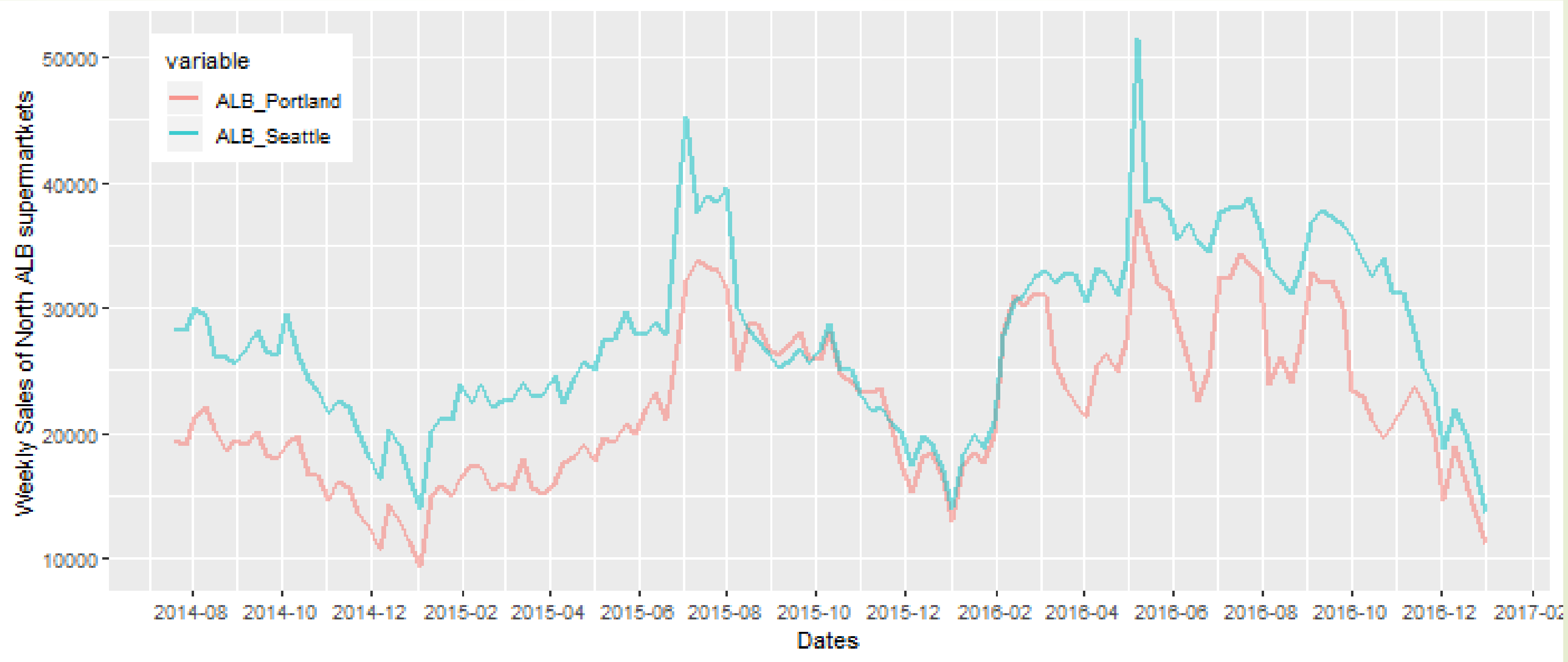
Clif bar sales Forecasting by Applying Time Series Analyses

Pei-Chun Lai
Email: pclaitamu@gmail.com
Department of Statistics, Texas A&M University, College Station, Texas

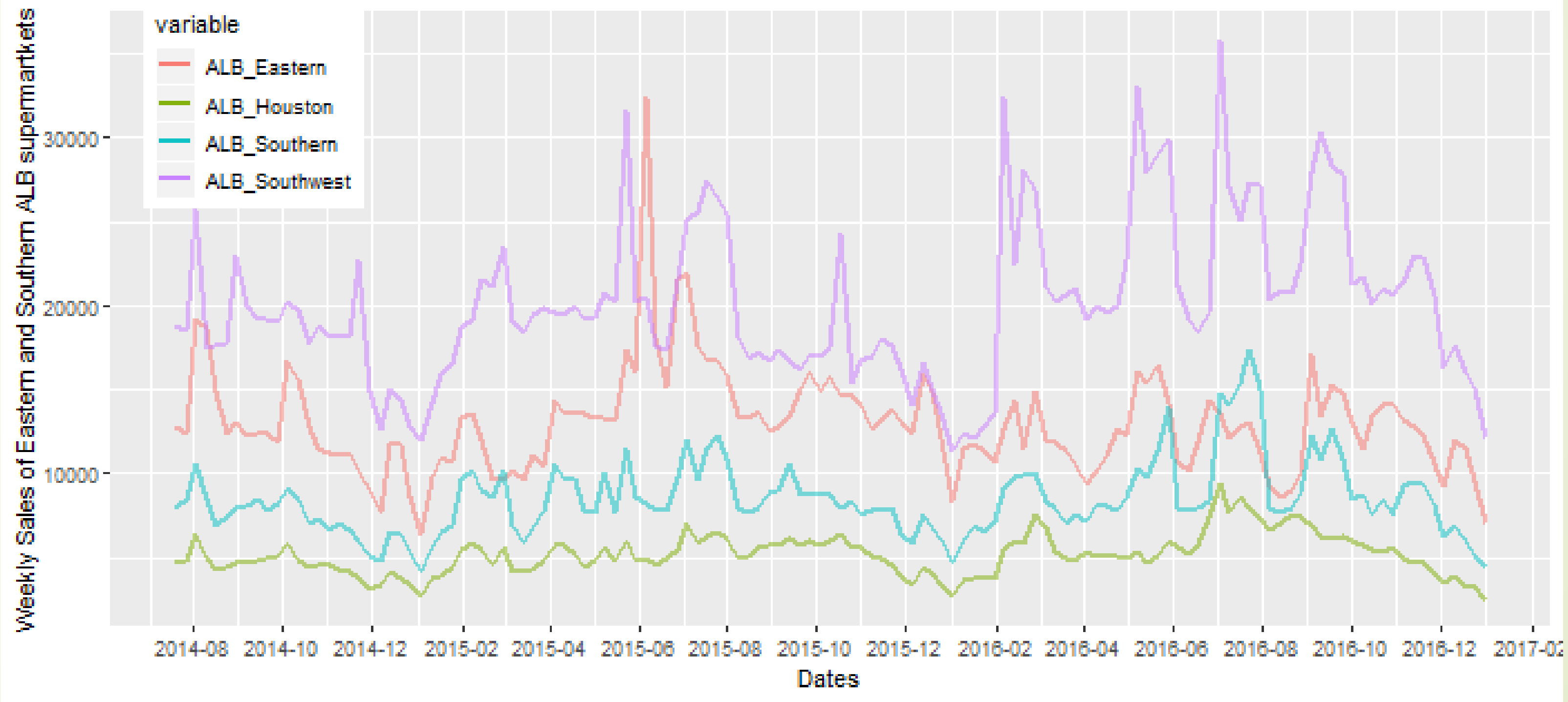
Research Question

- How to get the most accurate prediction of weekly sales of each grocery store?

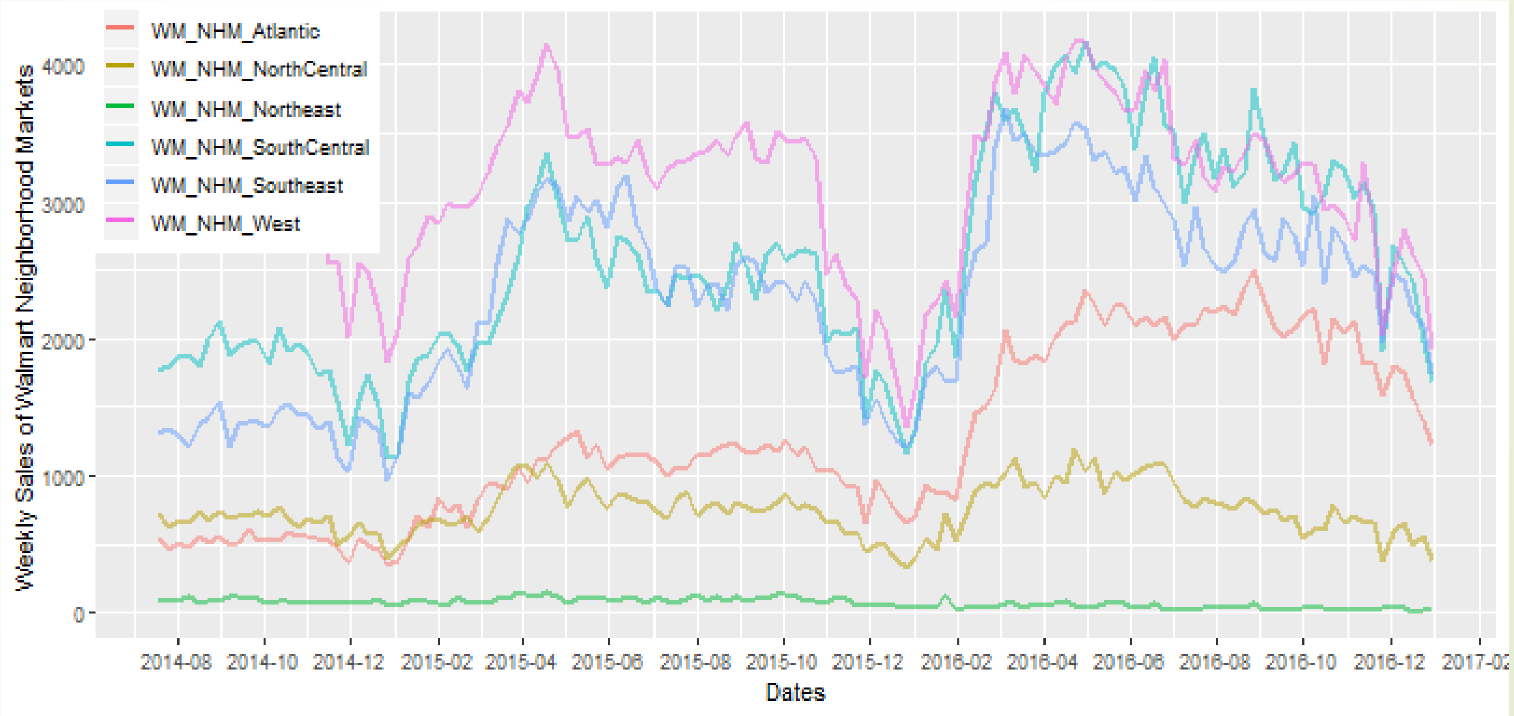
Clif Bar Sales at Albertsons Supermarkets in Portland and Seattle



Clif Bar Sales at Albertsons Supermarkets in Eastern and Southern US

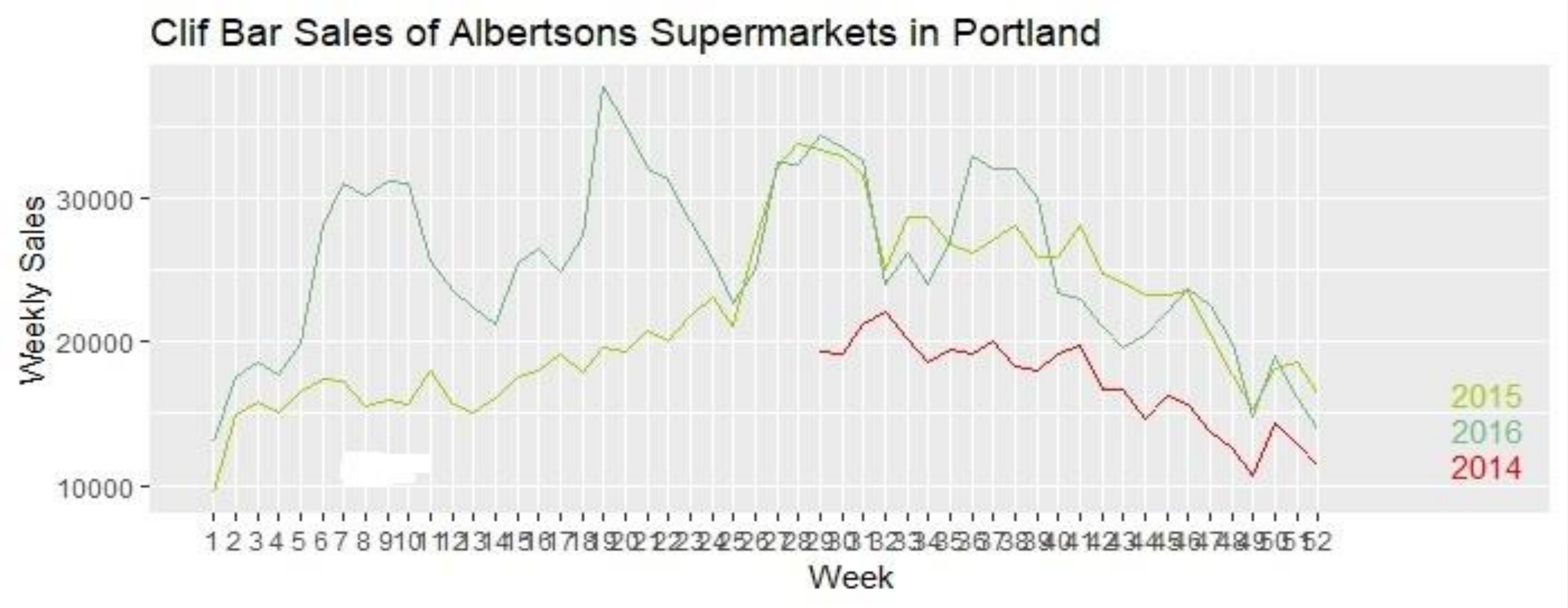


Clif Bar Sales at Walmart Neighborhood Markets in the US



We always can see Clif bar sales suddenly drop in these supermarkets at the end of 2014 and 2015.

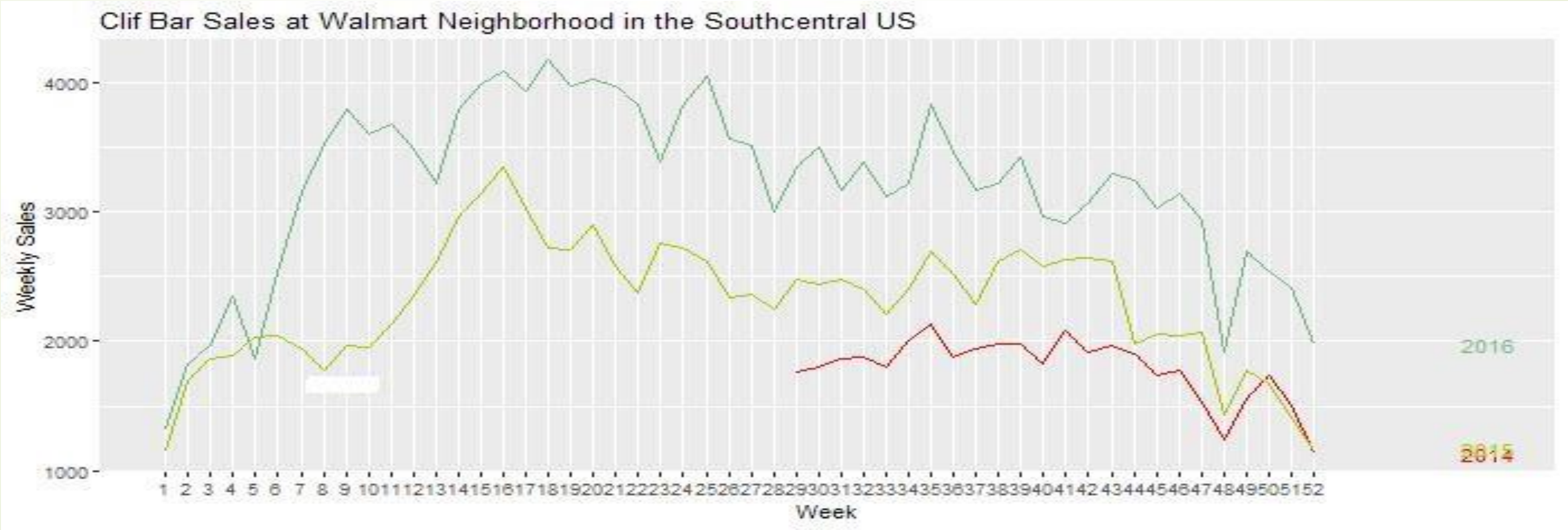
Seasonal plot of weekly Clif bar sales at Albertsons Supermarkets in Portland



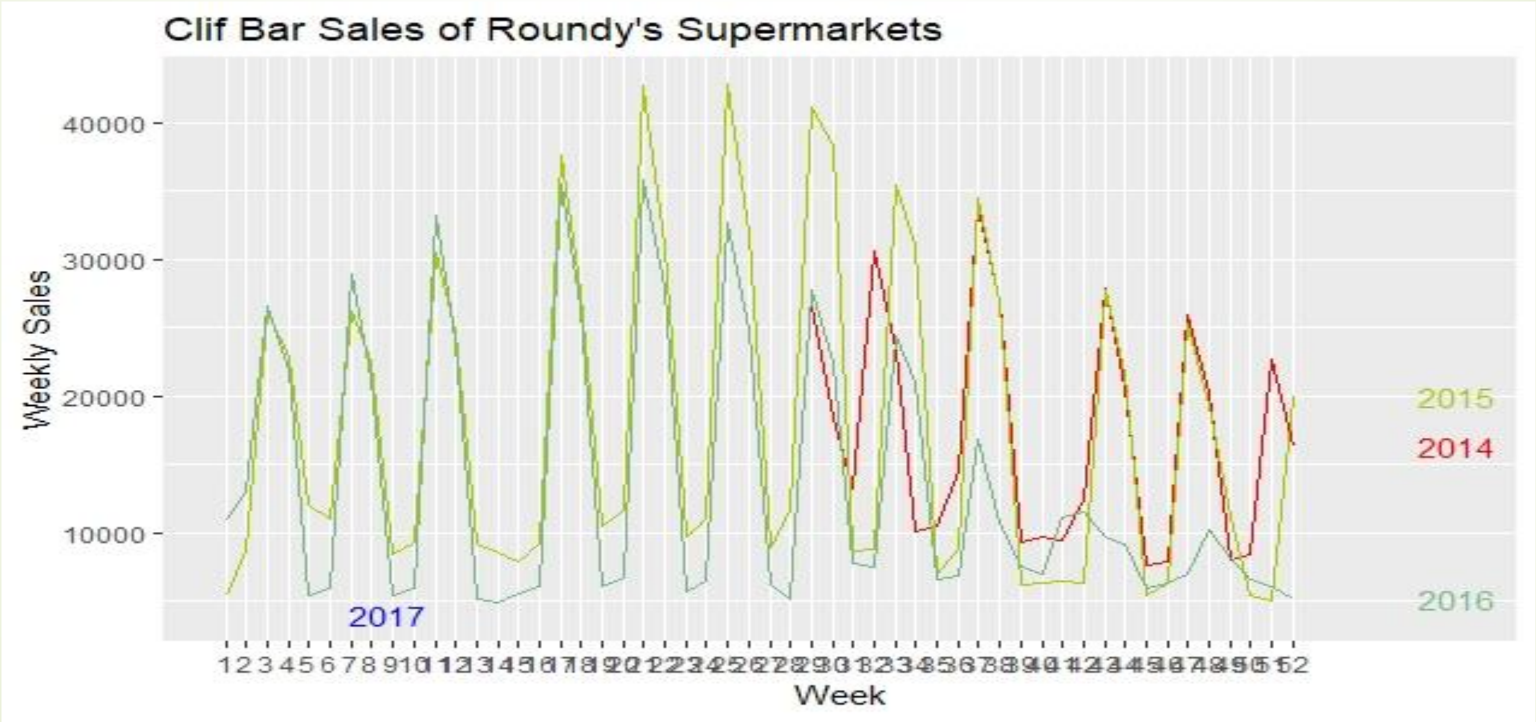
Compared to 2014, Clif Bar has more sales in the middle of the year to the end of the year in 2015 and 2016.

Seasonal Plot of Weekly Clif Bar Sales at Kroger Supermarkets in Dallas

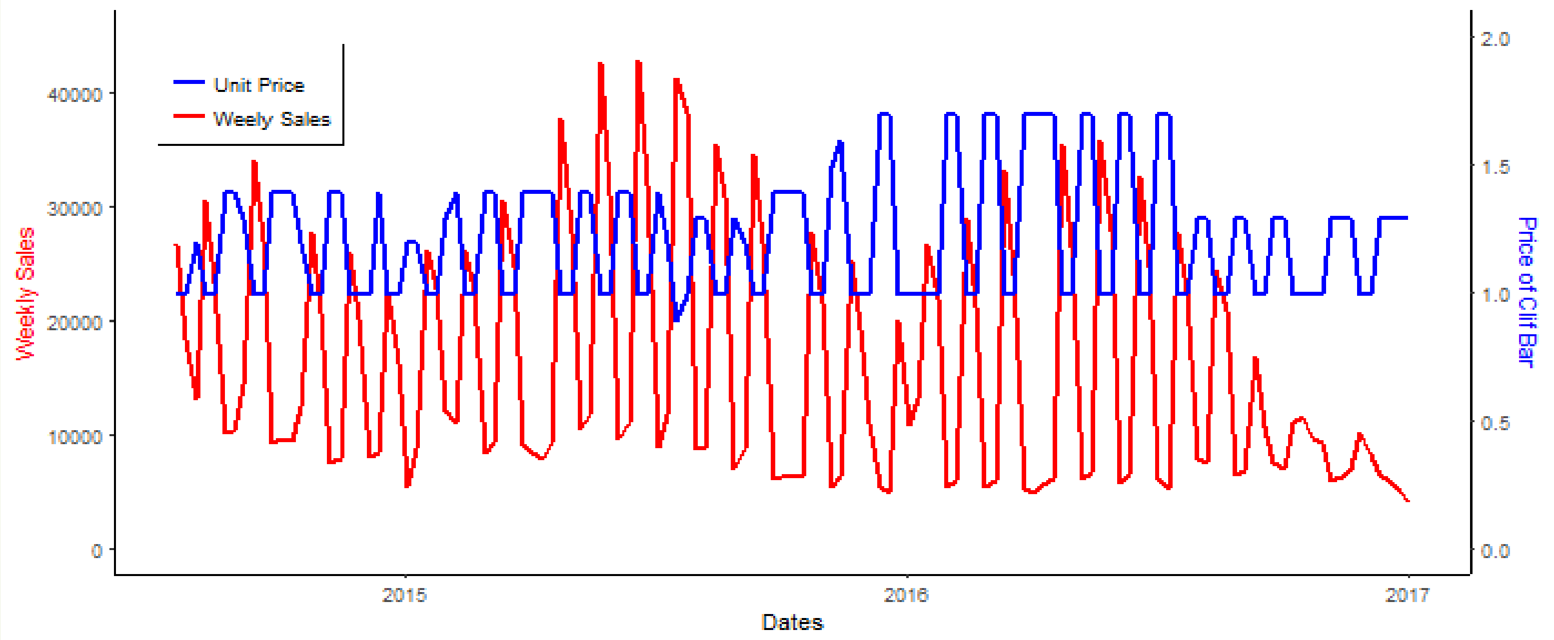
Seasonal Plot of Weekly Clif Bar Sales at Walmart Neighborhoods in the Southcentral US



Seasonal Plot of Weekly Clif Bar Sales at Roundy's Supermarkets



Plots of Weekly Clif Bar Price and Sales at Roundy's Supermarkets



Summaries of the Plots

- We see that sales of Clif Bars have consistently increased over the years at Walmart Neighborhood Markets in the Southcentral US.
- In the Roundy’s Supermarkets, the Pearson correlation coefficient between Clif Bar sales and prices is -0.74. It indicates a strong negatively linear relationship. Also, the Spearman correlation coefficient between Clif Bar sales and prices is -0.82.

Time Series Patterns

- The real-world data are often governed by a (deterministic) trend and may have (deterministic) cyclical or seasonal components in addition to the irregular/remainder (stationary process) component.
- Trend: A trend exists when there is a long-term increase or decrease in the data. Sometimes we refer to a trend as "changing direction."
- Seasonal: A seasonal component exists when a series exhibits regular fluctuations based on the season (e.g., every month/quarter/year). Seasonality is always of a fixed and known period.

Time Series Patterns, con't

- If the frequency is unchanging and associated with some aspect of the calendar, then the pattern is seasonal.

Time Series Decomposition

- Additive Decomposition: $y_t = T_t + S_t + R_t$,

  where y_t is the data, T_t is the trend component at time *t*, which reflects <u>the long-term increase or decrease</u> of the series, S_t is the seasonal component at time *t*, reflecting seasonality, and R_t is noise at time *t*.

- Seasonally adjusted data ($A_t = T_t + R_t$): If the seasonal component is removed from the original data, the resulting values are the "seasonally adjusted" data. For an additive decomposition, the seasonally adjusted data are given by $y_t - S_t$.

Exponential Smoothing

- Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older.
- Simple exponential smoothing is suitable for forecasting data with no clear trend or seasonal pattern.
- Simple exponential smoothing has a flat forecast function:

$$\hat{y}_{T+h} = \hat{y}_{T+1|T} = l_T, \quad \mathrm{h} = 2,3,\ldots$$

All forecasts take the same value, equal to the last level component (l_T).

Simple exponential smoothing

- The simple exponential smoothing could be represented by

  Forecast equation: $\hat{y}_{t+h|t} = l_t$

  Smoothing equation: $l_t = \alpha y_t + (1-\alpha) l_{t-1}$,

  Thus,

$$\hat{y}_{T+h|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0$$

- If we know the value of α, smoothing parameters, and l_0 , the initial component, we can compute all forecasts.
- $0 \leq \alpha \leq 1$, a smaller value of α would lead to smaller change over time, and so the series of fitted values would be smoother.

Simple exponential smoothing, con’t

- The level equation gives an estimate of the local mean, or "level" of the data-generating process (DGP), at this time.

Holt's Linear Trend Method

- This trend method involves these three equations:

  Forecast equation: $\hat{y}_{t+h|t} = l_t + hb_t$

  Level equation: $l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$

  Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1},$

  where β^* is the smoothing parameter for the trend, $0 \leq \beta^* \leq 1.$ The h-step-ahead forecast is equal to the last estimate level plus *h* times the last estimated trend value. In other words, the forecasts are a linear function of *h* and b_t denotes the slope at time t.

- Disadvantage: The forecasts display a constant trend indefinitely into the future which causes over-forecast, especially for longer forecast horizons.

Damped Trend Methods

- Solution: Gardner & McKenzie introduced a parameter that dampens the trend to a flat line some time in the future.
- The three equations including a damping parameter $0 < \varphi < 1$ are:

  Forecast equation: $\hat{y}_{t+h|t} = l_t + (\varphi + \varphi^2 + \cdots + \varphi^h)b_t$

  Level equation: $l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + \varphi b_{t-1})$

  Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\varphi b_{t-1}$.
- For $0 < \varphi < 1$, φ dampens the trend so that it approaches a constant some time in the future.

Holt-Winters' seasonal method

- The additive method involves these four equations:

  Forecast equation: $\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$

  Level equation: $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$

  Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$

  Seasonality equation: $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 -$

Holt-Winters' seasonal method, con't

- The level equation shows a weighted average between the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(l_{t-1} +$

Classification of exponential smoothing methods

Trend Component	Seasonal Component		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

Short hand	Method
(N,N)	Simple exponential smoothing
(A,N)	Holt's linear method
(A_d,N)	Additive damped trend method
(A,A)	Additive Holt-Winters' method
(A,M)	Multiplicative Holt-Winters' method
(A_d,M)	Holt-Winters' damped method

Classification of exponential smoothing methods, con't

- Formulas for recursive calculations and point forecasts.

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
$\mathbf{A_d}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$

Exponential Smoothing State Space Model (ETS)

- For each method there exist two models: one with additive errors and one with multiplicative errors. The state space model is labelled as ETS(, , ,) for (Error, Trend, Seasonal).
- ETS(A,N,N) represents simple exponential smoothing with additive errors.
- In simple exponential smoothing,

  $l_t = \alpha y_t + (1-\alpha) l_{t-1} = l_{t-1} + \alpha(y_t - l_{t-1}) = l_{t-1} + \alpha e_t,$

  where $e_t = y_t - l_{t-1} = y_t - \hat{y}_{t|t-1}$ is the residual at time *t*.
- We can also write $y_t = l_{t-1} + e_t$.

Simple Exponential Smoothing with Additive Errors

- For a model with additive errors, we assume that e_t are normally distributed white noise with mean 0 and variance σ^2, e.g., $e_t = \varepsilon_t \sim \mathrm{NID}(0,\ \sigma^2)$.
- Measurement equation: $y_t = l_{t-1} + \varepsilon_t$

  State equation: $l_t = l_{t-1} + \alpha\varepsilon_t$
- l_{t-1} is the predictable part of y_t, and ε_t is the unpredictable part of y_t.
- High values of α allow rapid changes in the level; low values of α lead the smooth changes.

Simple Exponential Smoothing with Multiplicative Errors

- The multiplicative error is defined as

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}},$$

where $\varepsilon_t \sim \mathrm{NID}(0,\ \sigma^2)$.

- The multiplicative form of the state space models are

  Measurement equation: $y_t = l_{t-1}(1 + \varepsilon_t)$

  State equation: $l_t = l_{t-1}(1 + \alpha\varepsilon_t)$

STLF() Function with ETS method in R

- In my settings, ETS models with additive trends are allowed.
- Apply each model that is appropriate to the data.
- Optimize parameters and initial values using maximum likelihood estimators (or some other criterion).
- Select best method using corrected Akaike's Information Criterion (AICc). AICc is defined as

$$AICc = AIC + \frac{2(k+2)(k+3)}{T-k-3},$$

where $AIC = -2\log(L) + 2k$, T is the number of observations used for estimation, L is the likelihood of the model, and k is the total number of parameters and initial states $(l_0, b_0, s_0, \ldots)$ that have been estimated.

STLF() Function with ETS method in R, con't

- The appropriate model is selected automatically using the information criterion.
- Produce forecasts by using best method. Use the underlying state space model to obtain the forecast intervals.

STLF command

- Decomposes time series into a trend, seasonal, and noise component using STL decomposition.
- STL is an acronym for "Seasonal and Trend decomposition <u>using Loess</u>", while Loess is a method for estimating nonlinear relationships.
- To forecast a decomposed time series, we forecast the seasonal component, $\widehat{S}_t$, and the seasonally adjusted components $\widehat{A_t} = \widehat{T}_t + \widehat{R_t}$, <u>separately</u>.
- Use ETS or ARIMA to forecast the seasonally adjusted series.

STLF command, con’t

- Combine the forecasted seasonal component and the forecasted seasonally adjusted components to get forecasts for original series.
- s.window (seasonal window): T<u>he number of consecutive years</u> to be used in estimating each value in the seasonal component. If we think the seasonal pattern is identical across years, we should set this parameter to a big value or periodic. If the seasonal pattern evolves quickly, we should set this parameter to smaller values so that the analysis is not affected by a previous seasonal pattern.

STLF command, con't

- t.window (trend-cycle window): The number of consecutive observations to be used when estimating the trend-cycle. In most cases, t.window is chosen automatically.

Exponential Smoothing vs ARIMA

- While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

Singular Value Decomposition

- A 'singular value decomposition' (SVD) of M has the form: $M = U\sum V^*$, where V^* is the conjugate transpose of V. The columns of U and the columns of V are called the left-singular vectors and right-singular vectors of M, respectively.
- SVD is a decomposition method of singular spectrum analysis.

ARIMA Models

- The order of the ARIMA model is selected by minimizing the BIC, although that is done within the auto.arima() function.

How does auto.arima() work?

- Non-seasonal version:
- Step1: Select current model (with smallest AICc) from: ARIMA(2, d, 2)
  - ARIMA(0, d, 0)
  - ARIMA(1, d, 0)
  - ARIMA(0, d, 1)
- Step 2: Consider variations of current model:
  - Vary one of p, q, from current model by ±1;
  - p, q both vary from current model by ±1;
  - Include/exclude c from current model.
- The model with the lowest AICc becomes the current model. Repeat Step 2 until no lower AICc can be found.

References

- 1. Hyndman, Rob J., and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.