

A top-down photograph of a Mexican meal. In the center, there's a stack of three soft tortillas. To the left, a white bowl contains yellow corn kernels. Above the tortillas is a small wooden bowl filled with red salsa. To the right, there are several tacos filled with meat, cheese, and vegetables like corn and onions. Scattered around the meal are fresh ingredients: a whole ear of corn on the cob, a bunch of cherry tomatoes, a lime, a red chili pepper, and a bunch of cilantro. A yellow bell pepper is partially visible in the bottom left corner.

# Tortilla prices in Mexico

---

A BUSINESS, ECONOMICS AND FINANCIAL DATA PROJECT

By Giorgia Barzan, Kaan Daibasoglu, Giacomo Schiavo

# Goal of the project

Analyze and interpret the previous market prices with relative factors to accurately predict the prices of tortillas in Mexico.

But why **TORTILLAS?**



# Why a dataset about Tortillas?

- It is a very *well-established* product, so the demand should be **stable**.
- Easy to produce with basic ingredients (water, flour, oil and salt only).
- Thanks to the basic ingredients we can see the **effects of side expenses** like heating, wages and transportation.

# Contents

- Data collection
- Exploratory analysis
- Data preprocessing
- Data modeling



# Data collection

## Minimum Wages



until 2022



for 2023



for 2024

## Main Dataset with Prices

- 270k+ records of tortilla prices from Mexico's national System of Information and Market Integration:
  - 53 cities
  - 384 mom-and-pop stores
  - 120 retail stores
- Price: Mexican pesos/kilogram of tortilla

## Ingredients

Commodity Price Data  
(The Pink Sheet)

Monthly prices  
from 1960 to present



THE WORLD BANK



# EXPLORATORY ANALYSIS

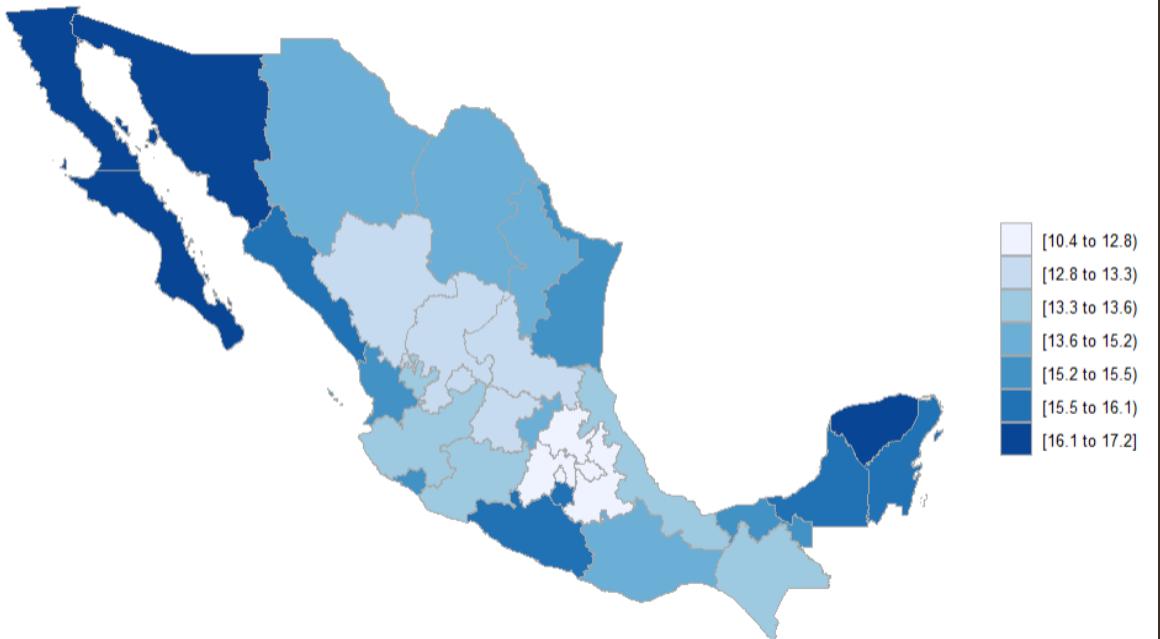
---



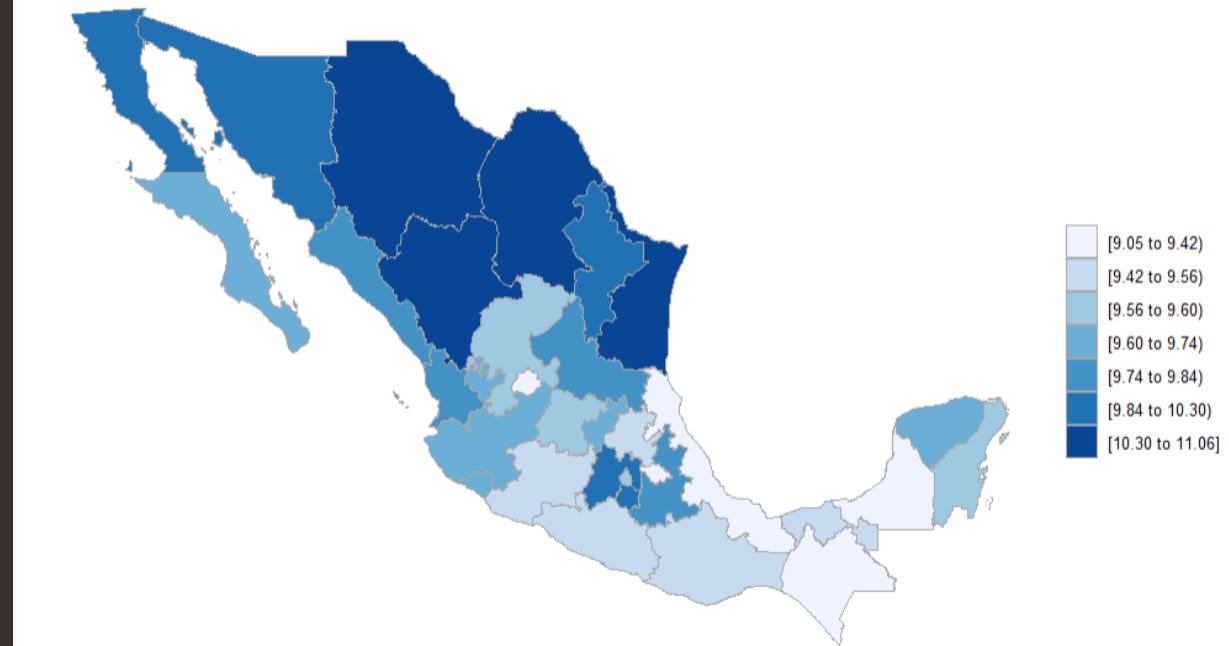
# Exploratory analysis

## Average Prices per State

Average Mom and Pop Store Prices, by state



Average Big Retail Prices, by state



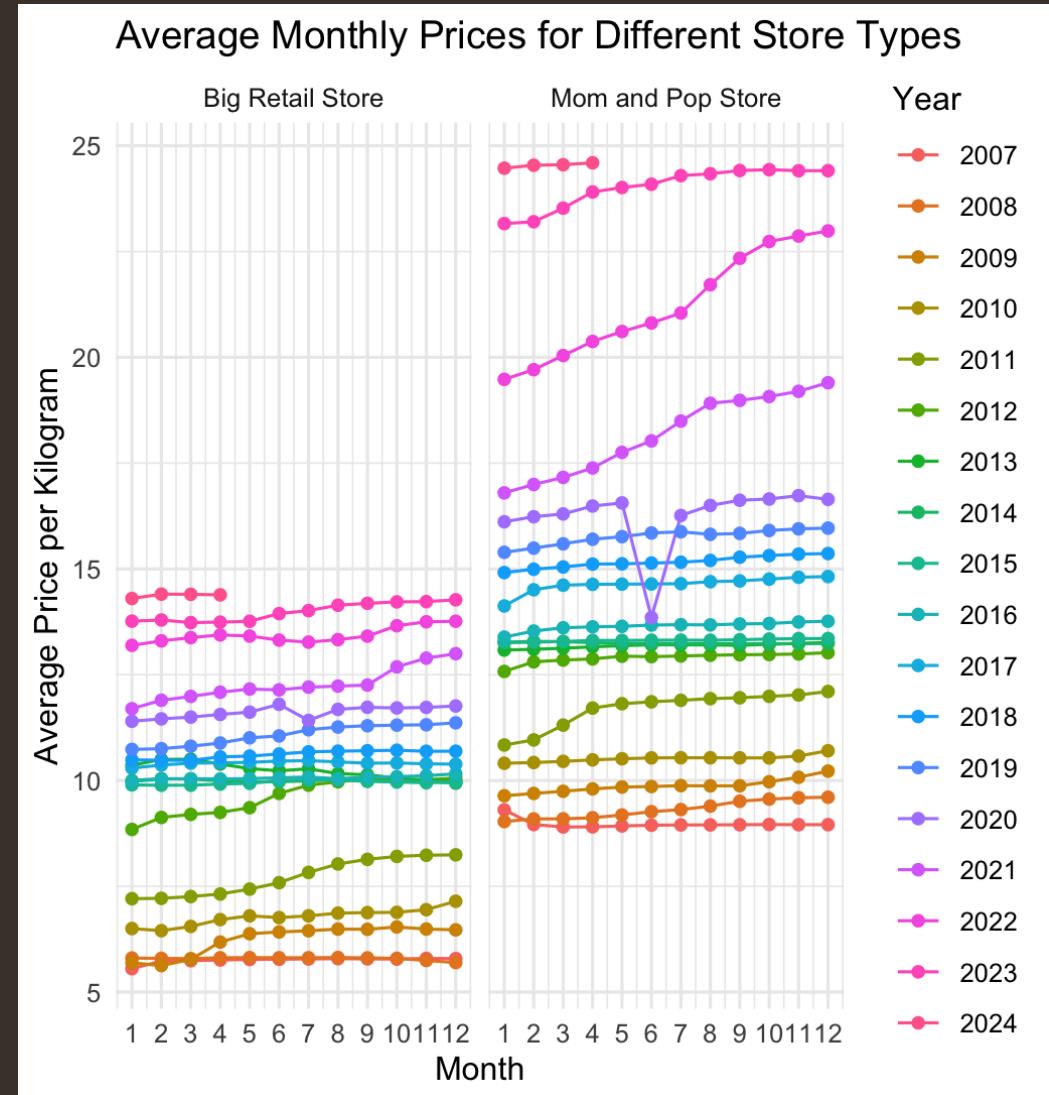


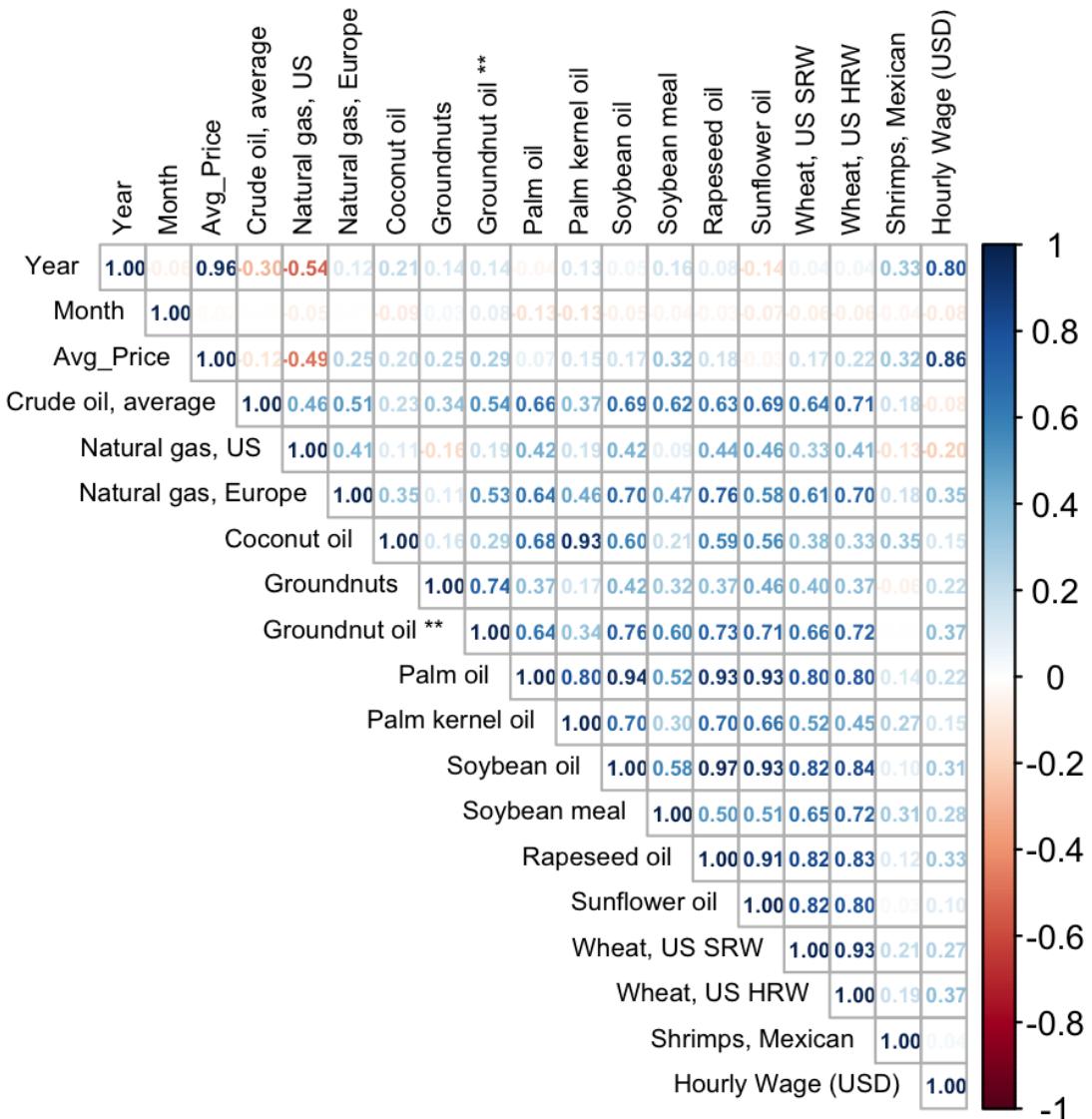
# Exploratory analysis

- Data with **upgoing trend**.
- No visible seasonality effect.
- One jump point in data (might be due to Covid-19).

# Different Store Types Analysis

- Big Retail has lower prices.
  - Drop in June 2020 on Mom and Pop Stores.
  - Rapid growth in Mom and Pop Store prices.



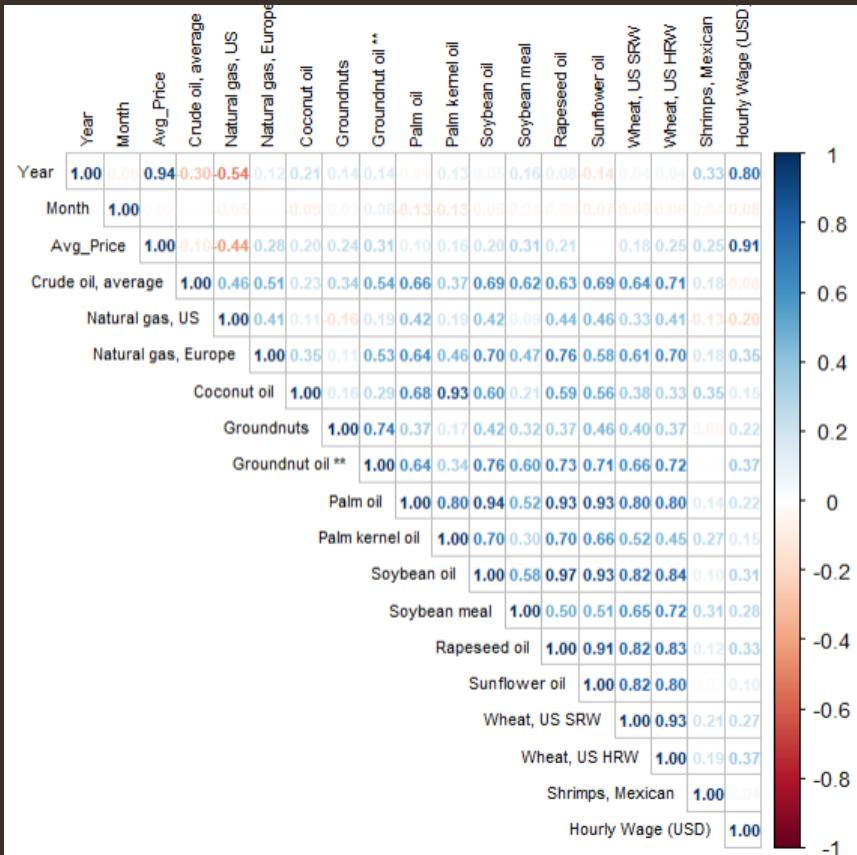


# Correlation Matrix

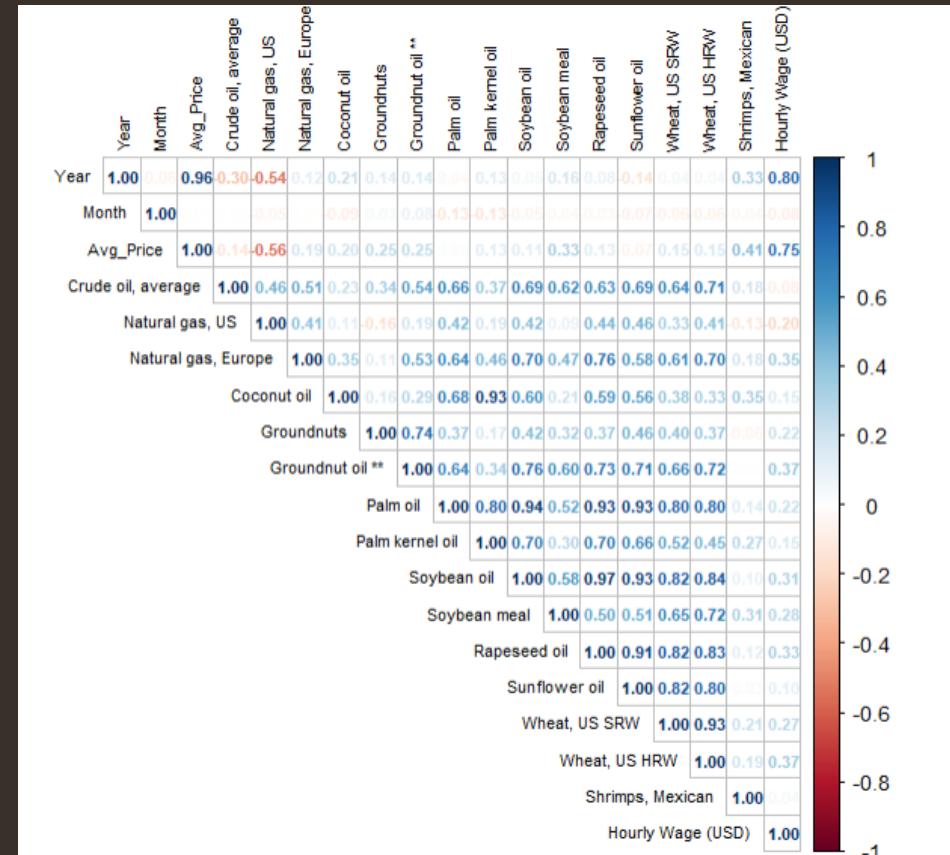
- Some relevant expenses.
- Expenses highly correlated.

# Correlation Matrix for different store types

Mom and Pop Stores



Big Retail Stores





## Possible reasons behind the relations

- *every oil* with the production of tortillas
- *natural gas* with the heat system used in baking tortillas and transportation costs
- *wheat* with the production of tortillas
- *hourly wage* with the personnel costs of the brands and stores
- *soybean meal* with the production of tortillas



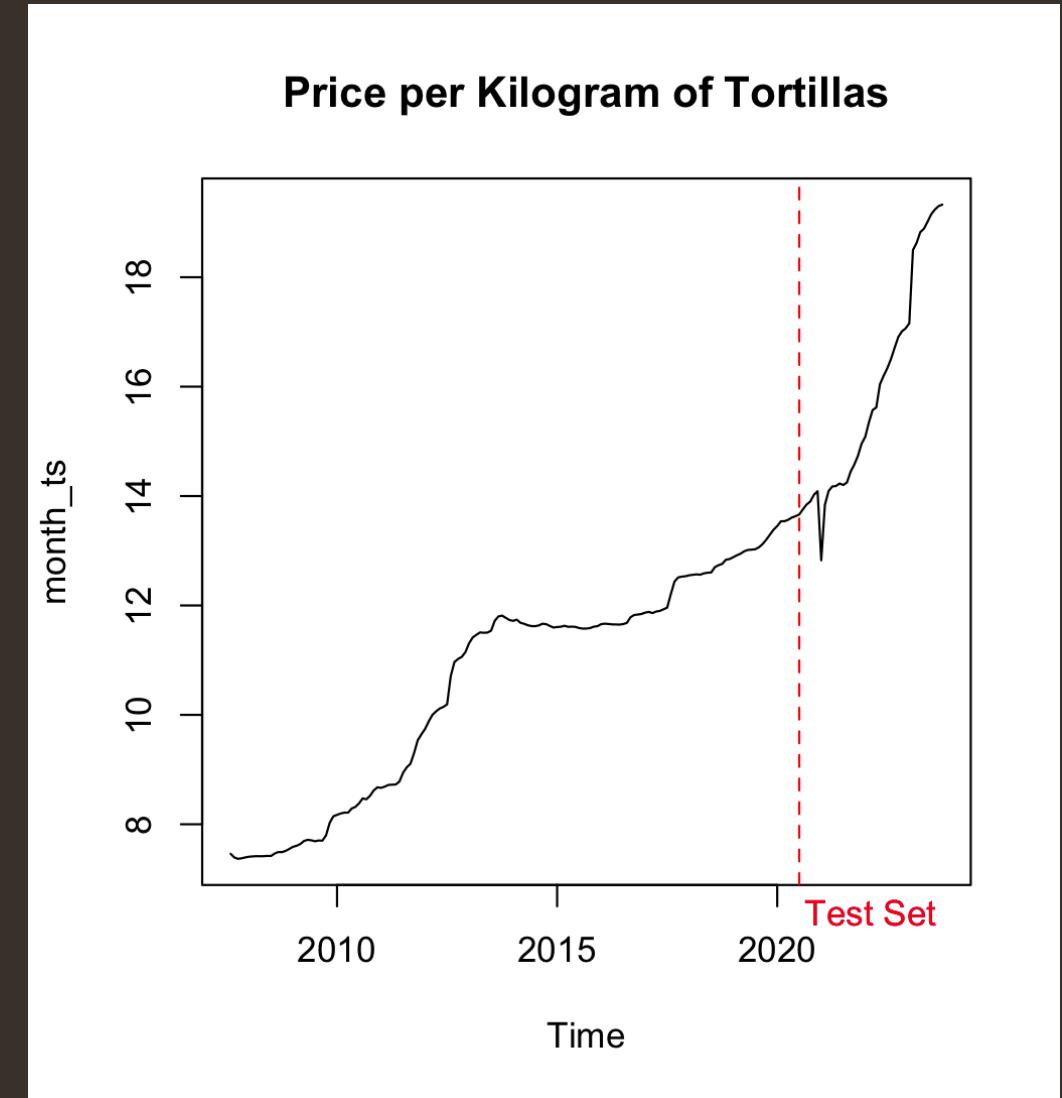
# Data Preprocessing

- All the rows with missing data (NA) are **removed**.
- **Wheat US/SRW** column converted to numerical.
- Different datasets are merged.
- For the year 2023, the daily minimum wage is converted to monthly equivalent.
- **Monthly average** is chosen over yearly and daily for ease of use.
- The expenses not correlated with average price are **removed** .
- **Train test split method** is chosen for more reliable evaluation.

# Train - test split

Train size: 80% - from 01/2007 to 12/2019

Test size: 20% - from 01/2020 to 10/2023



# Data Modeling

1. **TSLM**: to model *linear relationships* between monthly prices and expenses variables.
2. **ARIMA**: to capture non-stationarity, autocorrelation and moving averages in time series data.
3. **Holt**: exponential smoothing model to capture trends and seasonal patterns in time series data.
4. **Splines**: to capture non-parametric relationships of data.
5. **GAM**: to capture non-linear relationships between monthly prices and expenses variables.

# Time Series Linear Model (TSLM)

Fit a *linear regression* model on time series data, incorporating expenses variables as predictors.

---



# Multicollinearity Removal

VIF analysis of relevant expenses to remove variables with high multicollinearity.

Variables highly correlated:

*Sunflower oil*

*Rapeseed oil*

*Wheat, US HRW*

# TSLM with correlated expenses and trend

- Variables included:
- Natural gas (US)
  - Soybean meal
  - Wheat, US SRW
  - Trend

**Price per Kilogram of Tortillas**

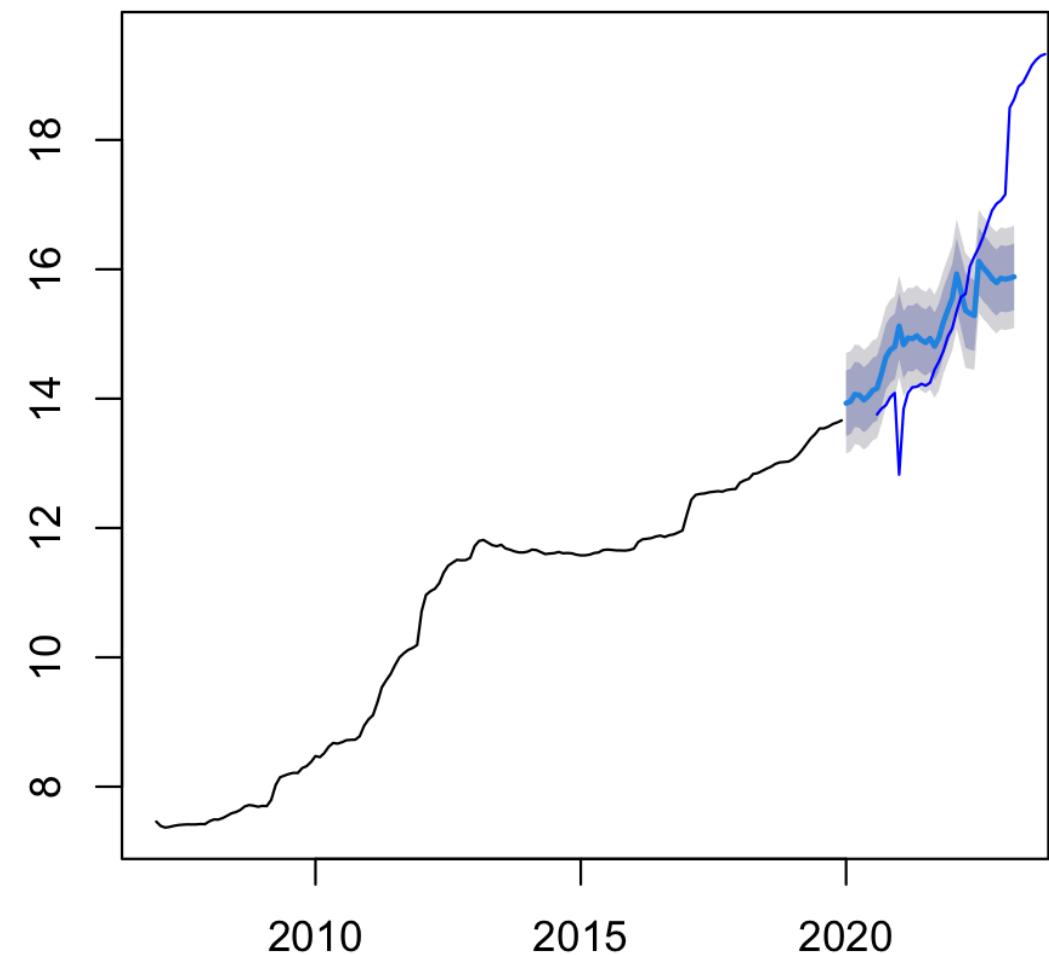


Forecast of TSLM with correlated expenses and trend

RMSE = 1.654

MPE = 4.791%

## Forecast of the Price per Kilogram of Tortillas



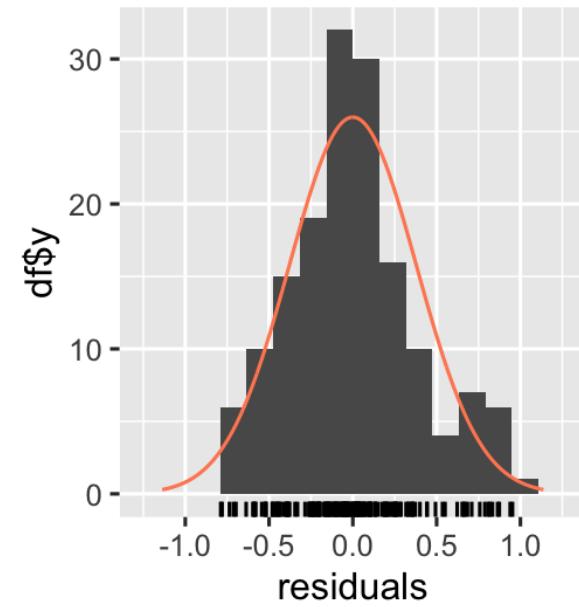
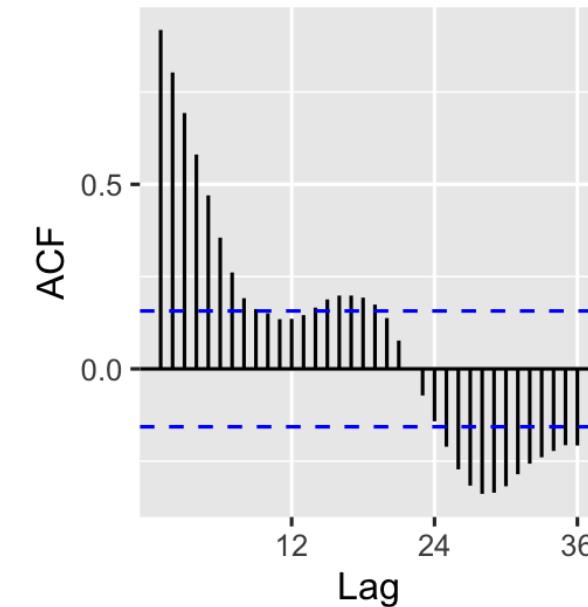
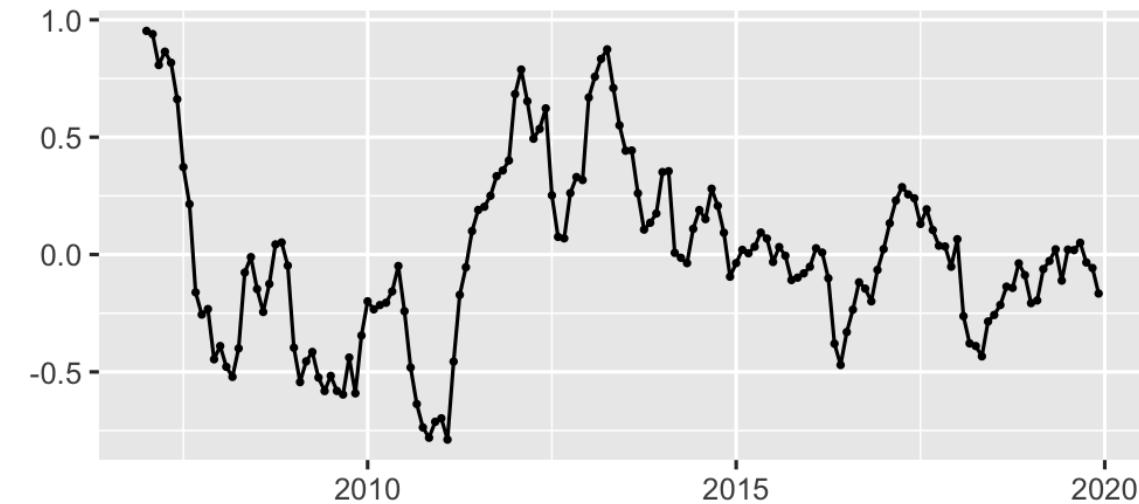
# Residuals from TSLM

AUTOCORRELATED  
RESIDUALS

LM TEST = 136.96, DF = 24,  
P-VALUE < 2.2E-16

DW = 0.1213523

Residuals from Linear regression model



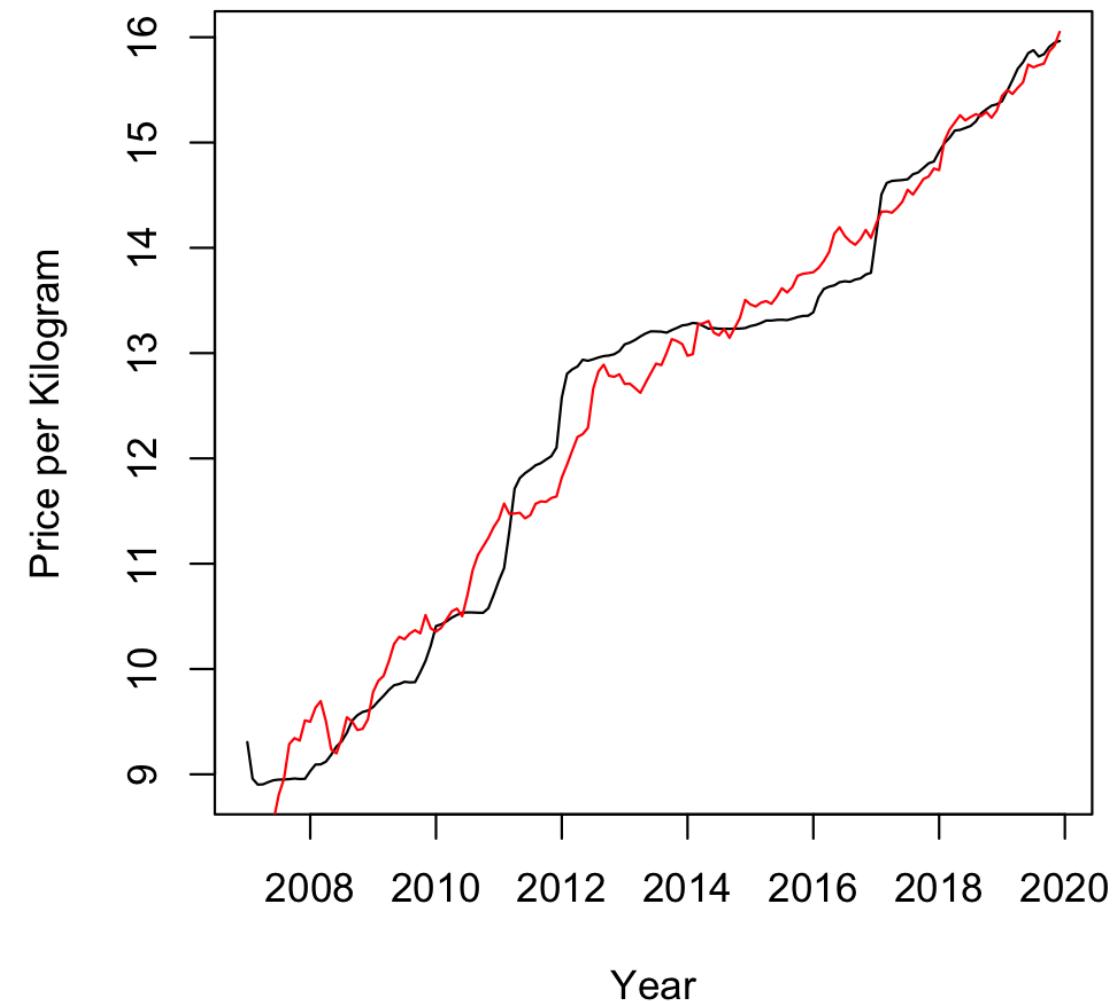
## MOM AND POP STORES

TSLM with correlated expenses and trend

Variables included:

- Natural gas (US)
- Soybean meal
- Wheat, US SRW
- Trend

**Price per Kilogram of Tortillas  
in Mom and Pop Stores**



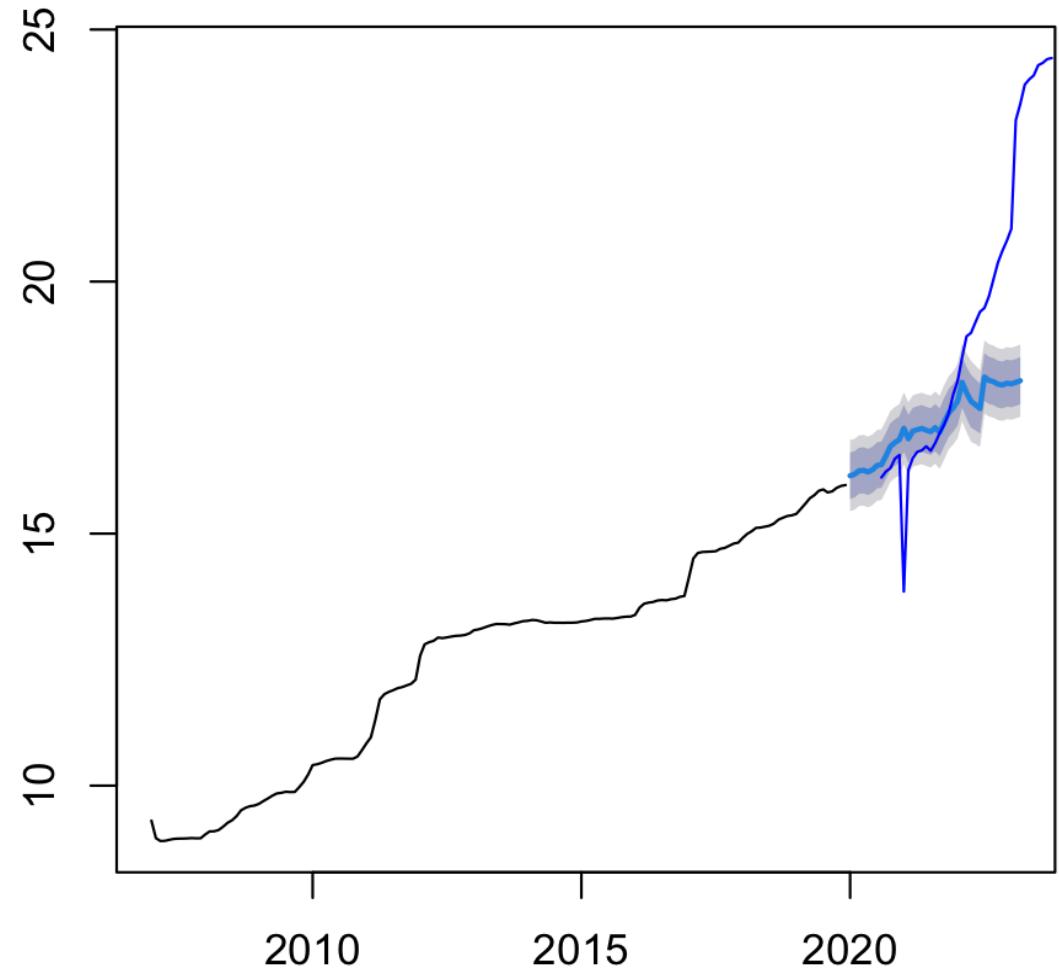
## MOM AND POP STORES

Forecast of TSLM with correlated expenses and trend

RMSE = 3.217

MPE = 9.23%

**Forecast of the Price per Kilogram of Tortillas for Mom and Pop store**



MOM AND POP STORES

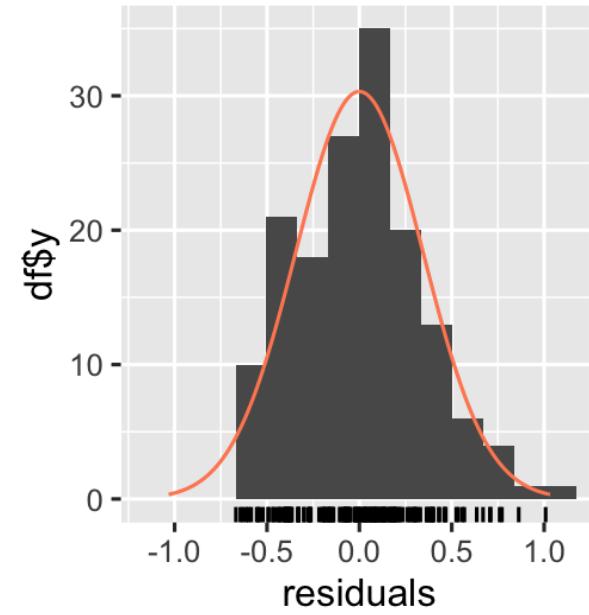
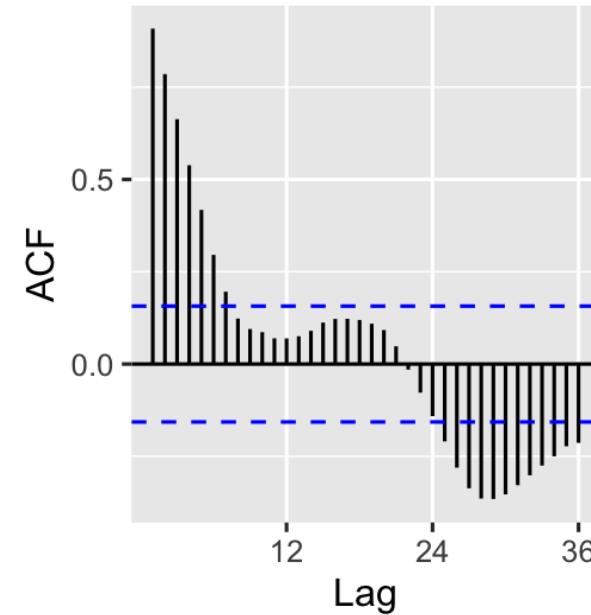
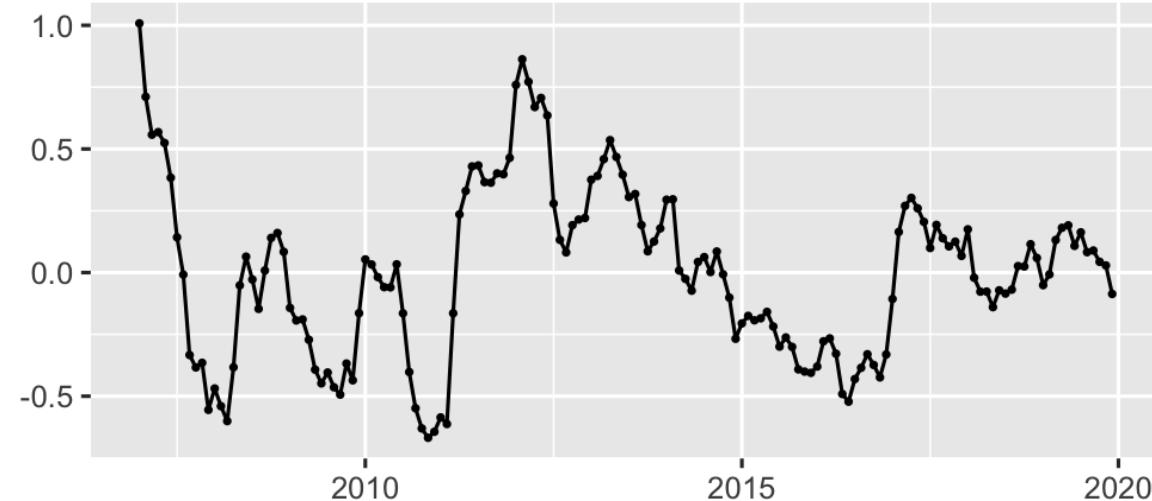
## Residuals from TSLM

AUTOCORRELATED  
RESIDUALS

LM TEST = 133.45, DF = 24,  
P-VALUE < 2.2E-16

DW = 0.1269811

Residuals from Linear regression model



## BIG RETAIL STORES

TSLM with correlated expenses and trend

Variables included:

- Natural gas (US)
- Soybean meal
- Wheat, US SRW
- Trend

### Price per Kilogram of Tortillas in Big Retail Stores



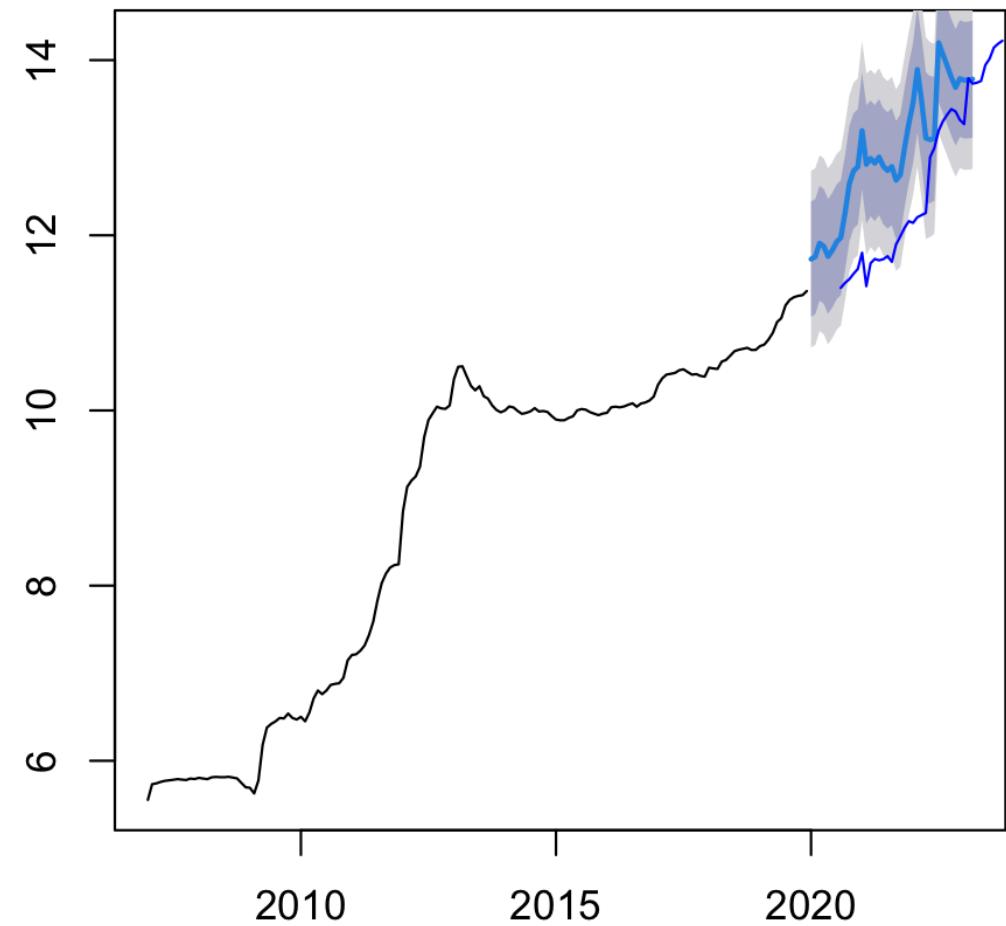
## BIG RETAIL STORES

Forecast of TSLM with correlated expenses and trend

RMSE = 0.543

MPE = -2.609%

**Forecast of the Price per Kilogram of Tortillas for Big Retail store**



# BIG RETAIL STORES

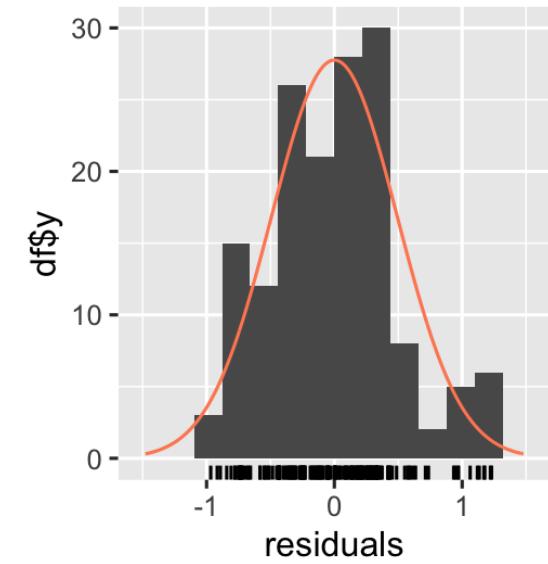
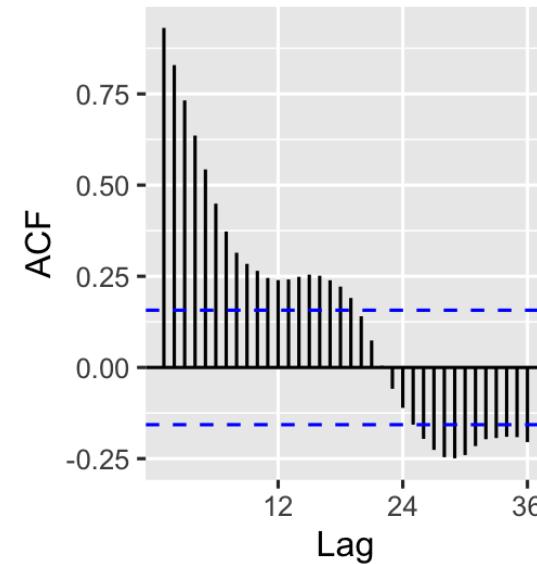
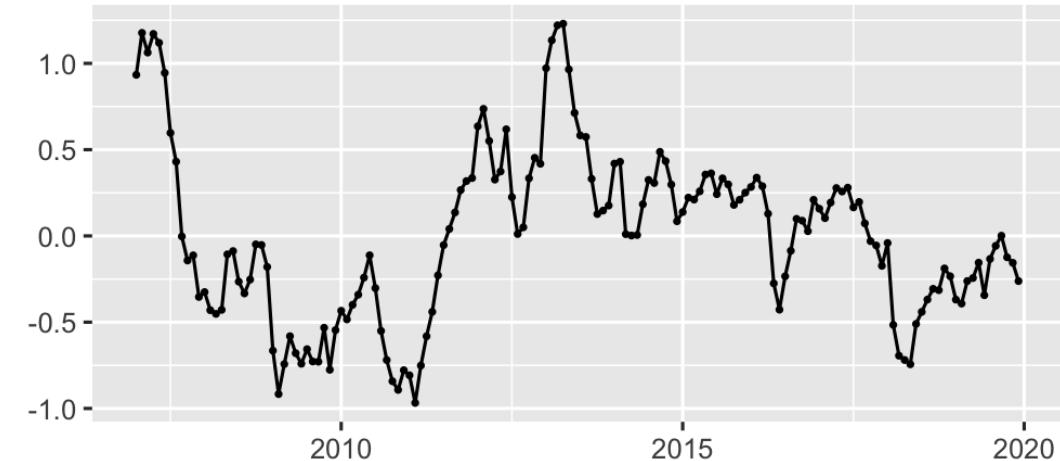
## Residuals from TSLM

AUTOCORRELATED  
RESIDUALS

LM TEST = 140.14, DF = 24,  
P-VALUE < 2.2 E-16

DW = 0.1132

Residuals from Linear regression model



# Autoregressive Integrated Moving Average Model (ARIMA)

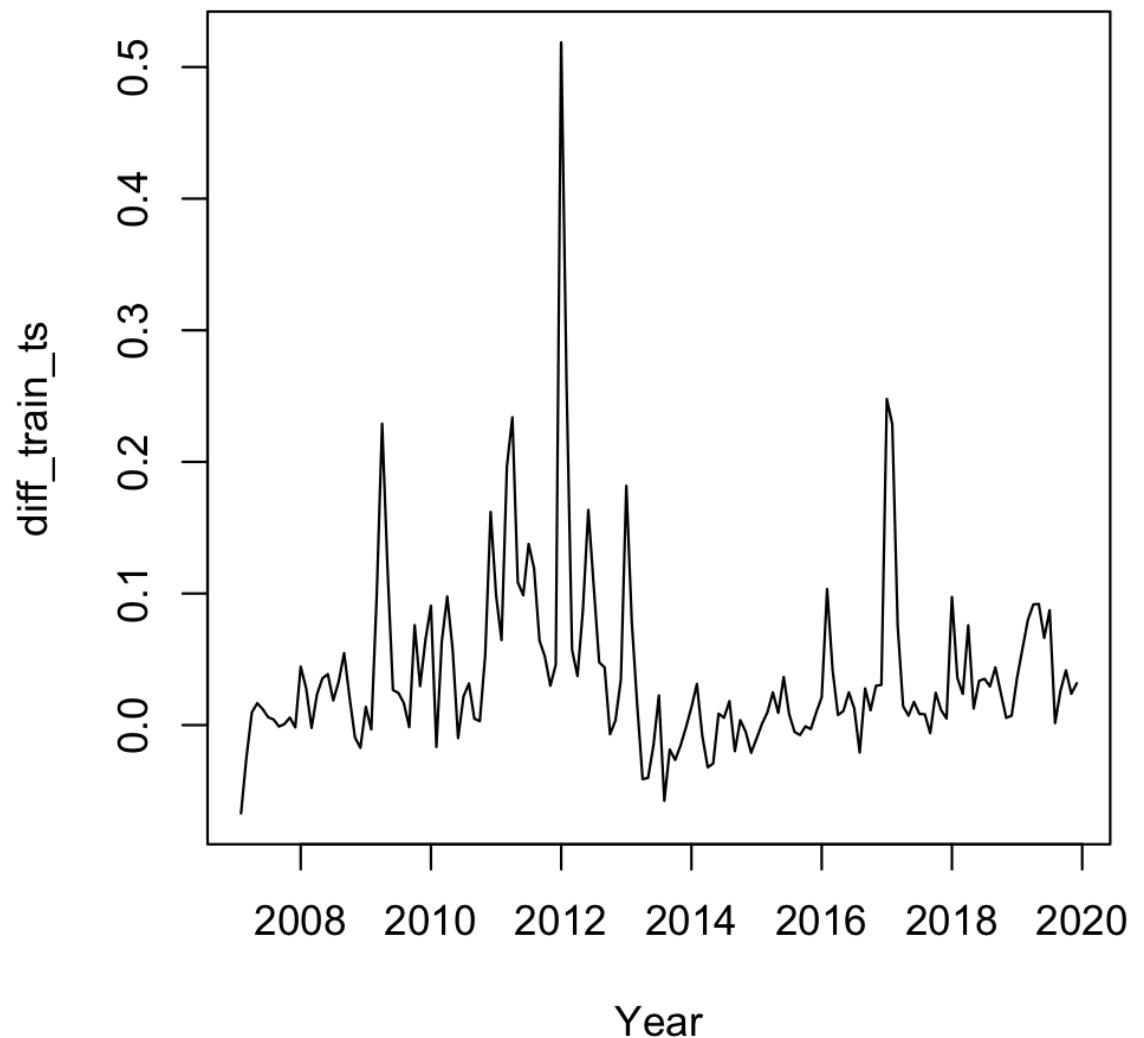
Fit a model that captures *autocorrelation* in time series data by combining *autoregression*, *differencing*, and *moving averages*.

---

# Differenced series

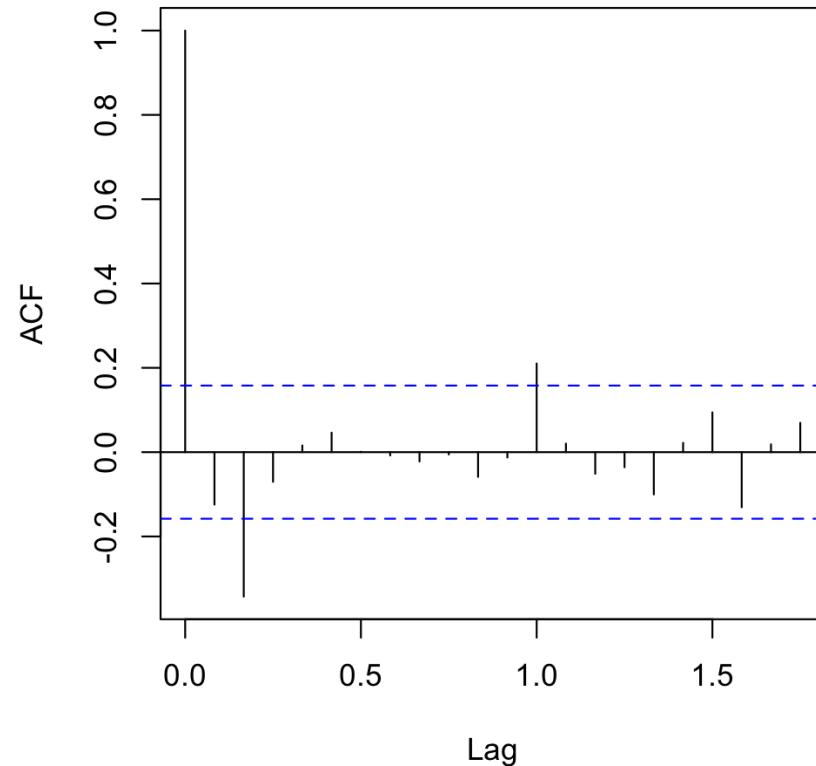
To remove trend and make the series stationary.

**Differenced series**

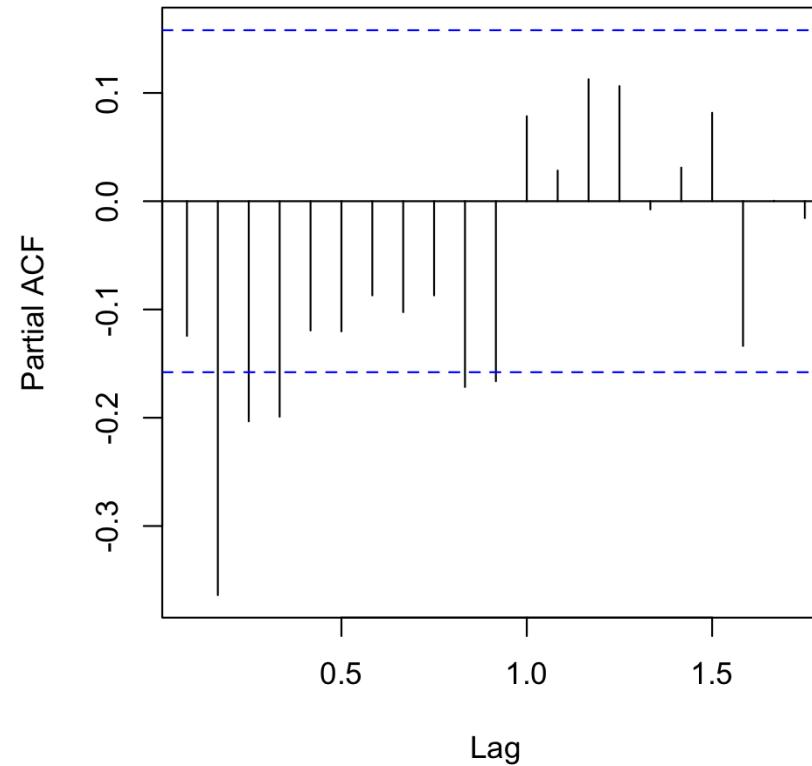


# ACF and PACF on Differenced Series

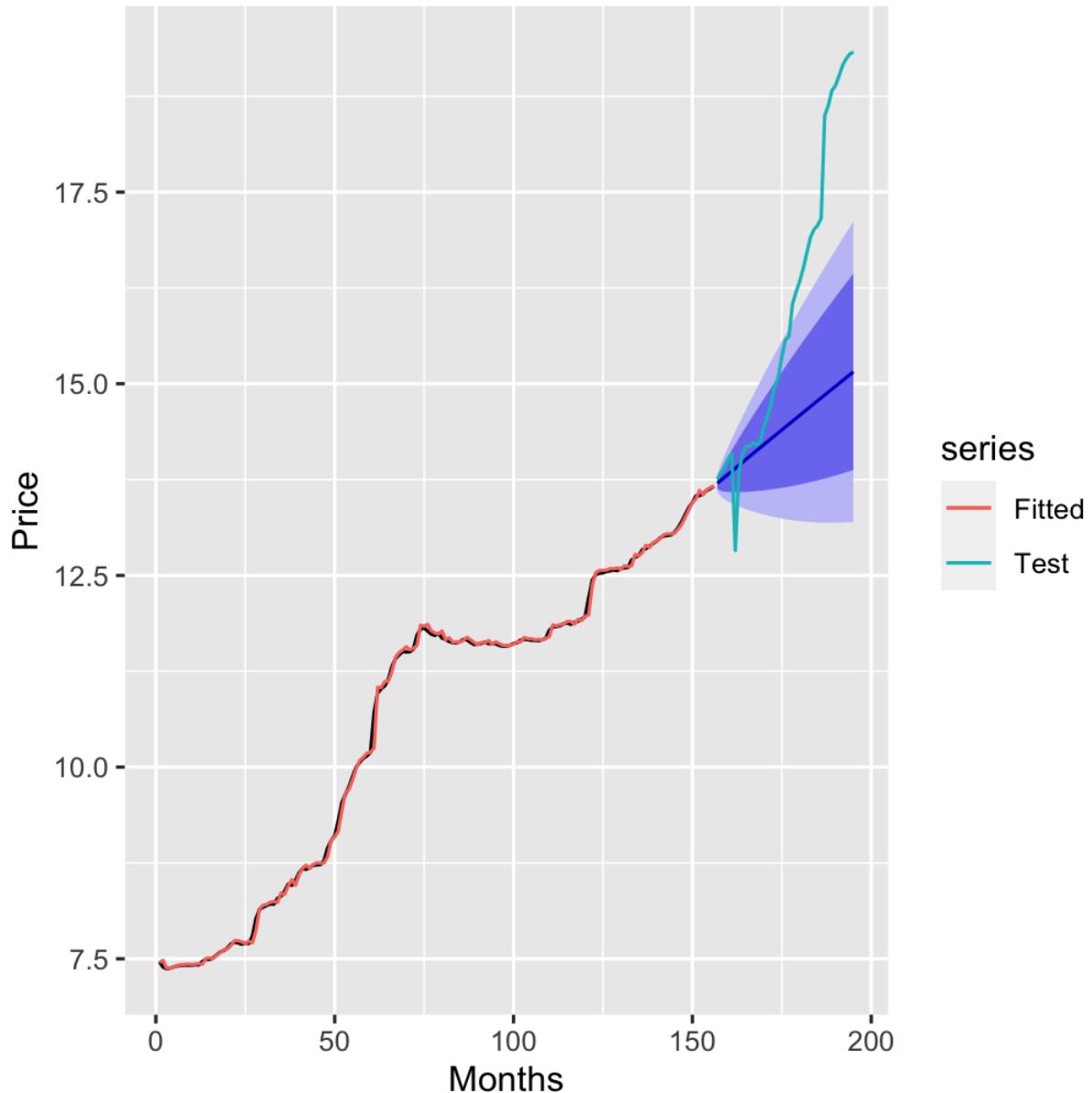
**ACF of Differenced Series**



**PACF of Differenced Series**



## Forecasts from ARIMA(1,1,2) with drift



ARIMA (1,1,2)  
with drift

RMSE = 2.182

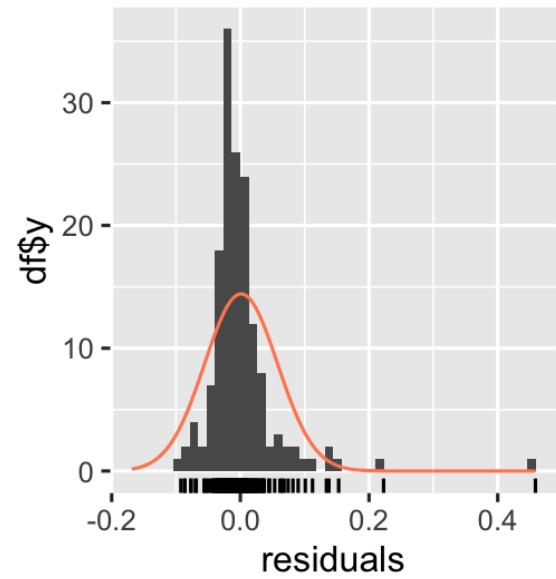
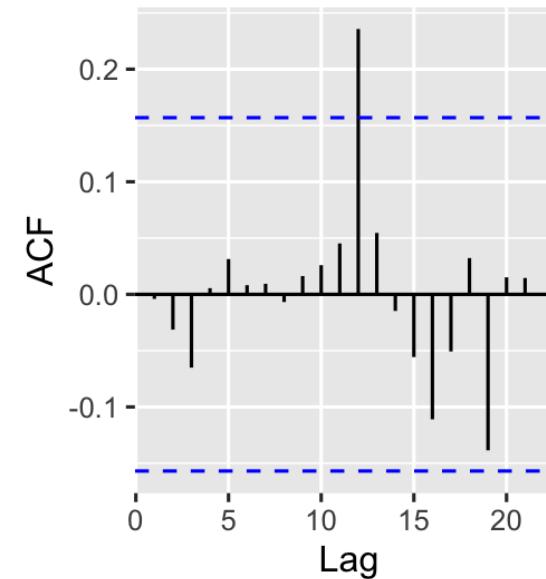
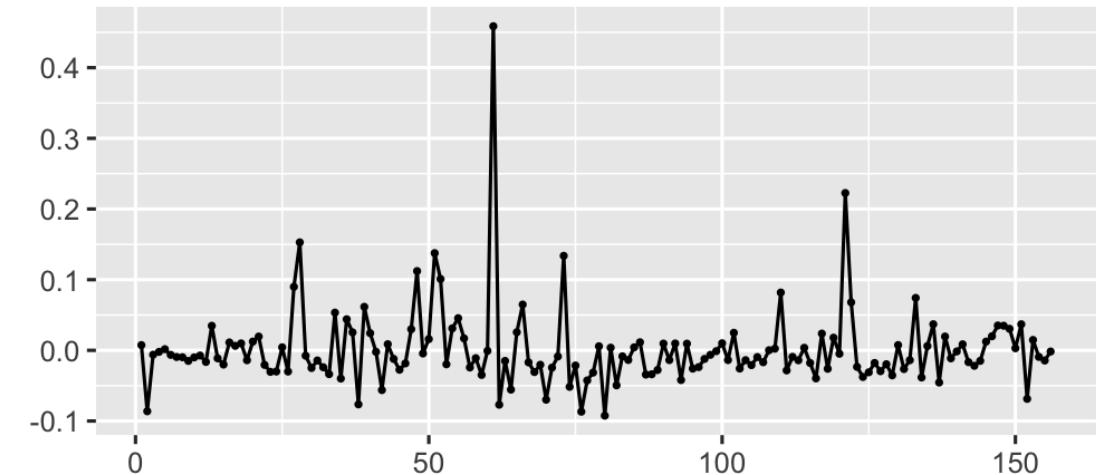
MPE = 8.562%

Forecast and fitted values

# Residuals from ARIMA (1,1,2) with drift

WHITE NOISE  
RESIDUALS

Residuals from ARIMA(1,1,2) with drift

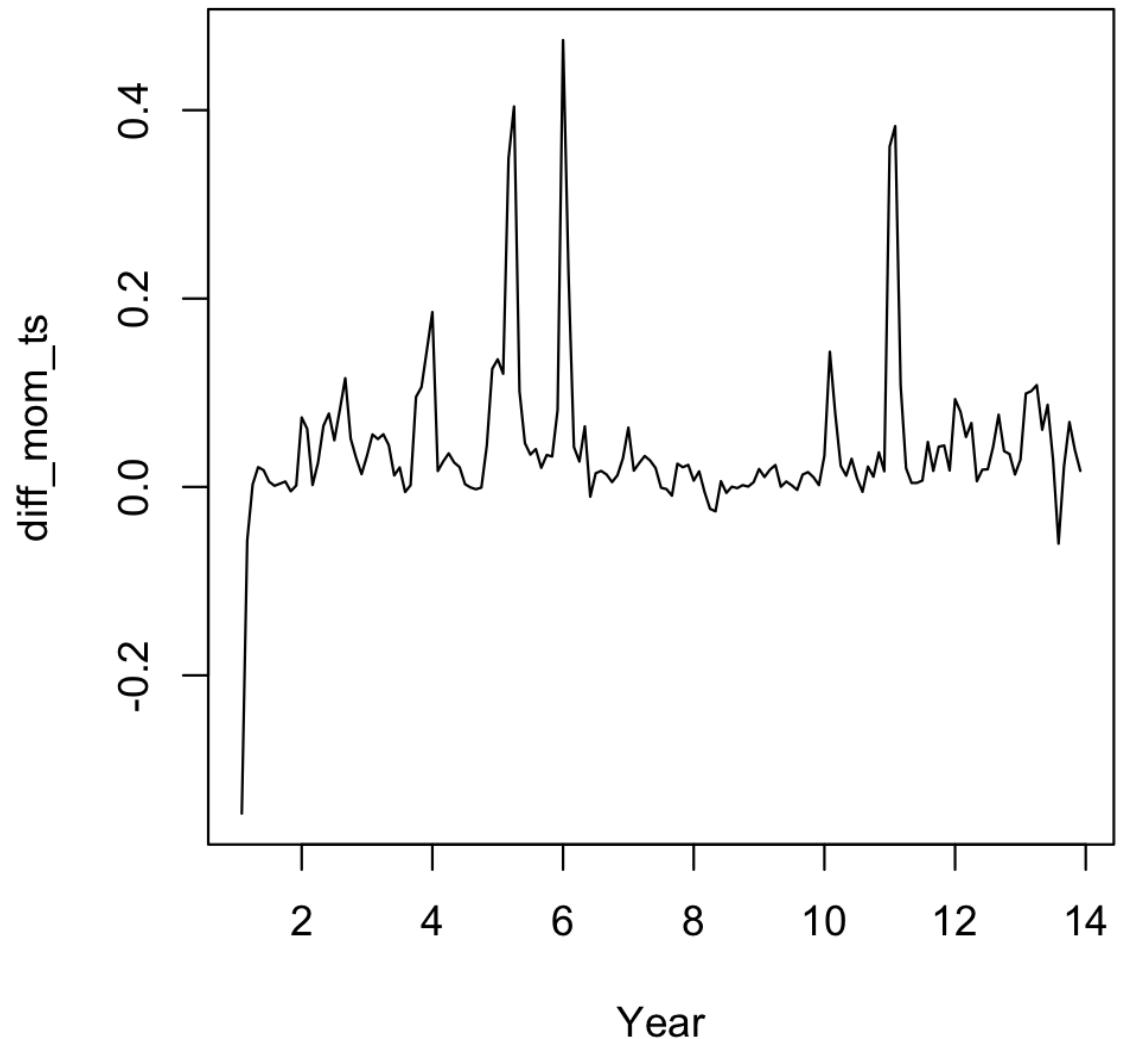


## MOM AND POP STORES

### Differenced series

To remove trend and make the series stationary.

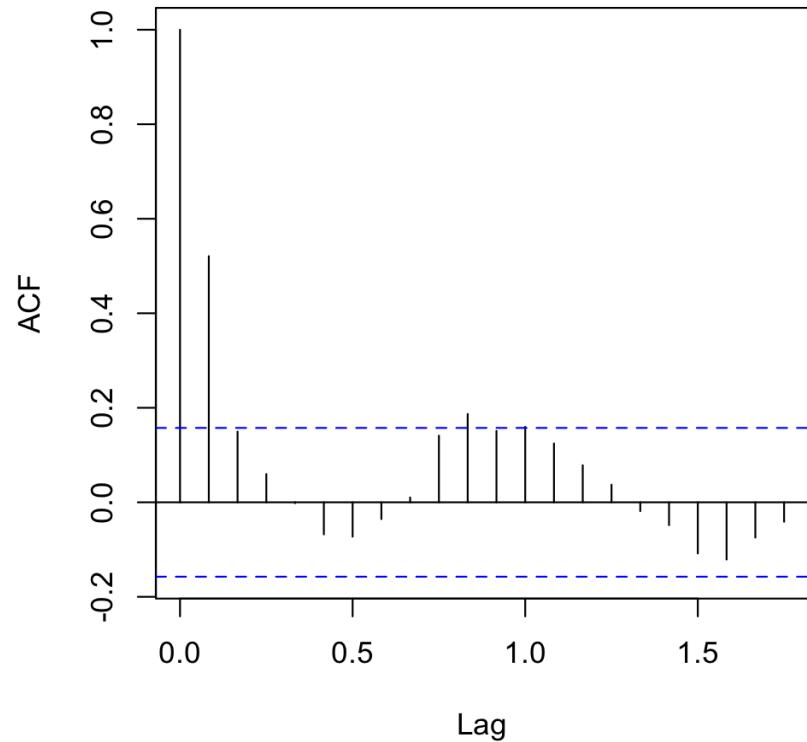
**Differenced series**



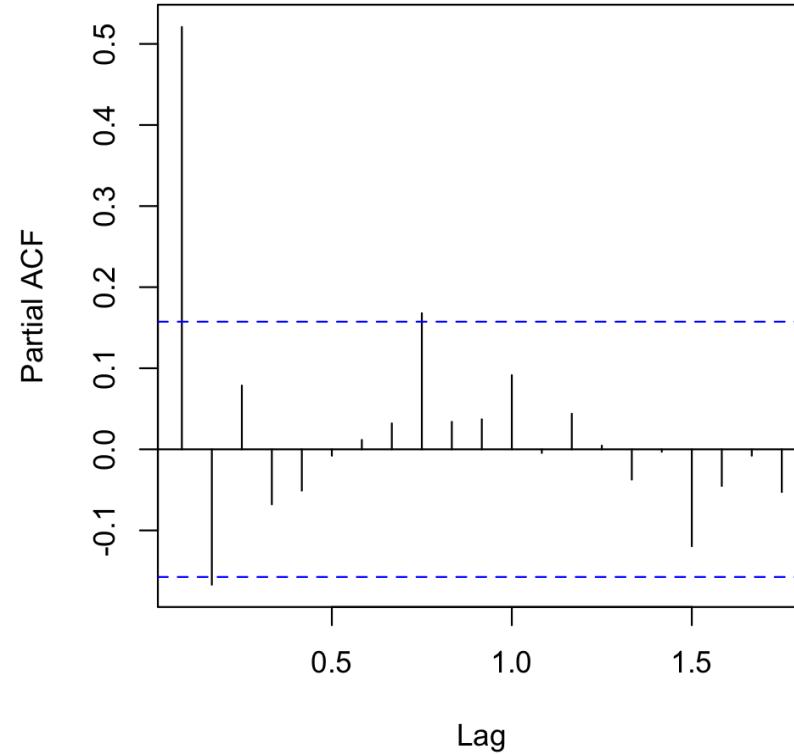
MOM AND POP STORES

# ACF and PACF on Differenced Series

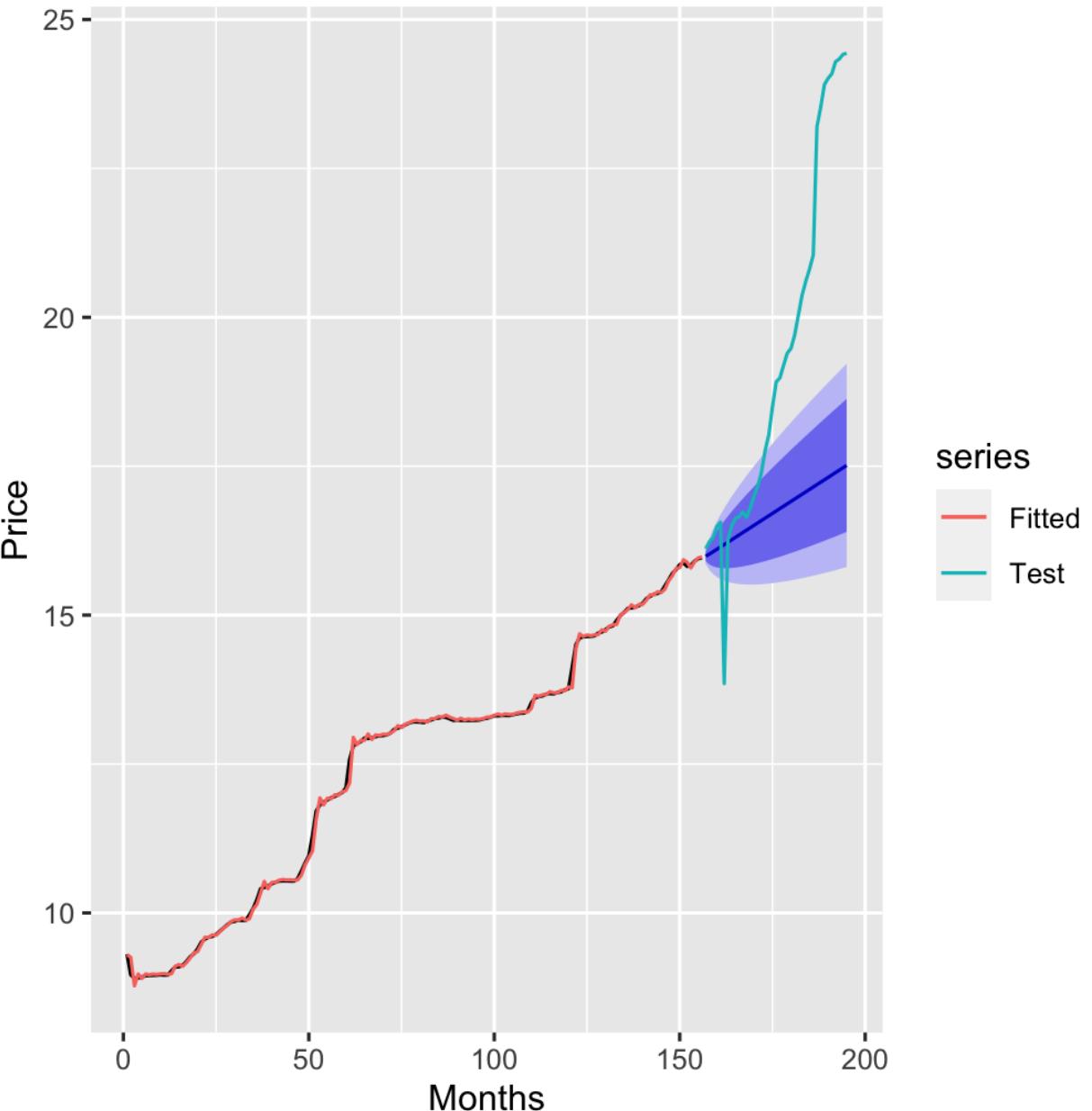
**ACF of Differenced Series**



**PACF of Differenced Series**



## Forecasts from ARIMA(1,1,1) with drift



# ARIMA (1,1,1) with drift

RMSE = 3.626

MPE = 11.467%

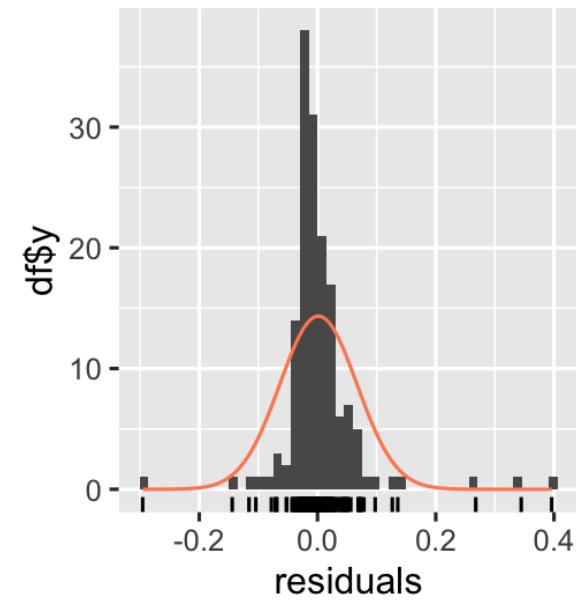
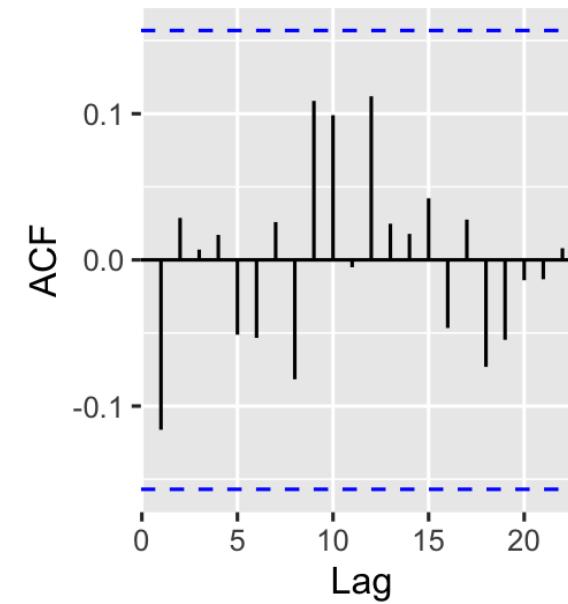
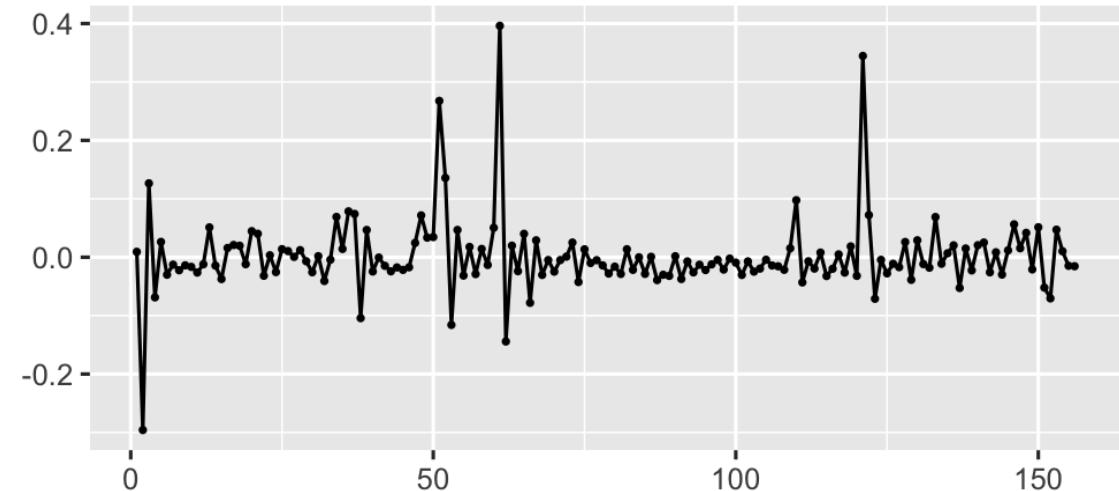
Forecast and fitted values

MOM AND POP STORES

Residuals form  
ARIMA (1,1,1)  
with drift

WHITE NOISE  
RESIDUALS

Residuals from ARIMA(1,1,1) with drift

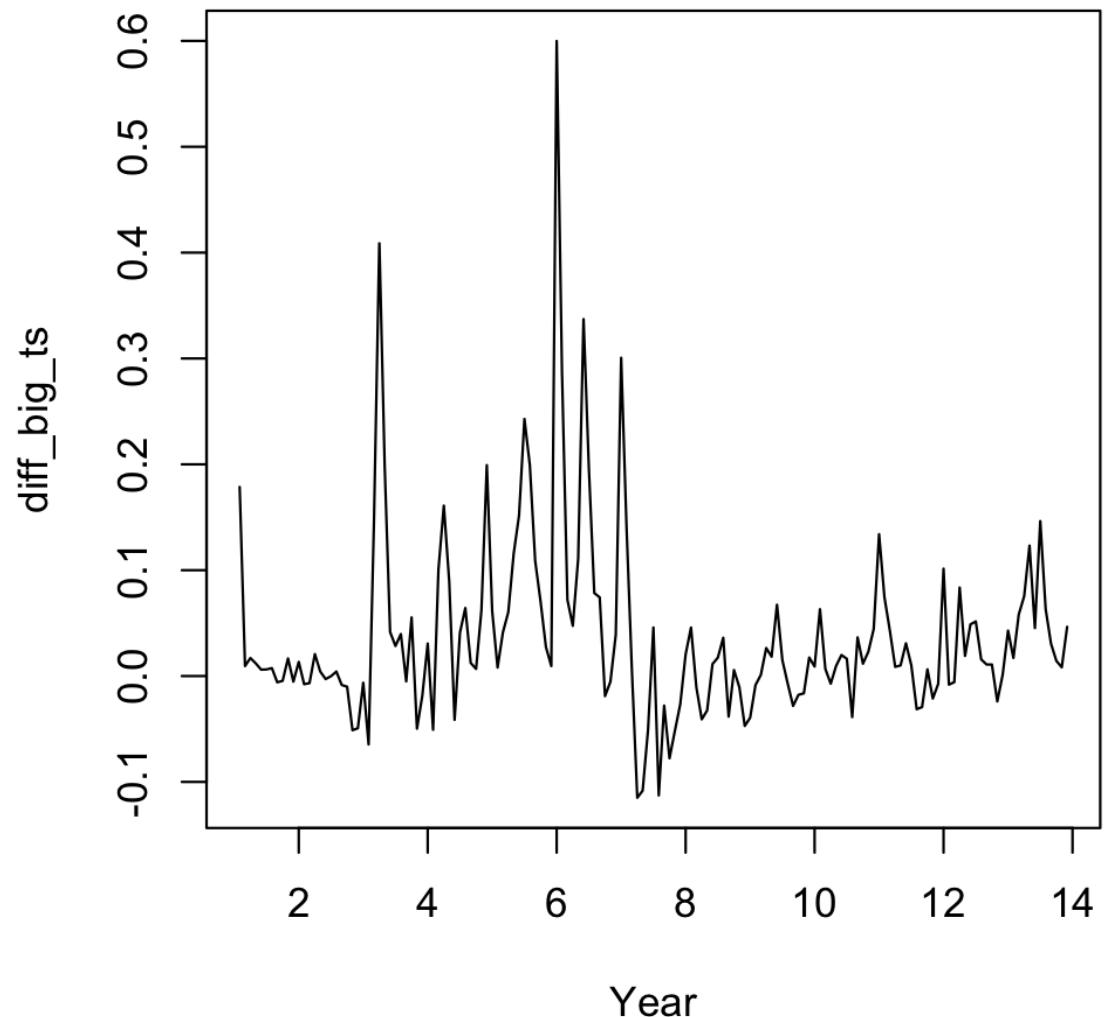


## BIG RETAIL STORES

### Differenced series

To remove trend and make the series stationary.

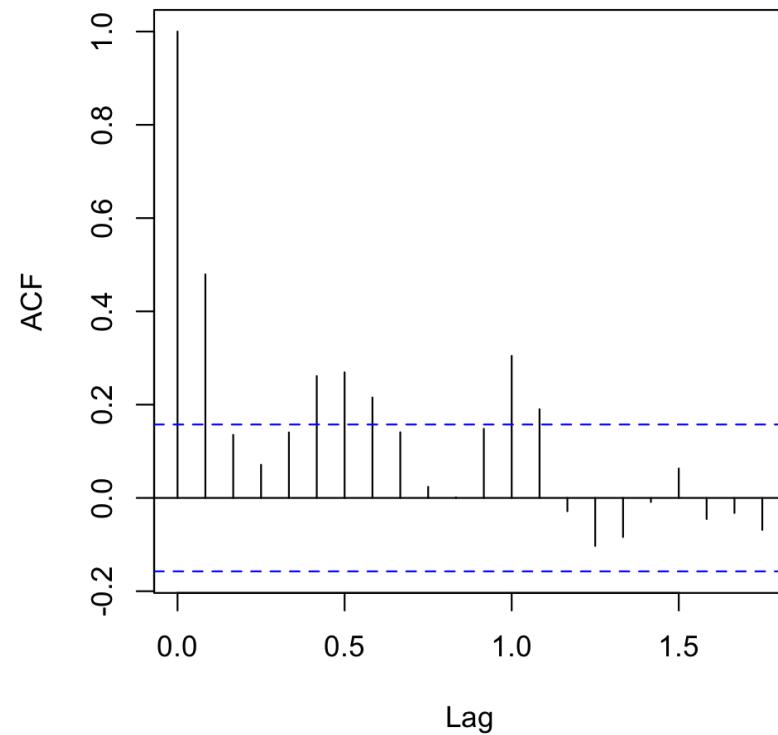
**Differenced series**



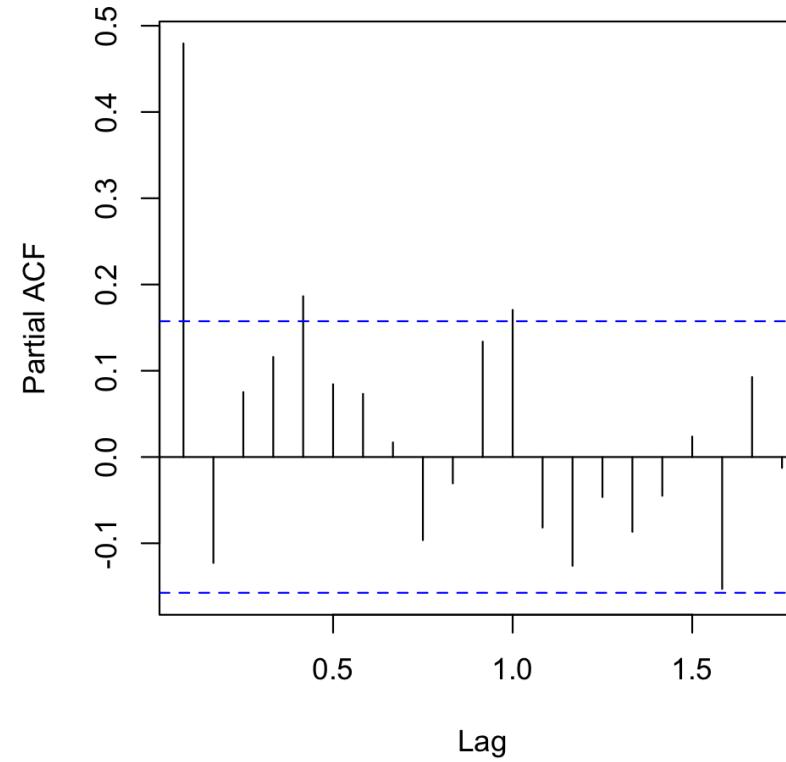
BIG RETAIL STORES

# ACF and PACF on Differenced Series

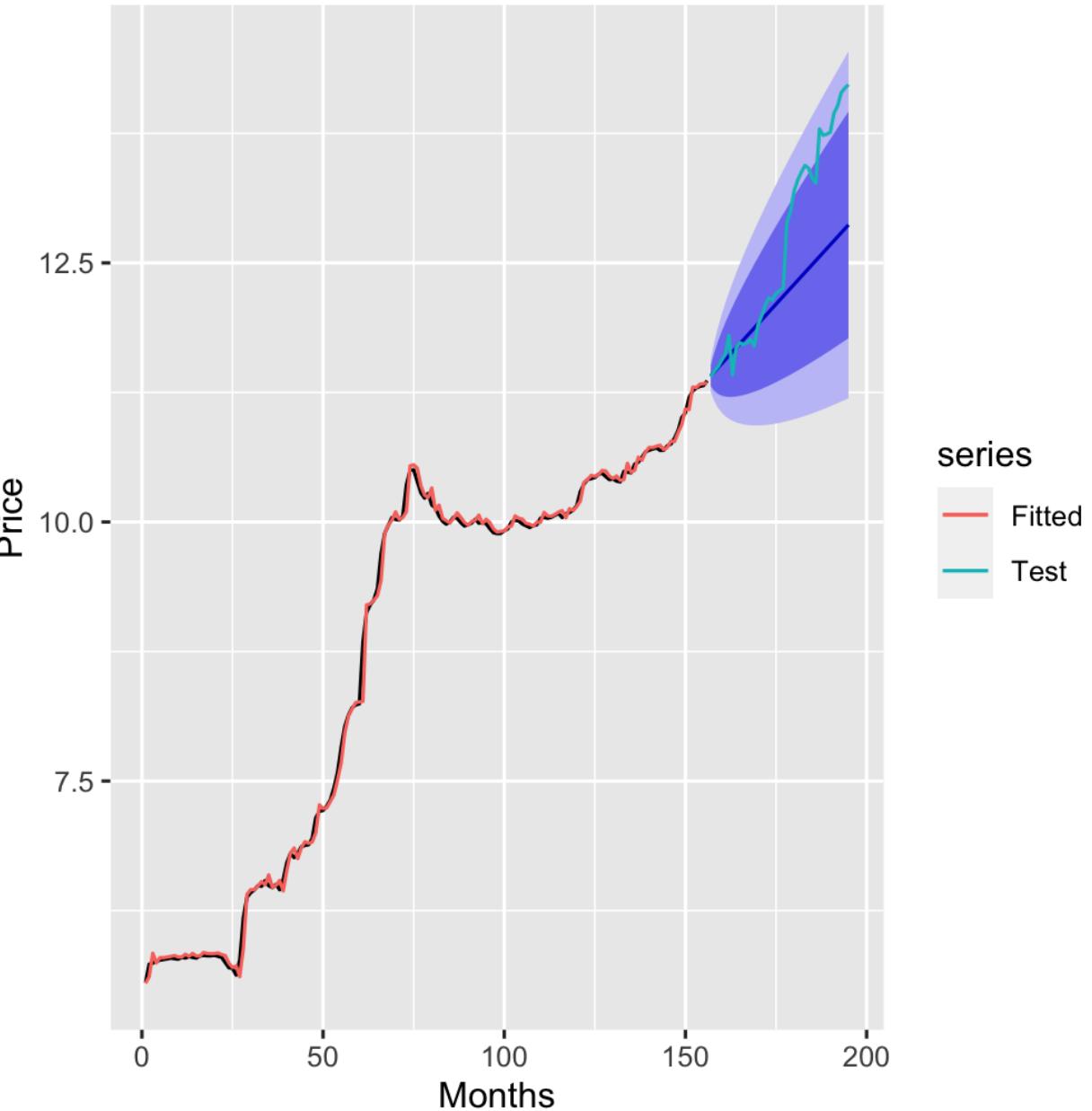
**ACF of Differenced Series**



**PACF of Differenced Series**



## Forecasts from ARIMA(1,1,1) with drift



# ARIMA (1,1,1) with drift

RMSE = 0.734

MPE = 3.625%

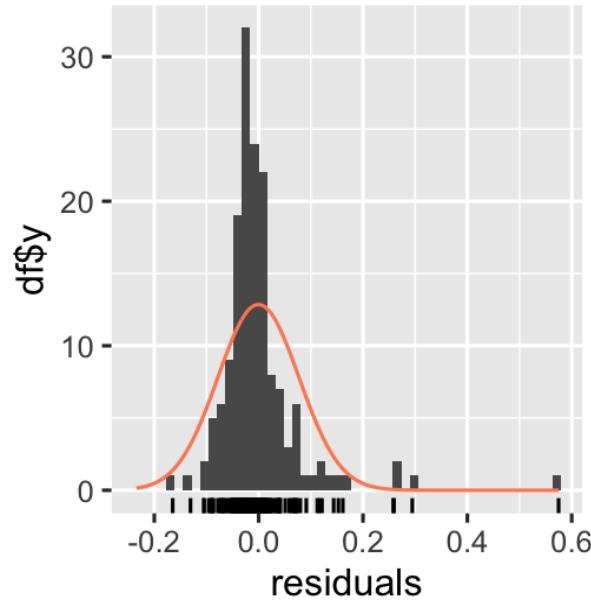
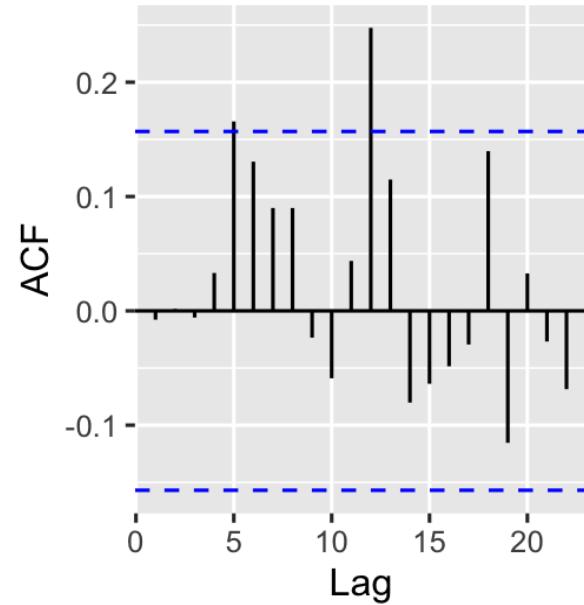
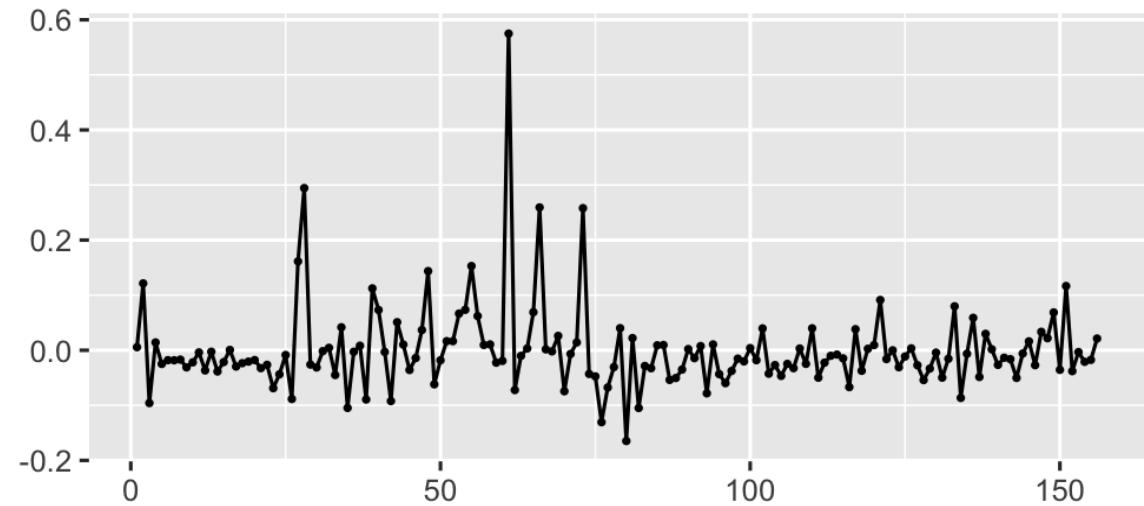
Forecast and fitted values

BIG RETAIL STORES

# Residuals from ARIMA (1,1,1) with drift

WHITE NOISE  
RESIDUALS

Residuals from ARIMA(1,1,1) with drift

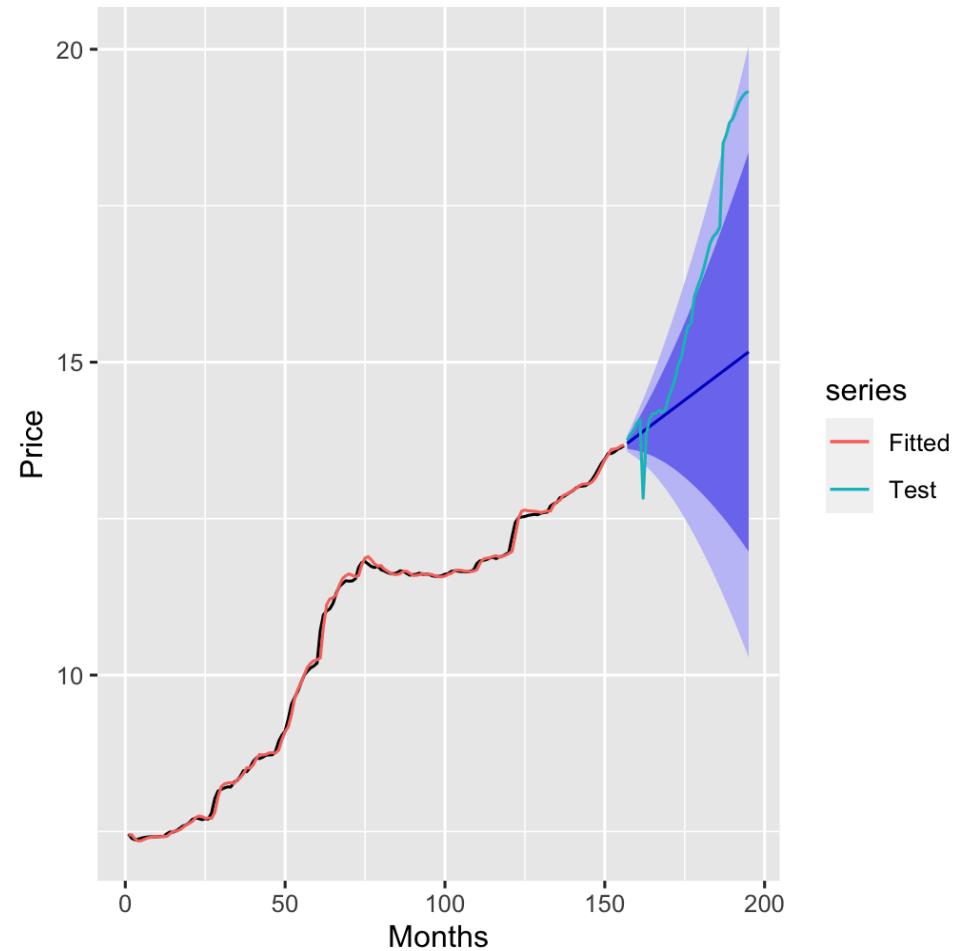


# Holt's Exponential Smoothing

Smooth time series data by combining *level* and *trend* components with *exponential smoothing* to forecast future values.

---

Forecasts from Holt's method



$$\begin{aligned}\alpha &= 0.9999, \\ \beta &= 0.2394\end{aligned}$$

# Holt's Exponential Smoothing

Forecast and fitted values  
of Holt's method.

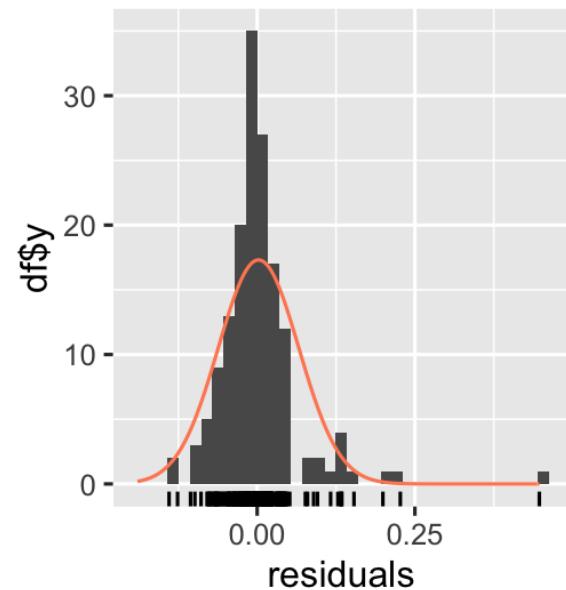
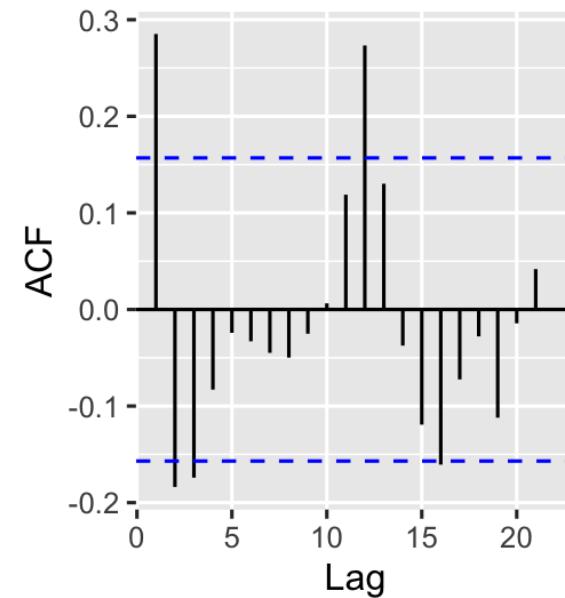
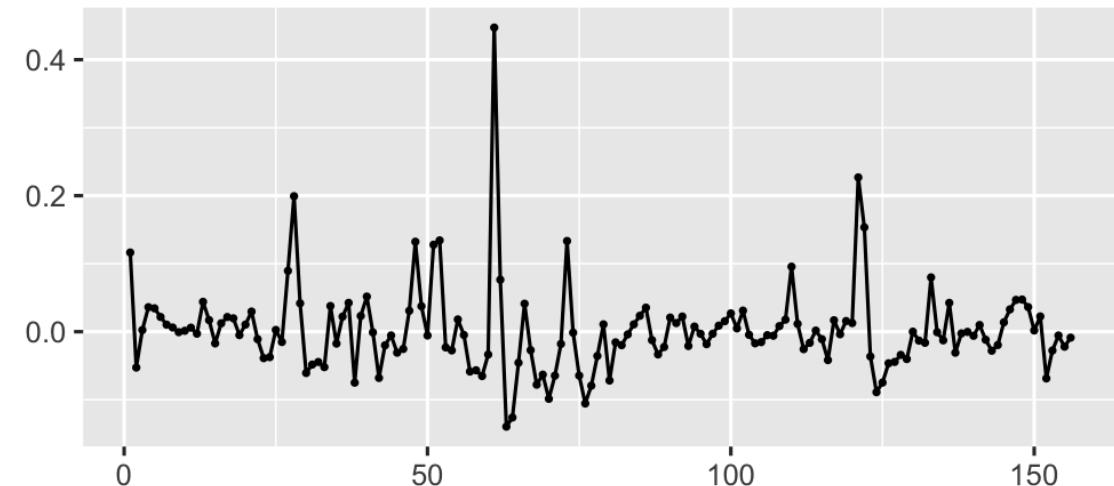
RMSE = 2.180

MPE = 8.57%

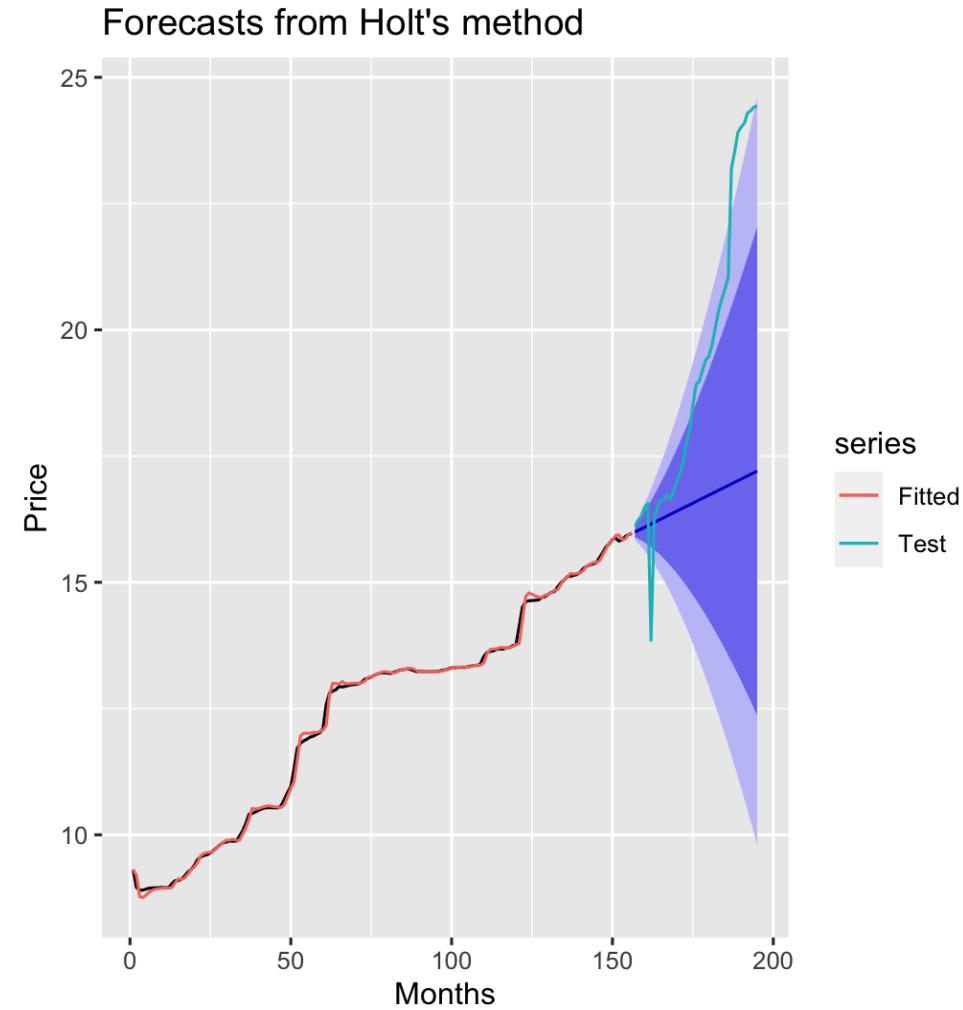
# Residuals from Holt's Method

AUTOCORRELATED RESIDUALS  
 $Q^* = 25.474$ , DF = 10,  
P-VALUE = 0.004516

Residuals from Holt's method



# Holt's Exponential Smoothing



$$\alpha = 0.9999, \beta = 0.3067$$

Forecast and fitted values  
of Holt's method.

RMSE = 3.795

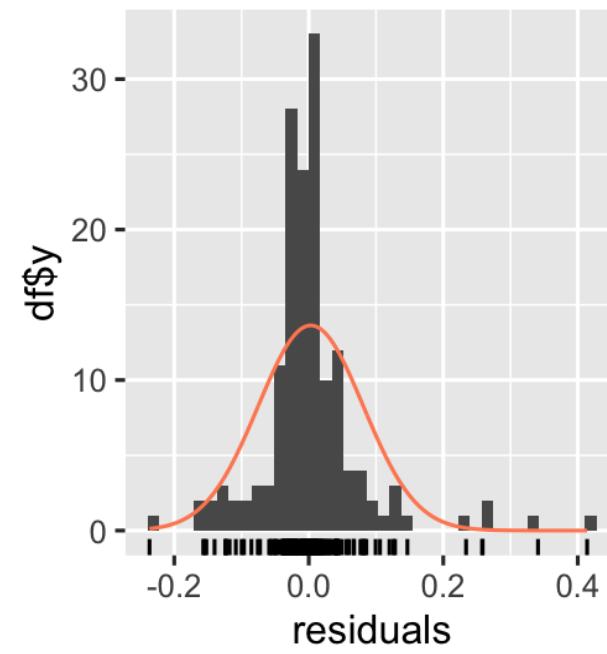
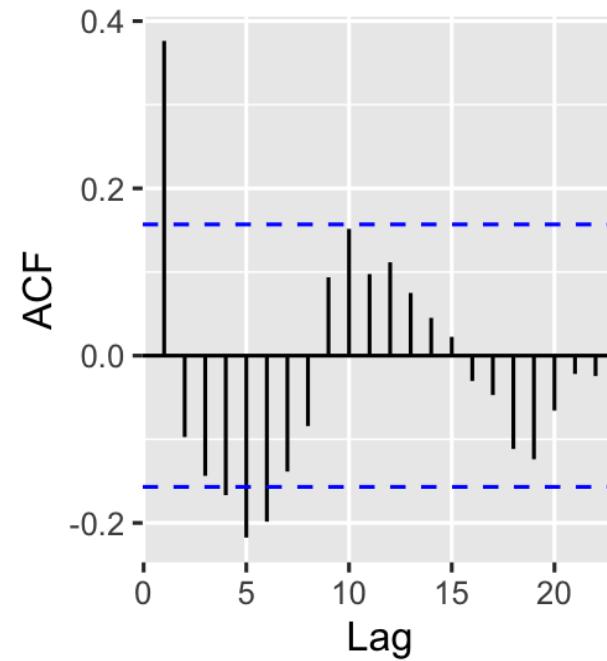
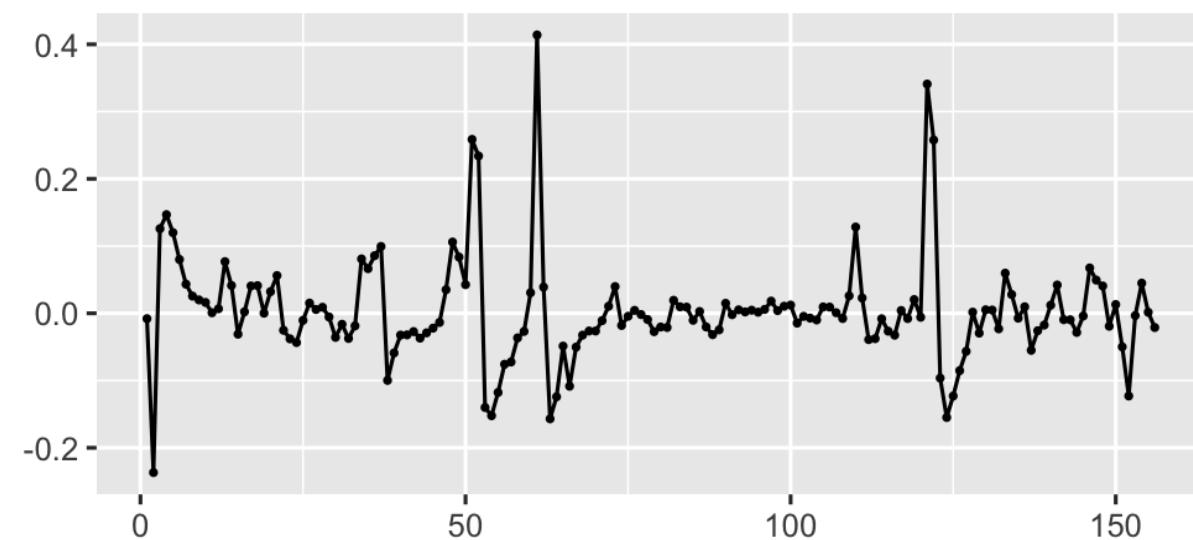
MPE = 12.187%

MOM AND POP STORES

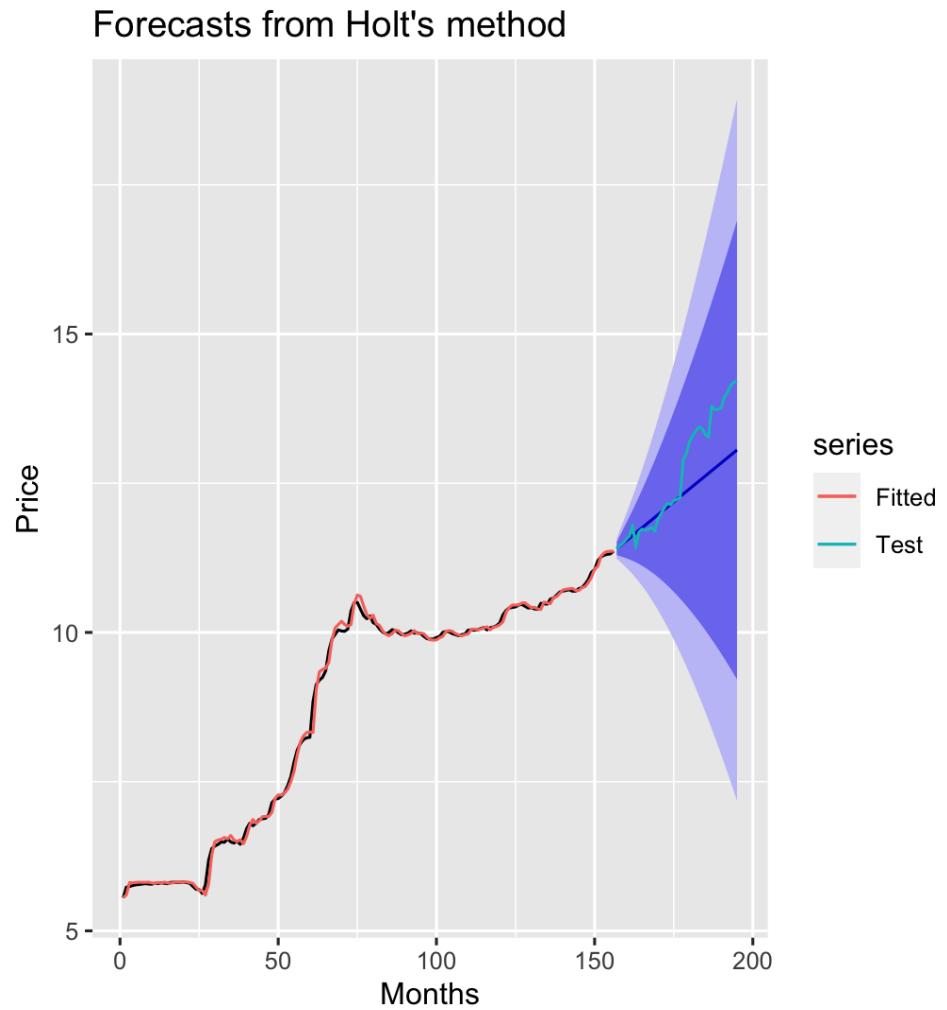
# Residuals from Holt's Method

AUTOCORRELATED RESIDUALS  
 $Q^* = 55.772$ , DF = 10,  
P-VALUE =  $2.265 \times 10^{-8}$

Residuals from Holt's method



# Holt's Exponential Smoothing



$$\alpha = 0.9999, \beta = 0.2152$$

Forecast and fitted values  
of Holt's method.

RMSE = 0.668

MPE = 3.176%

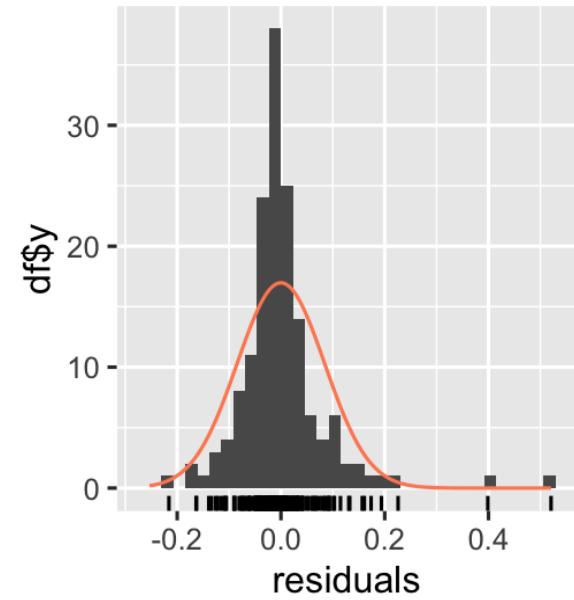
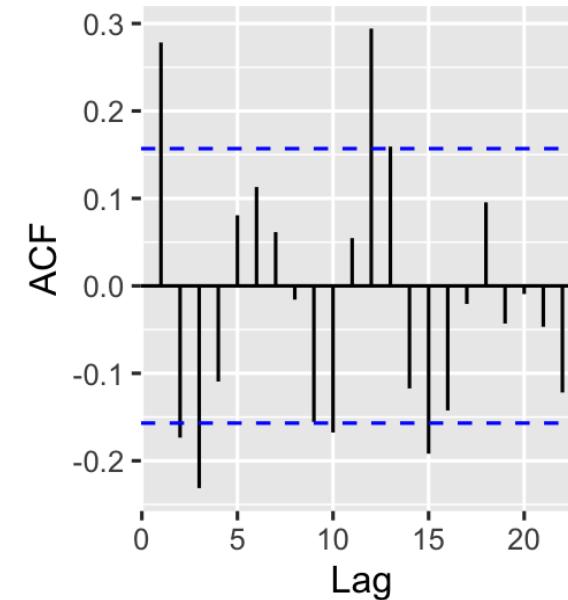
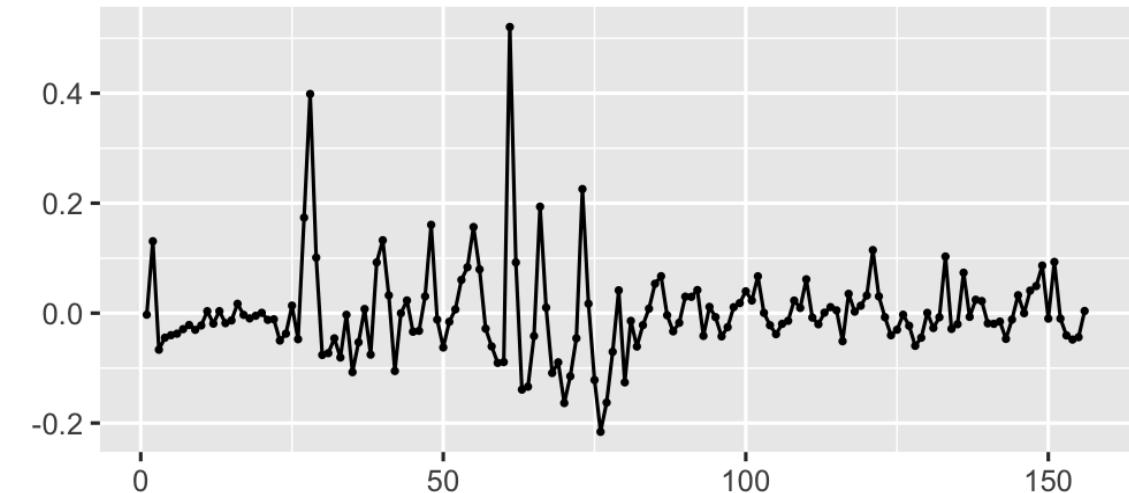
MOM AND POP STORES

# Residuals from Holt's Method

AUTOCORRELATED RESIDUALS

$Q^* = 40.324$ , DF = 10,  
P-VALUE =  $1.486 \times 10^{-5}$

Residuals from Holt's method

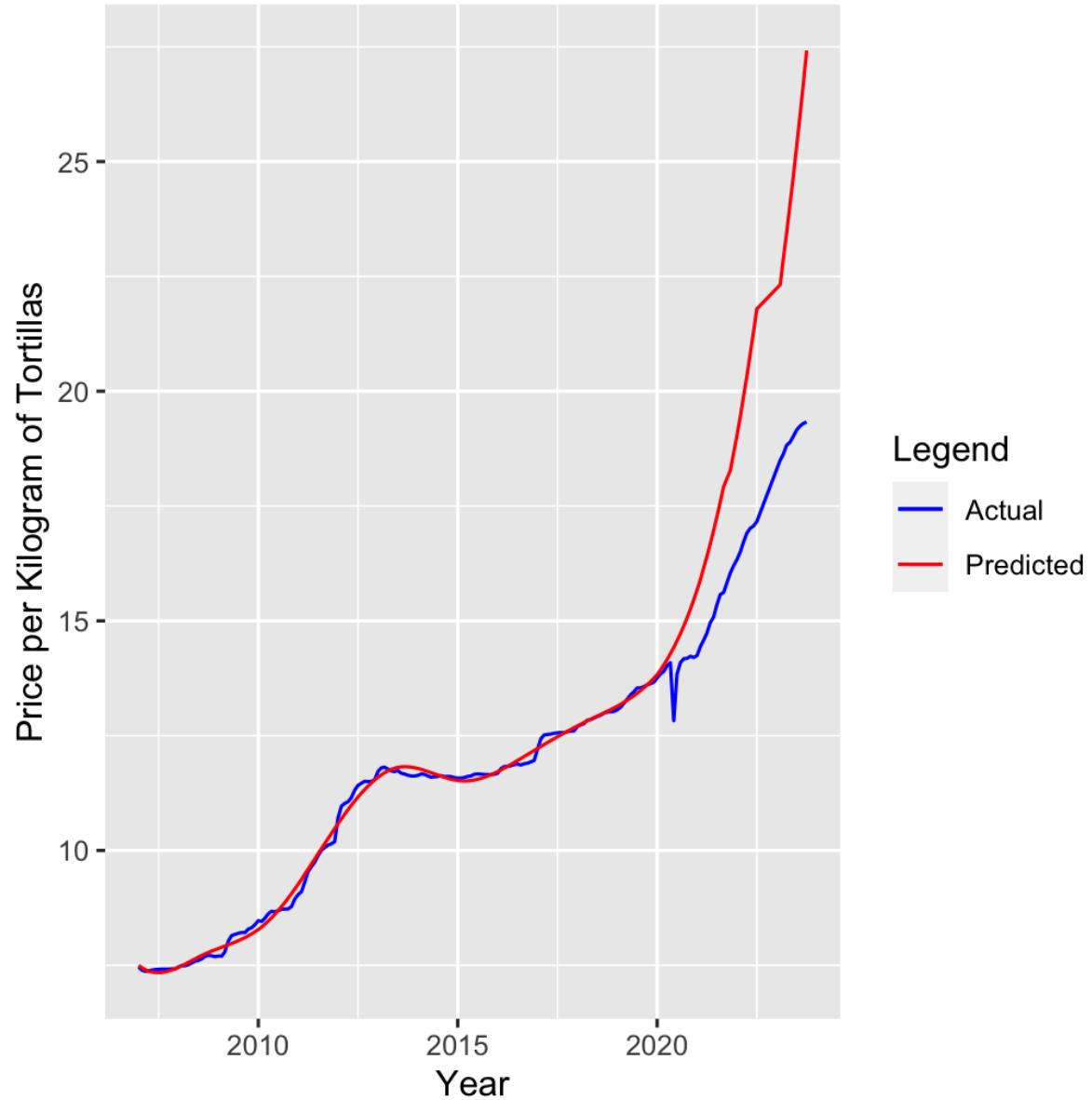


# Regression Splines

Fit *piecewise polynomials* to data, ensuring **smooth transitions** at specified *knots*.

---

## Forecast from Cubic Spline



# Regression Splines

Internal knots = 7

RMSE = 3.415

MPE = -15.622%

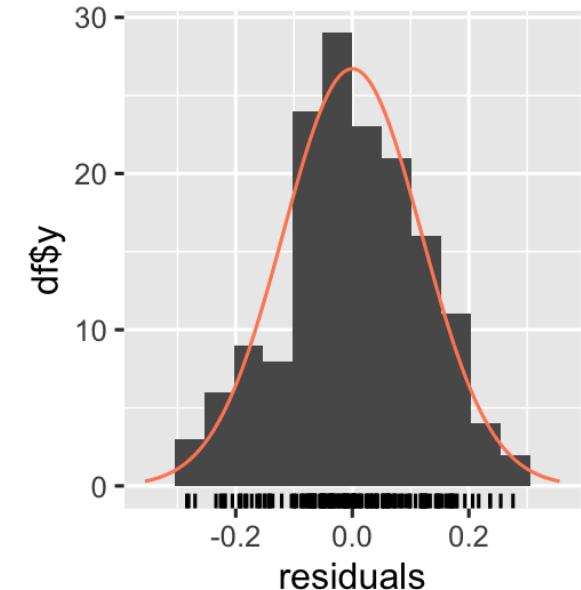
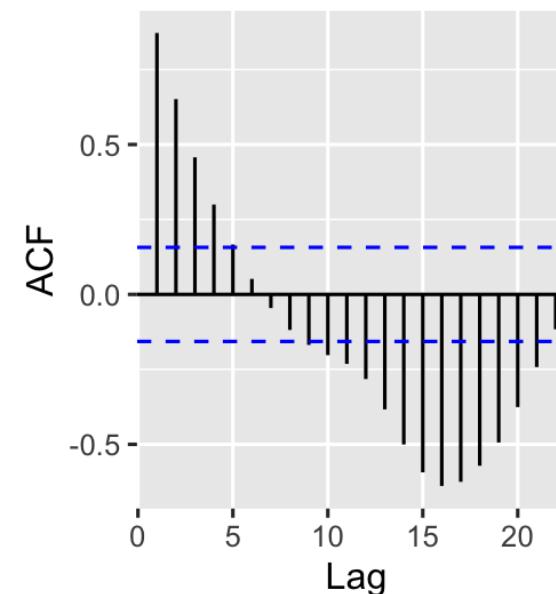
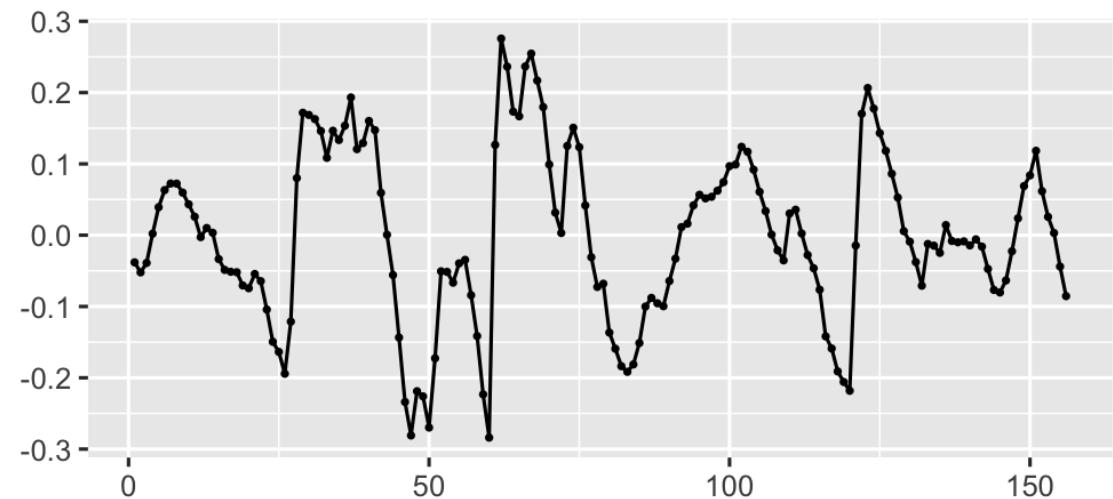
# Residuals from Regression Splines

AUTOCORRELATED  
RESIDUALS

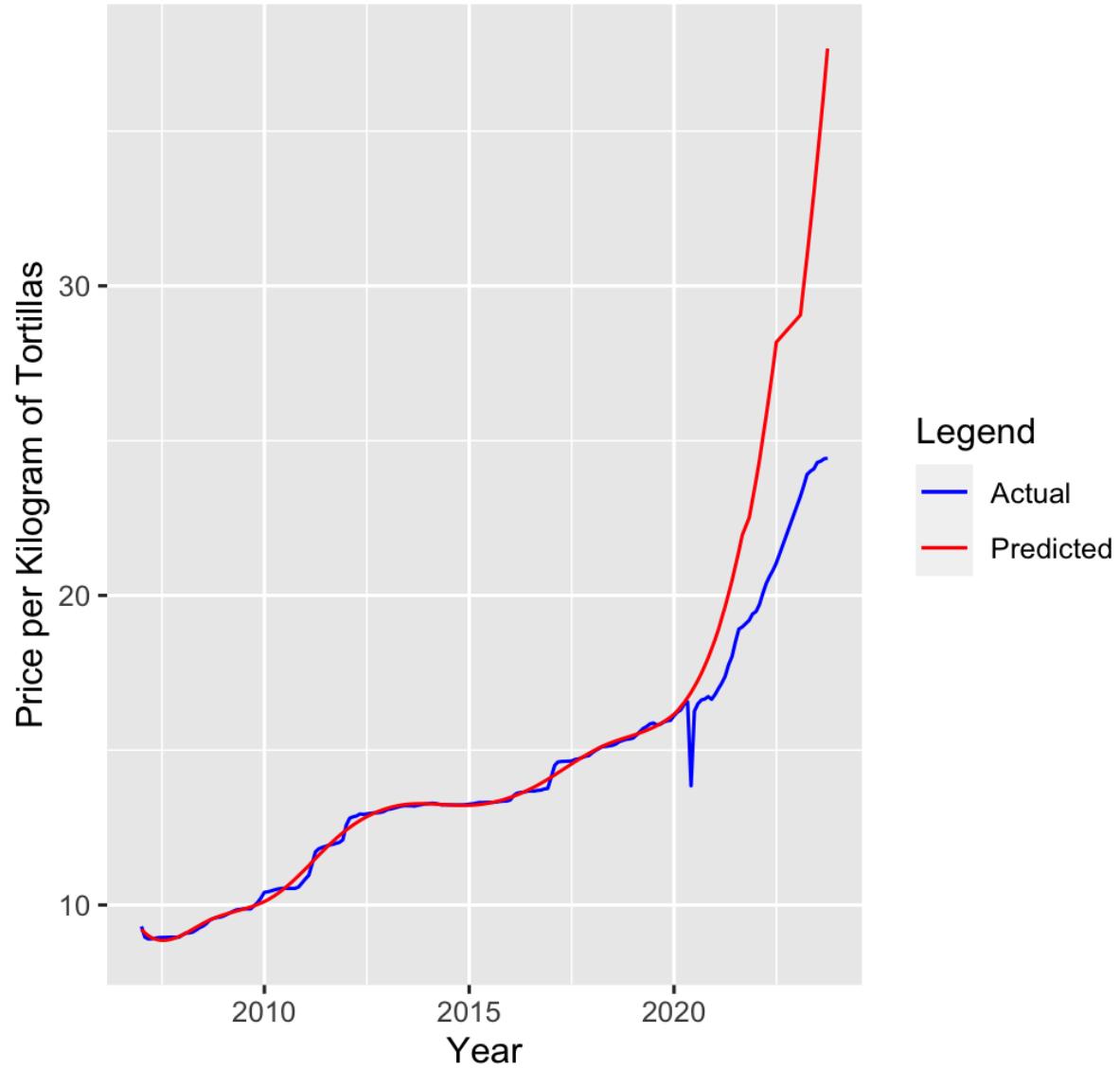
LM TEST = 132.88, DF = 14,  
P-VALUE < 2.2E-16

DW = 0.2520018

Residuals



## Forecast from Cubic Spline for Mom and Pop Store



MOM AND POP STORES

# Regression Splines

Internal knots = 7

RMSE = 5.346

MPE= -19.131%

MOM AND POP STORES

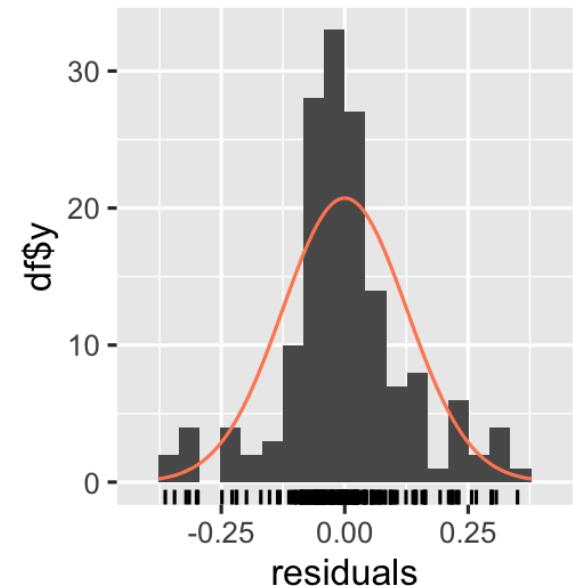
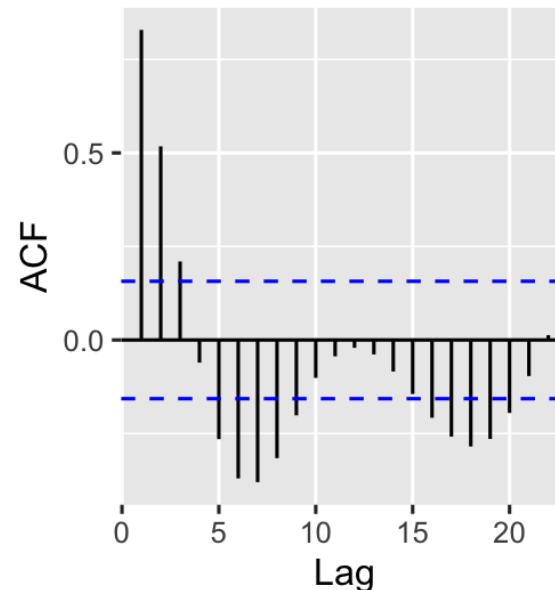
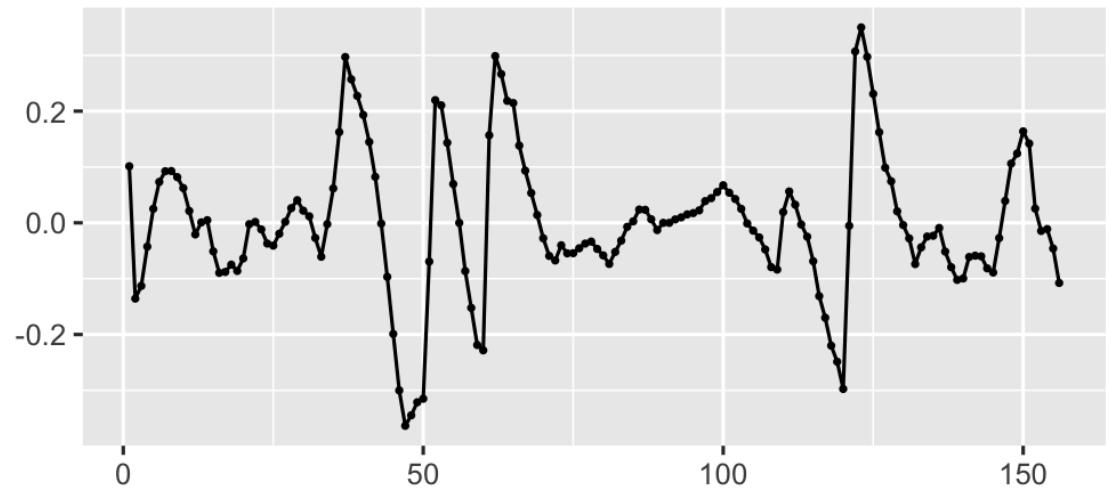
# Residuals from Regression Splines

AUTOCORRELATED  
RESIDUALS

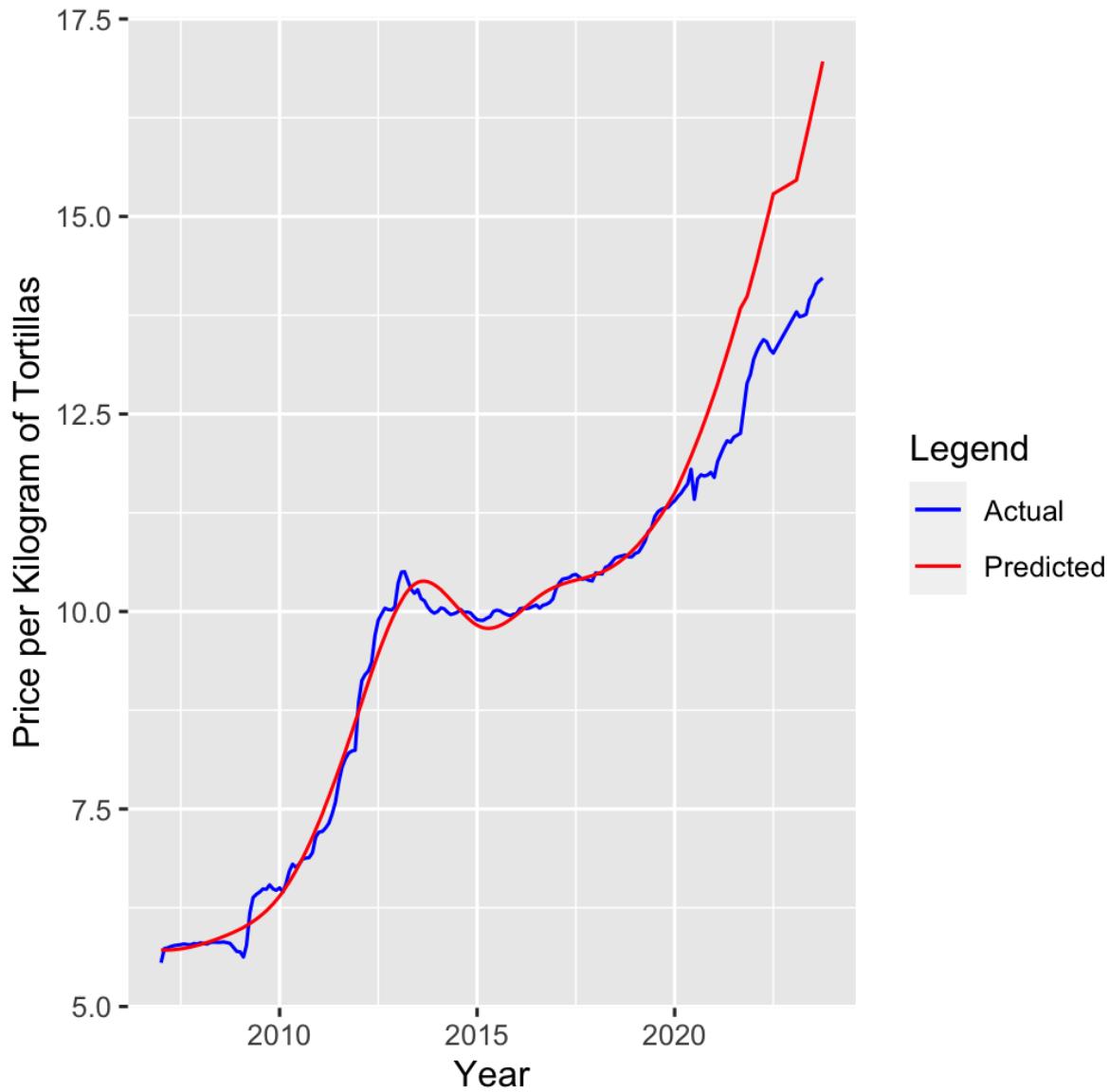
LM TEST = 125.75, DF = 14,  
P-VALUE < 2.2 E-16

DW = 0.3334799

Residuals



## Forecast from Cubic Spline for Big Retail Store



BIG RETAIL STORES

# Regression Splines

Internal knots = 7

RMSE = 1.4403

MPE = -9.531%

# BIG RETAIL STORES

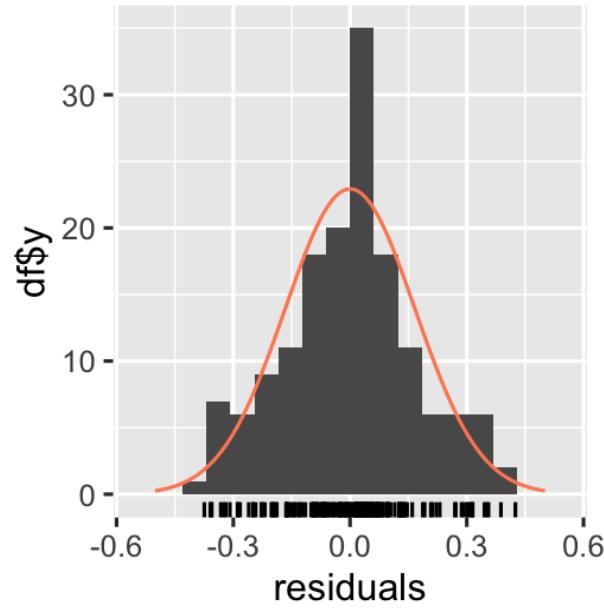
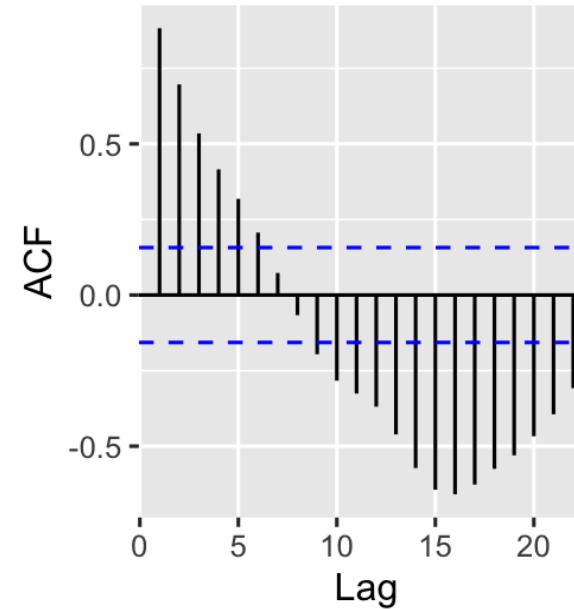
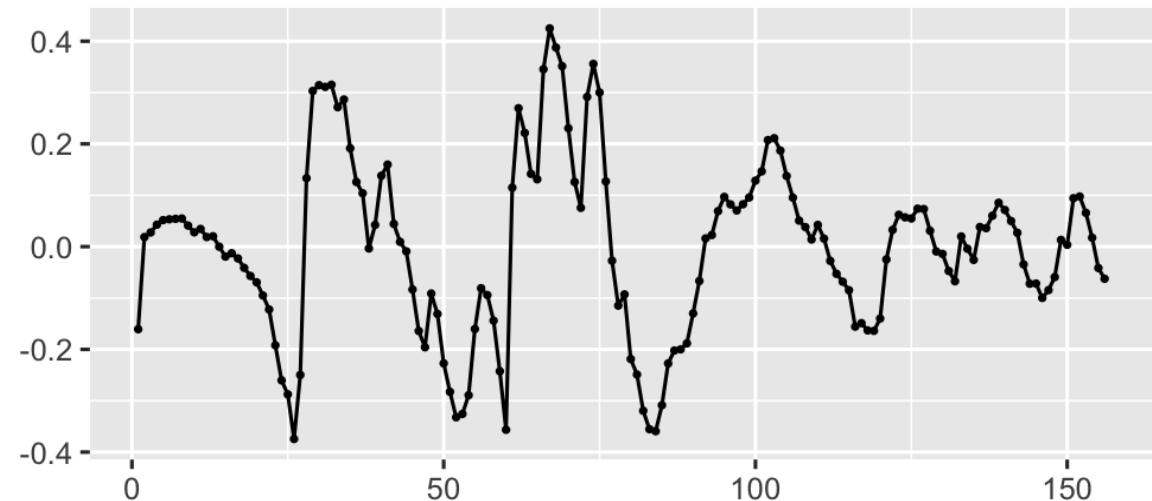
## Residuals from Regression Splines

AUTOCORRELATED  
RESIDUALS

LM TEST = 133.21, DF = 14,  
P-VALUE < 2.2E-16

DW = 0.2275

Residuals

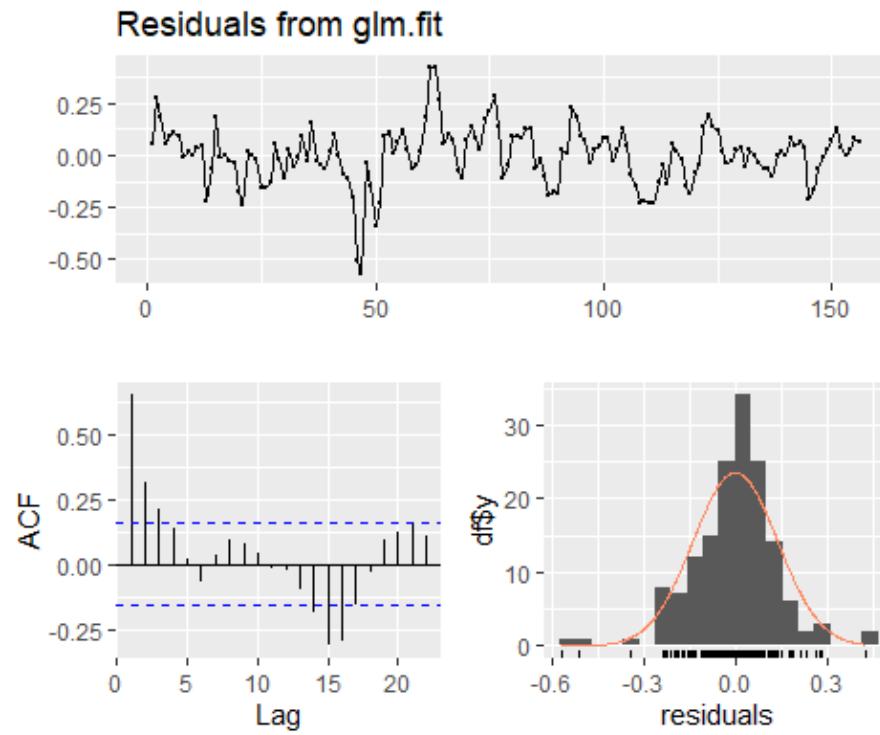
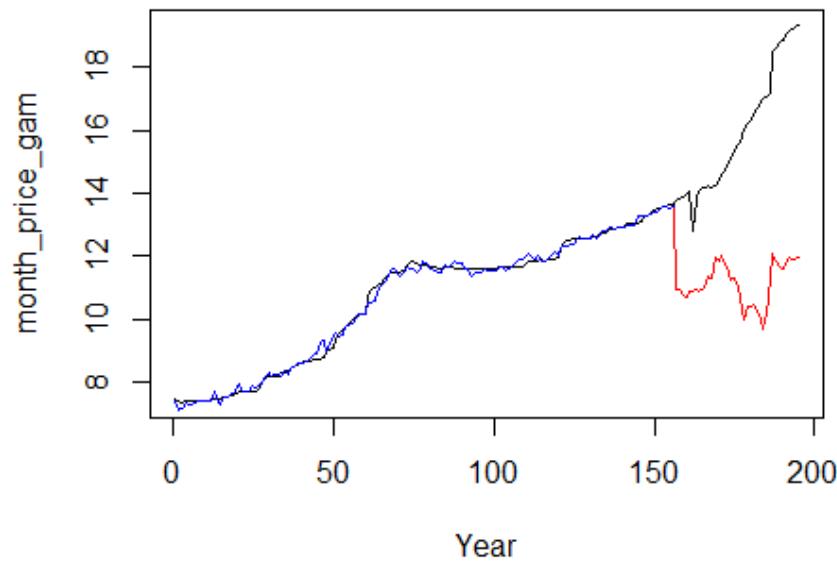


# Generative Additive Model (GAM)

Fit a *non-linear* function to **each variable individually**, and then combine them together.

---

# Stepwise-GAM

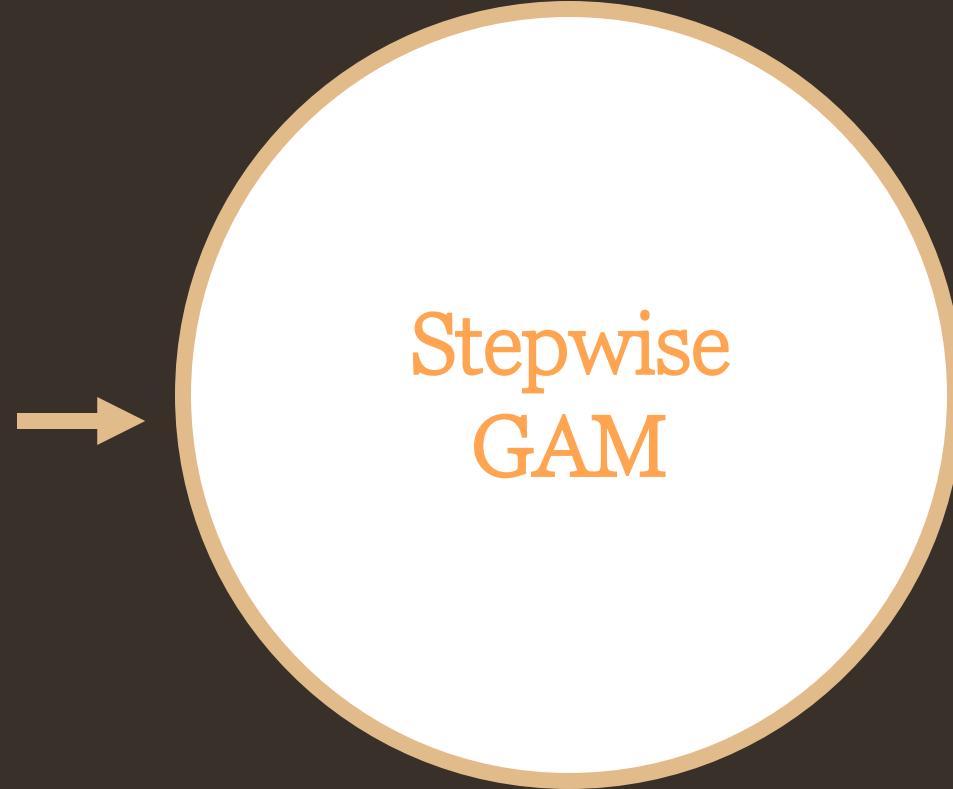


ALL RELEVANT  
EXPENSES

Year  
Month  
Palm\_oil  
Sunflower\_oil  
Rapeseed\_oil  
Natural\_gas\_US  
Wheat\_US\_HRW  
Wheat\_US\_SRW  
Hourly\_Wage\_USD  
Soybean\_meal  
time

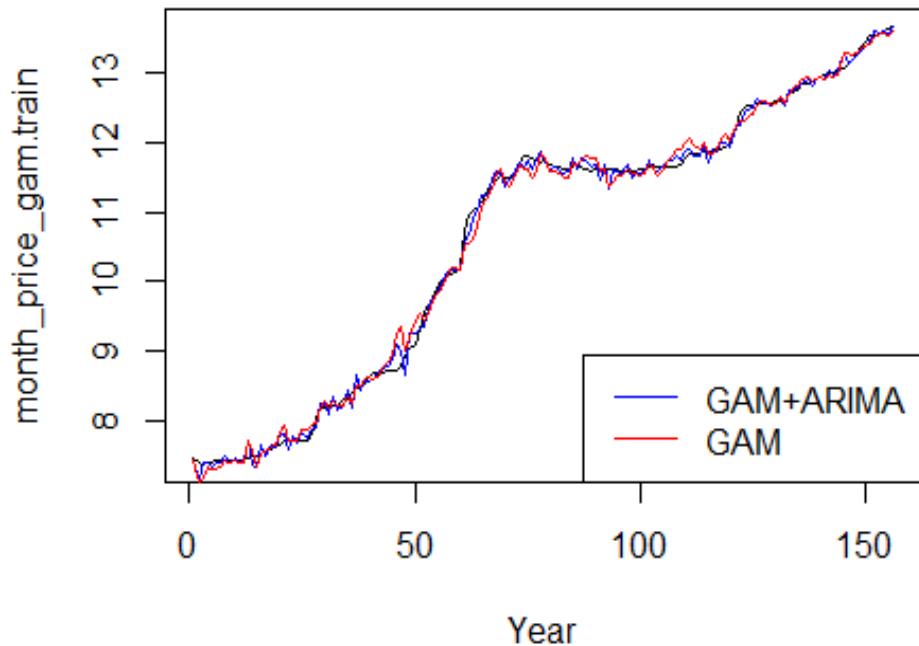
CHOSEN  
VARIABLES

Year  
Month  
 $s(\text{Palm\_oil})$   
Sunflower\_oil  
Rapeseed\_oil  
Natural\_gas\_US  
Wheat\_US\_HRW  
 $s(\text{Wheat\_US\_SRW})$   
 $s(\text{Hourly\_Wage\_USD})$   
 $s(\text{Soybean\_meal})$   
 $s(\text{time})$

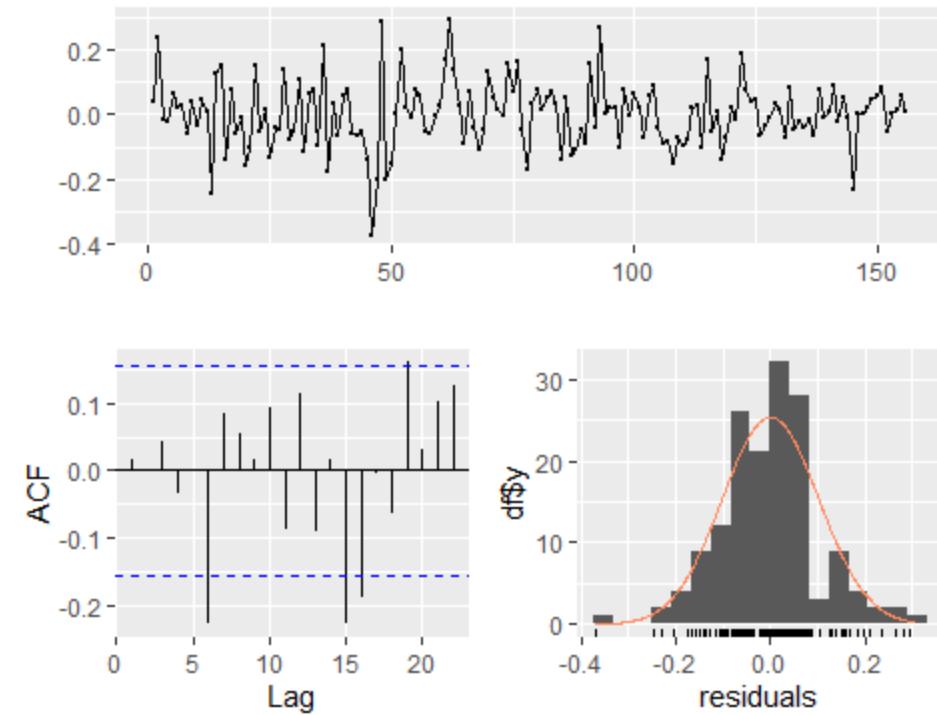


# Auto ARIMA on residuals

ARIMA(3,0,0)

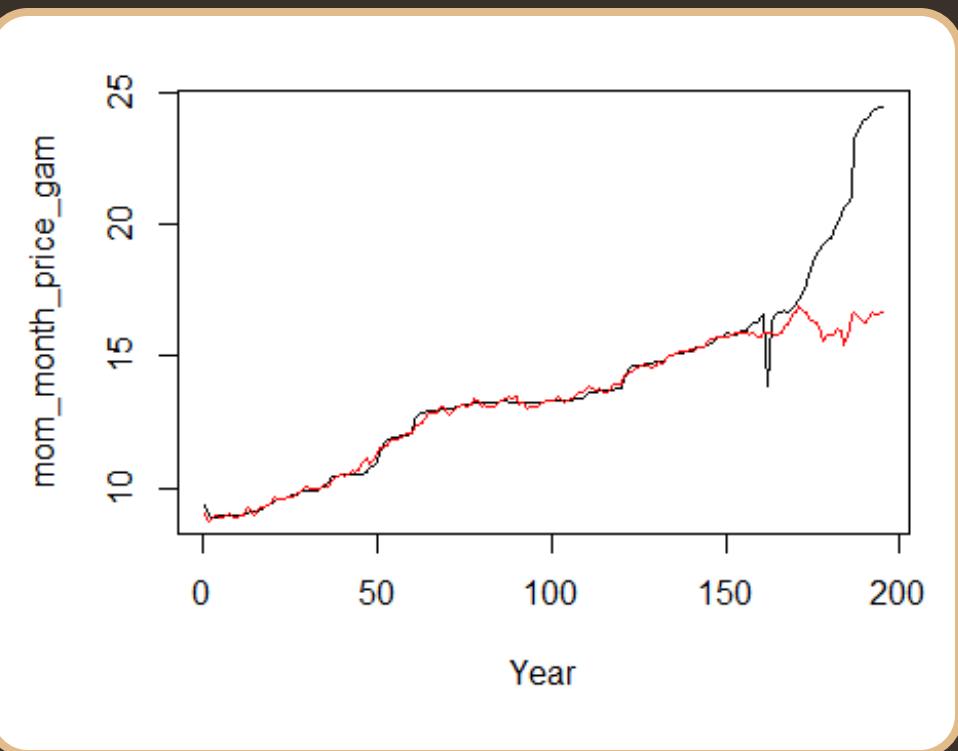


Residuals from ARIMA(3,0,0) with zero mean

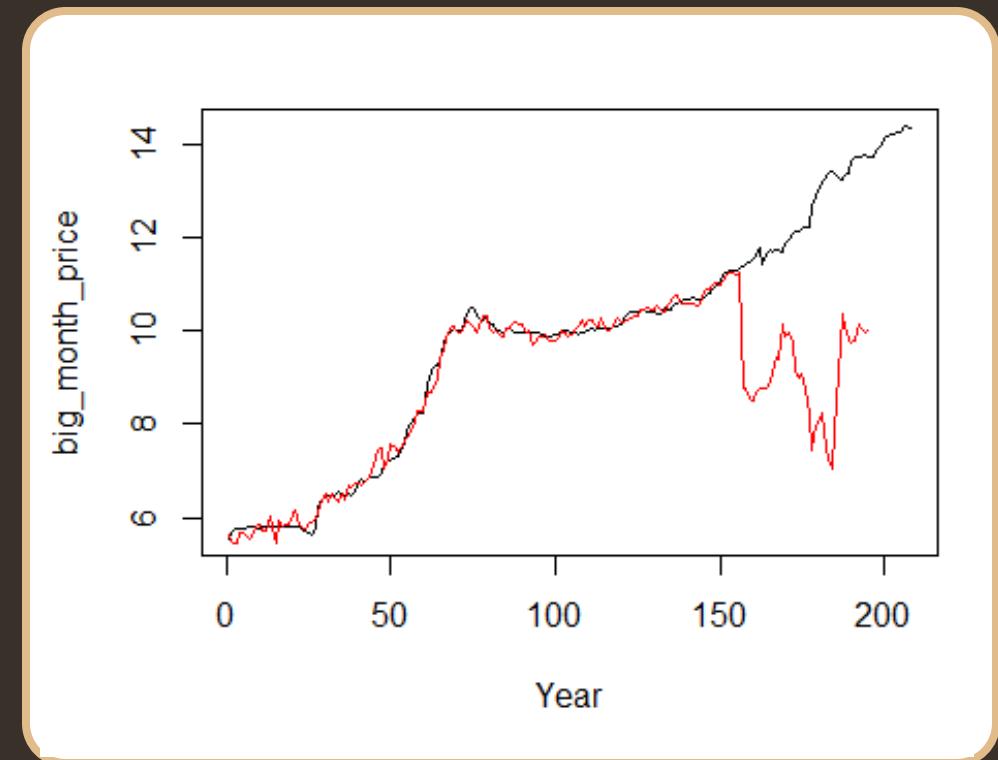


# Stepwise GAM for different store type

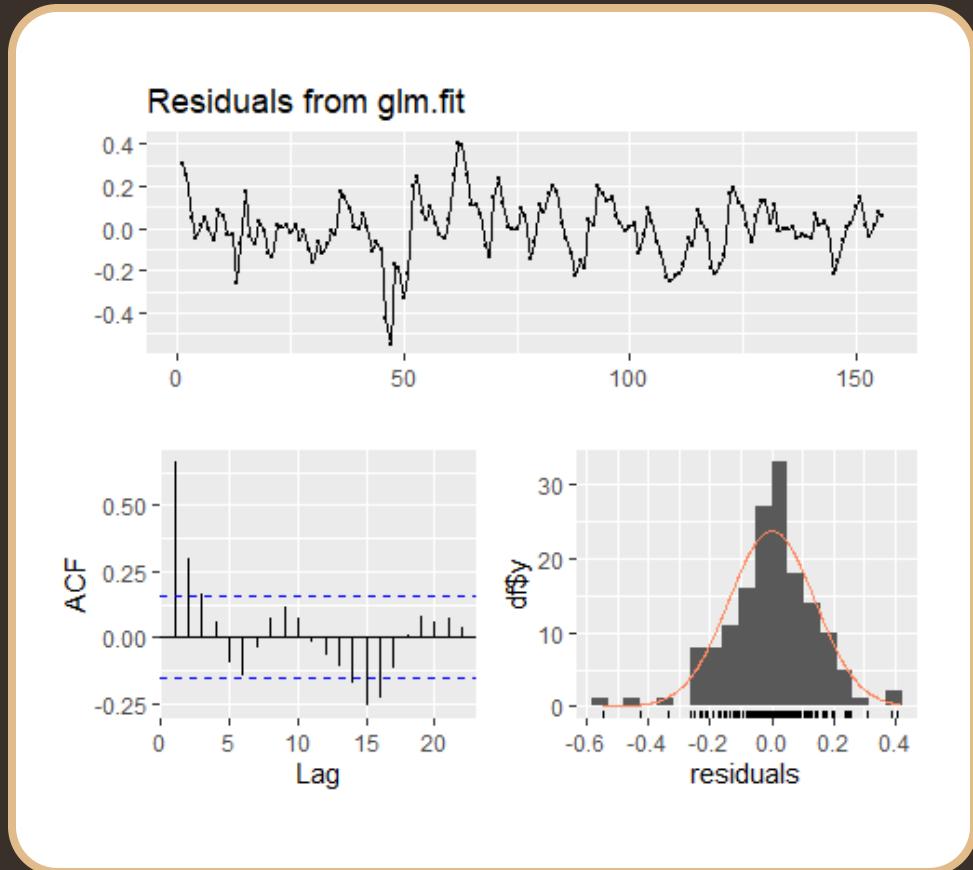
MOM AND POP STORE



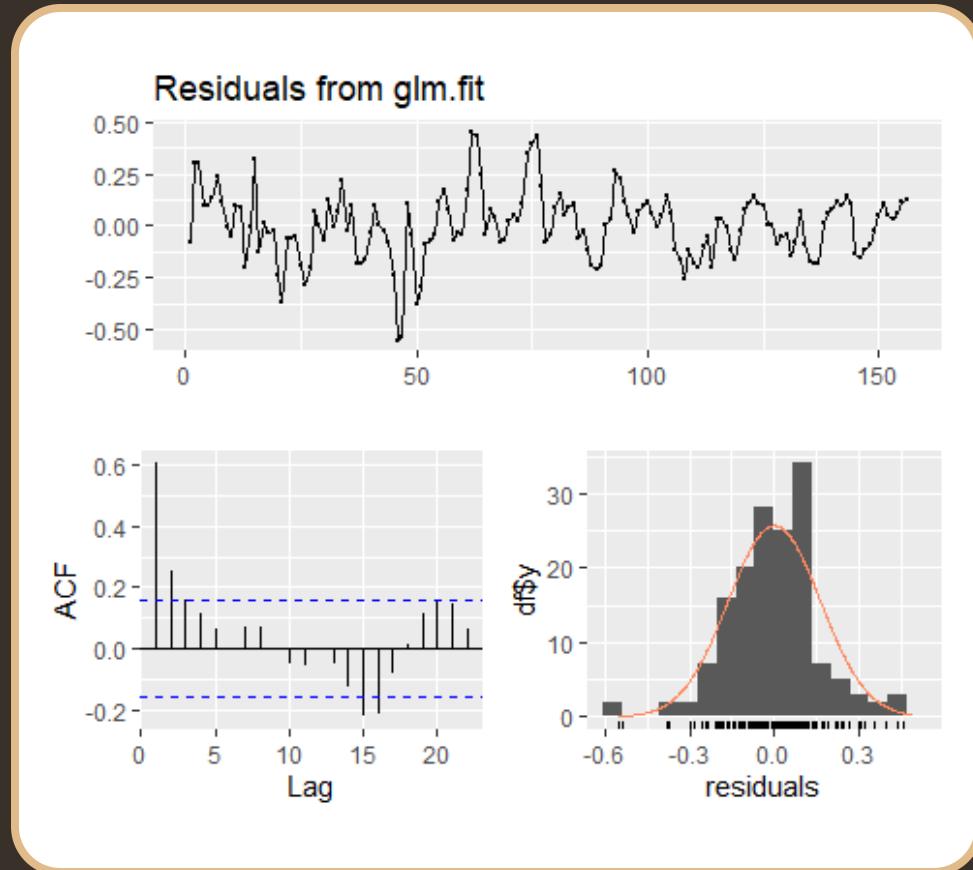
BIG RETAIL STORE



MOM AND POP STORE



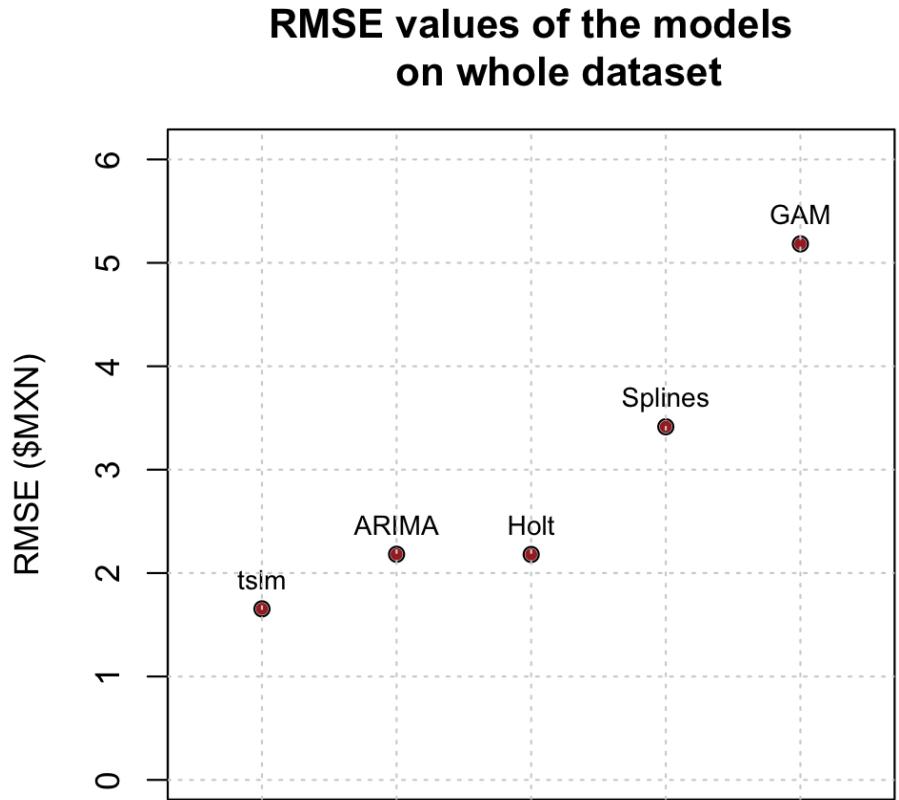
BIG RETAIL STORE



# Conclusions

FINAL THOUGHTS ABOUT THE RESULTS

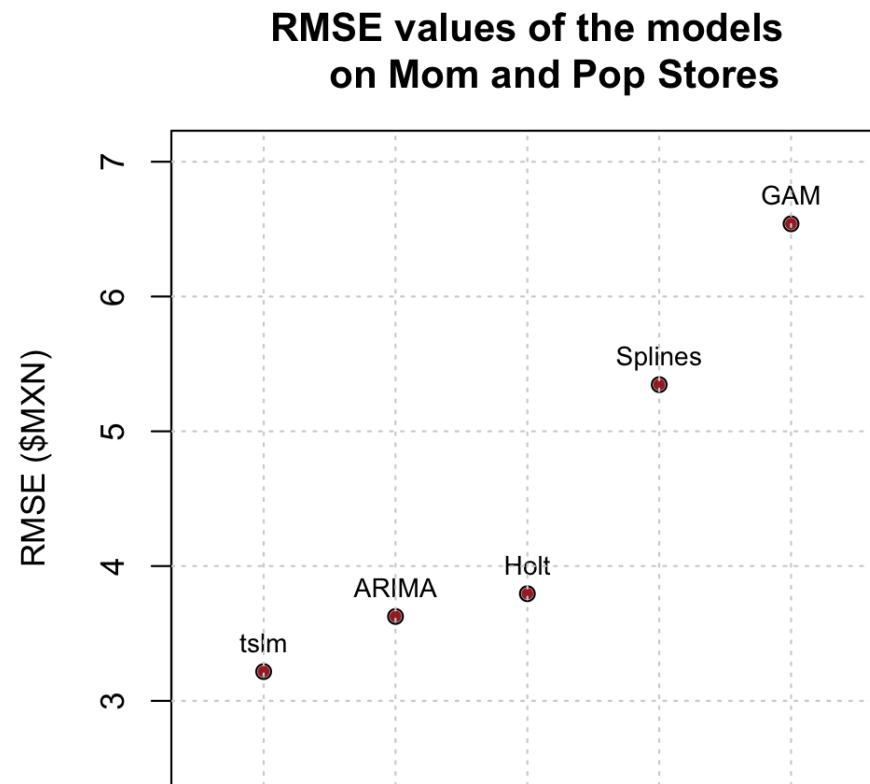
---



## RMSE Comparison

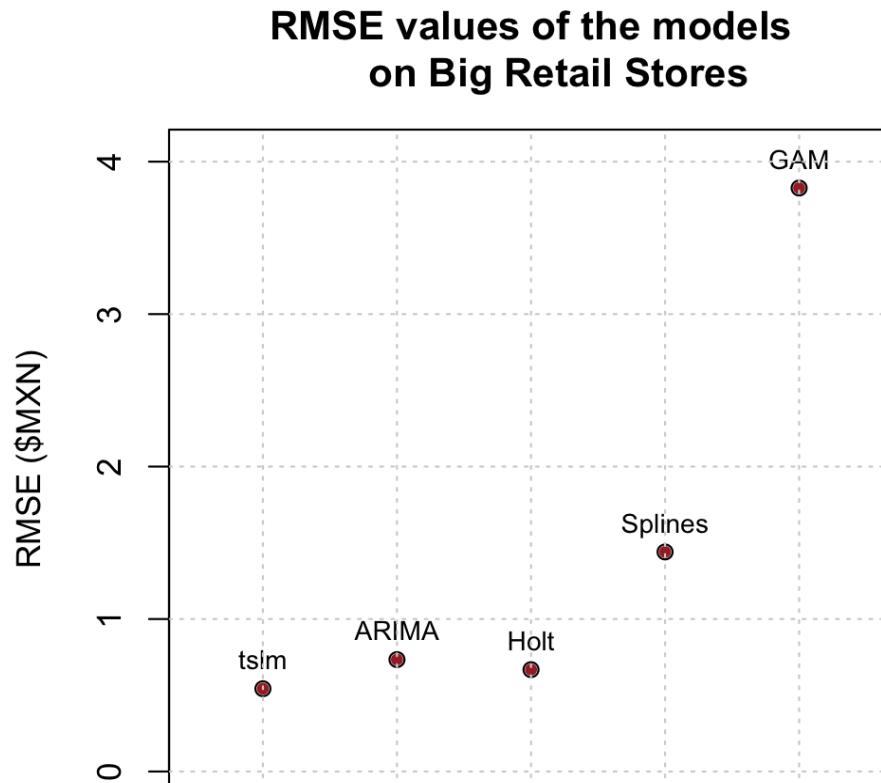
- Tslm has the best RMSE value, GAM the worst.
- Holt and ARIMA are similar

# RMSE Comparison: *Mom and Pop* *Stores*



- Tslm has the best RMSE value, GAM the worst.
- Holt and ARIMA similar

# RMSE Comparison: *Big Retail Stores*



- TSLM, Holt and ARIMA have the best values.
- GAM is the highest.



# Conclusions

- General prices are highly affected by Mom and Pop stores, especially the *drop* and *rapid growth* after 2020.
- TSLM, thanks to the expenses variables, provides better prediction.
- ARIMA and Holt adapt well to historic data.
- Splines adapt to the curve of data, but struggle to predict.
- GAM struggle to capture relationship between relevant expenses and prices.

# Best Model: TSLM

Taking into consideration expenses helps predicting better the general prices, Mom and Pop and Big Retail Store prices.

Arima also good for Big Retail Stores and has white noise residuals.



Thank you  
for your attention.