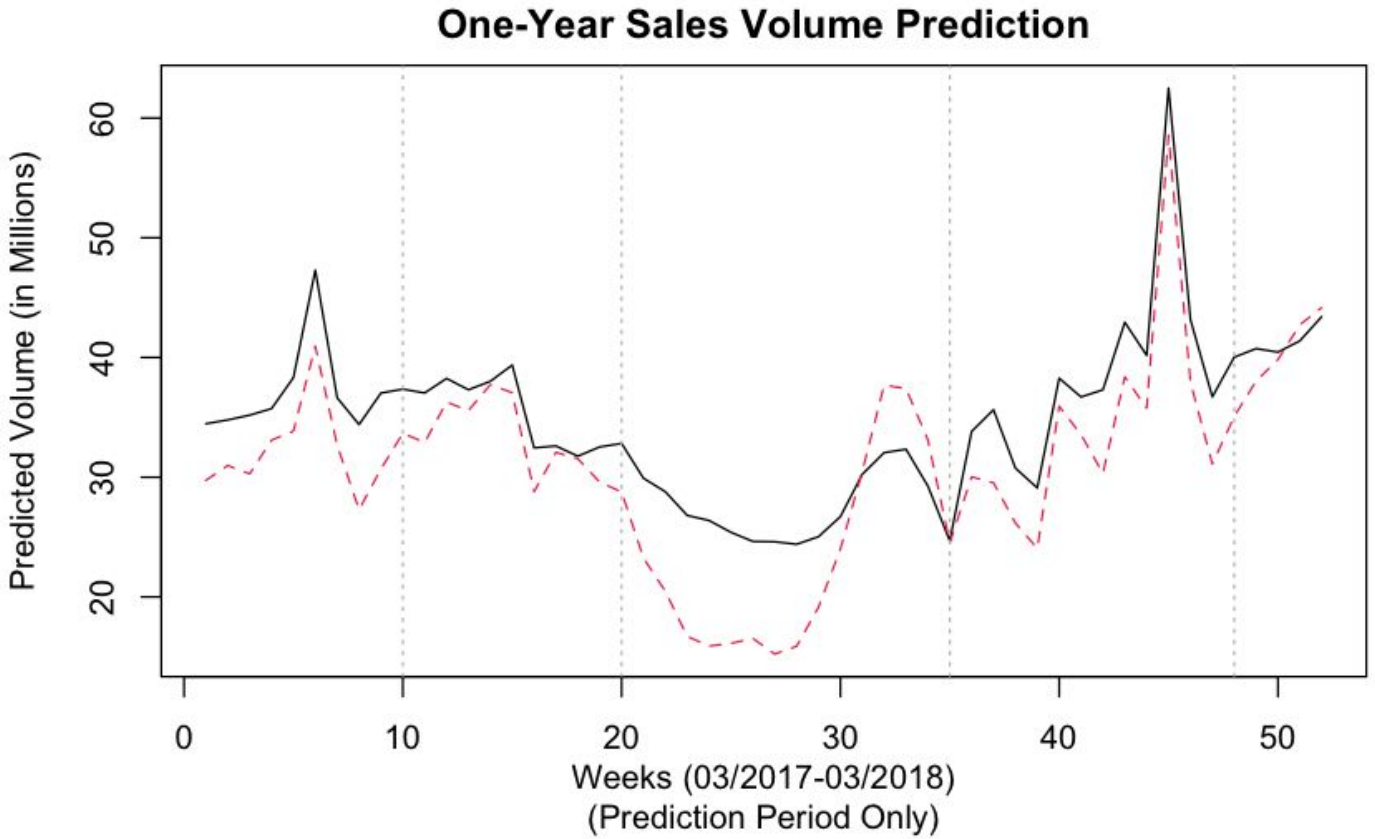
Weekly Data: OLS…Maybe Not?



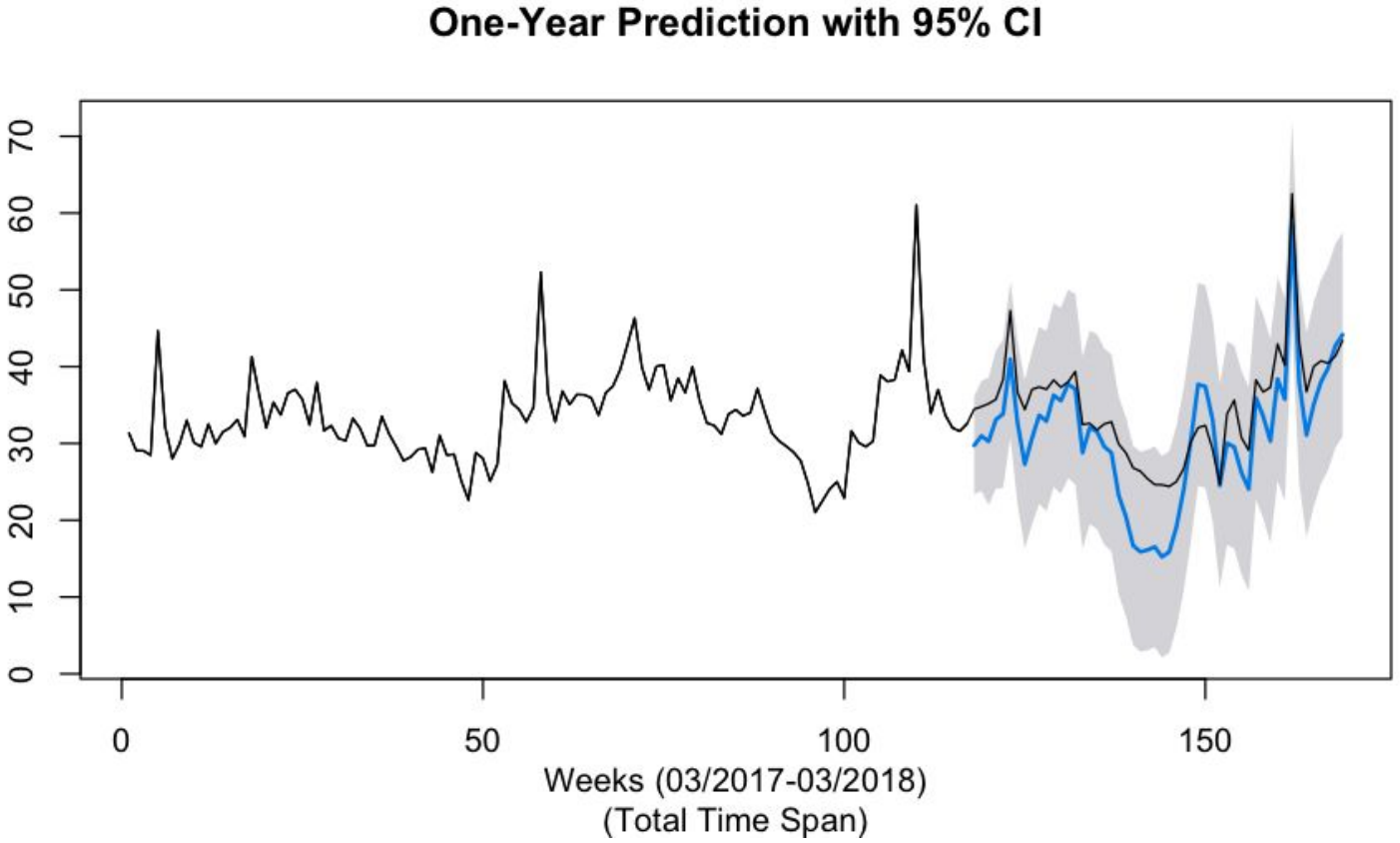
Weekly Data: A Closer Look at Sales Volume



Weekly Data: Results



Weekly Data: Prediction of Sales Volume (With Uncertainty)



Weekly Data: A Closer Look at Per-Unit Prices