Forecasting Lightweight Vehicle Sales

Sifan Wu, Jeremy Xu, Kairan Zhong, Felix Wang, Jun Tan

Time Series Analysis and Forecasting

May 24, 2023

THE UNIVERSITY OF CHICAGO

Outline

Introduction

Data Selection

Experimental results and Analysis

Model Selection

Conclusion and Future work

Introduction

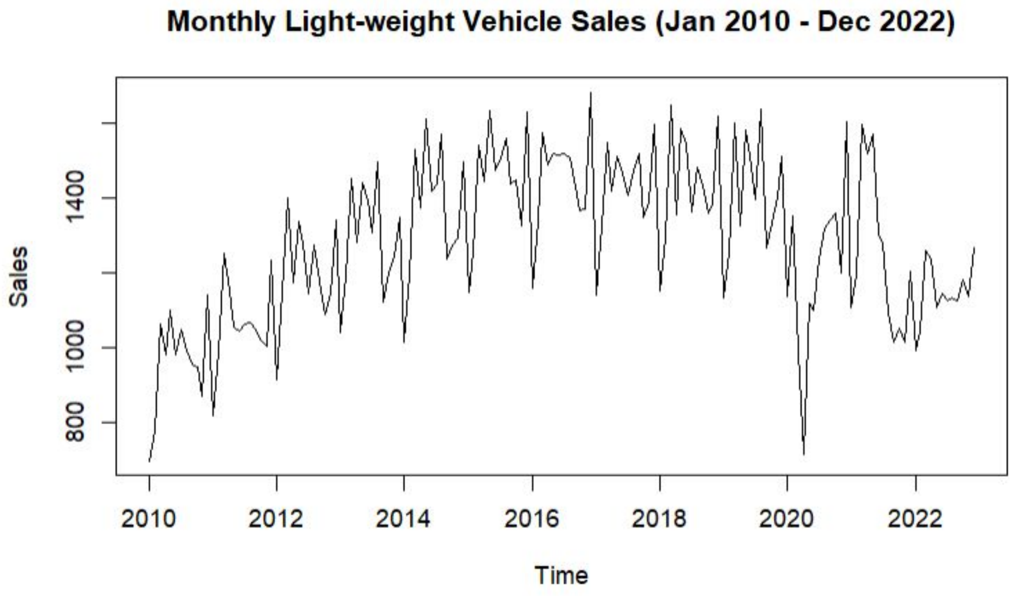
- Analyze the monthly lightweight vehicle sales data from Jan 2010 to Dec 2022 (https://fred.stlouisfed.org/series/LTOTALNSA & https://fred.stlouisfed.org/series/DSPIC96).

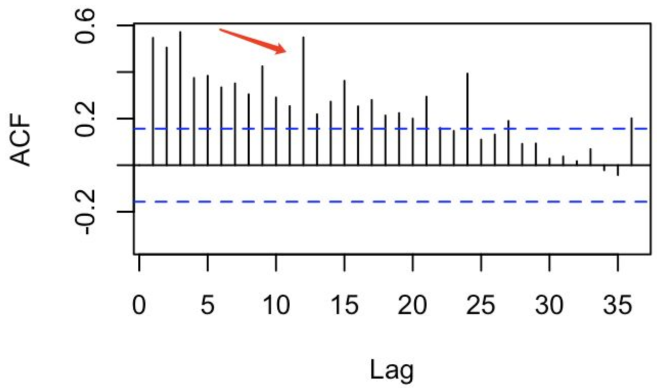
- To forecast the demand for lightweight vehicles over the next 24 months.
- To see if fuel-efficient and environmentally friendly vehicles will have a growing market
  - If so, help manufacturers prepare in advance for the increasing demand in the future
  - If not, encourage policymakers to make better policies of promoting the lightweight vehicles

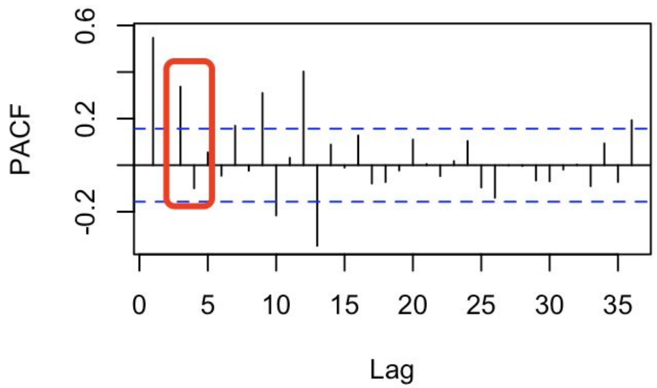
- Model used:
  - Linear regression - Sifan Wu
  - Exponential smoothing - Jeremy Xu
  - ARIMA - Kairan Zhong
  - Regression with ARMA Errors - Felix Wang
  - Intervention Analysis - Jun Tan

Data Selection

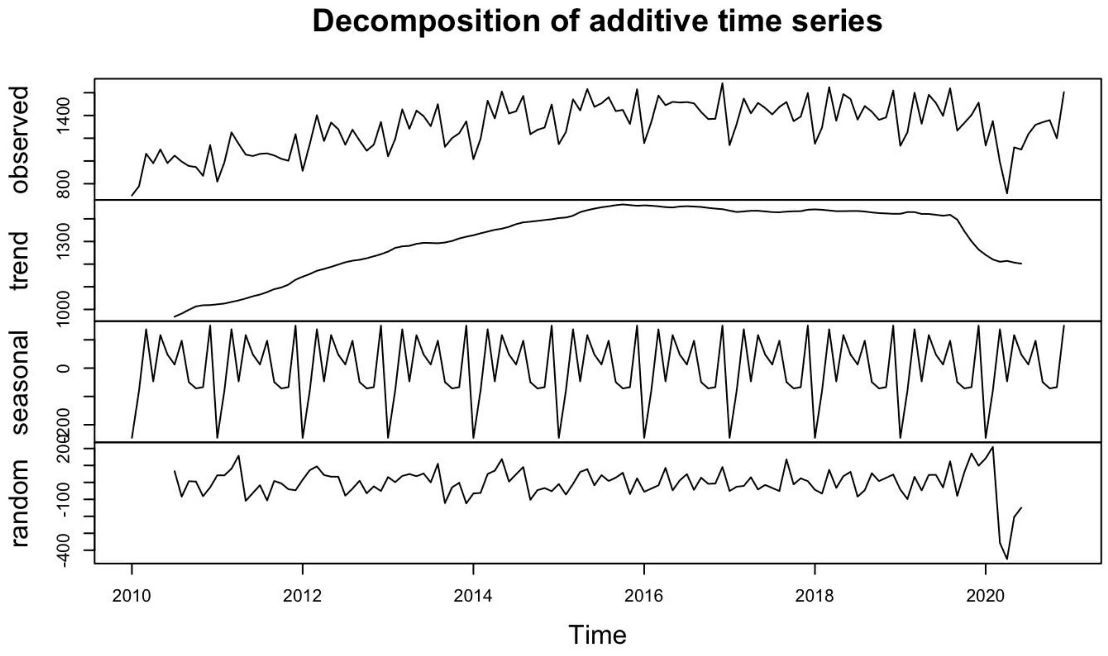
- Two variables
  - Light-weight Vehicle Sales
  - US Disposable Income
- The original data set
  - has seasonality with an upward trend.
  - is Non-stationary.
- Stationarize data
  - seasonal differencing -> first order differencing
- The sample ACF and the sample PACF seem to indicate that an $AR(3)$ model should be specified.
- Train set: 2020/01 - 2020/12   Test set: 2021/01 - 2022/12

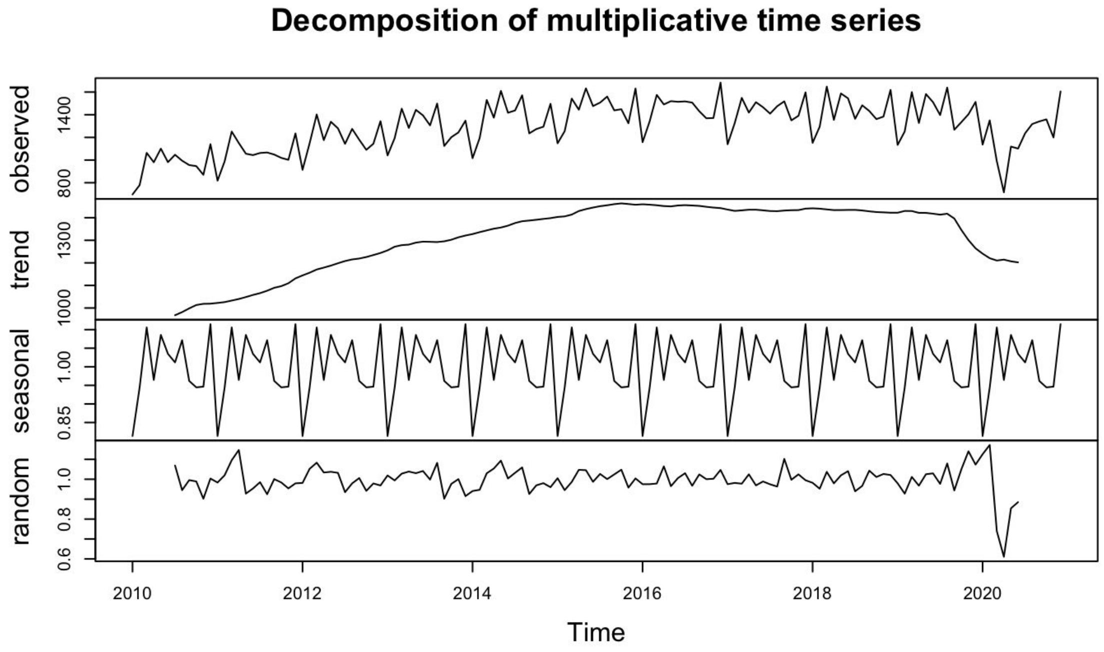






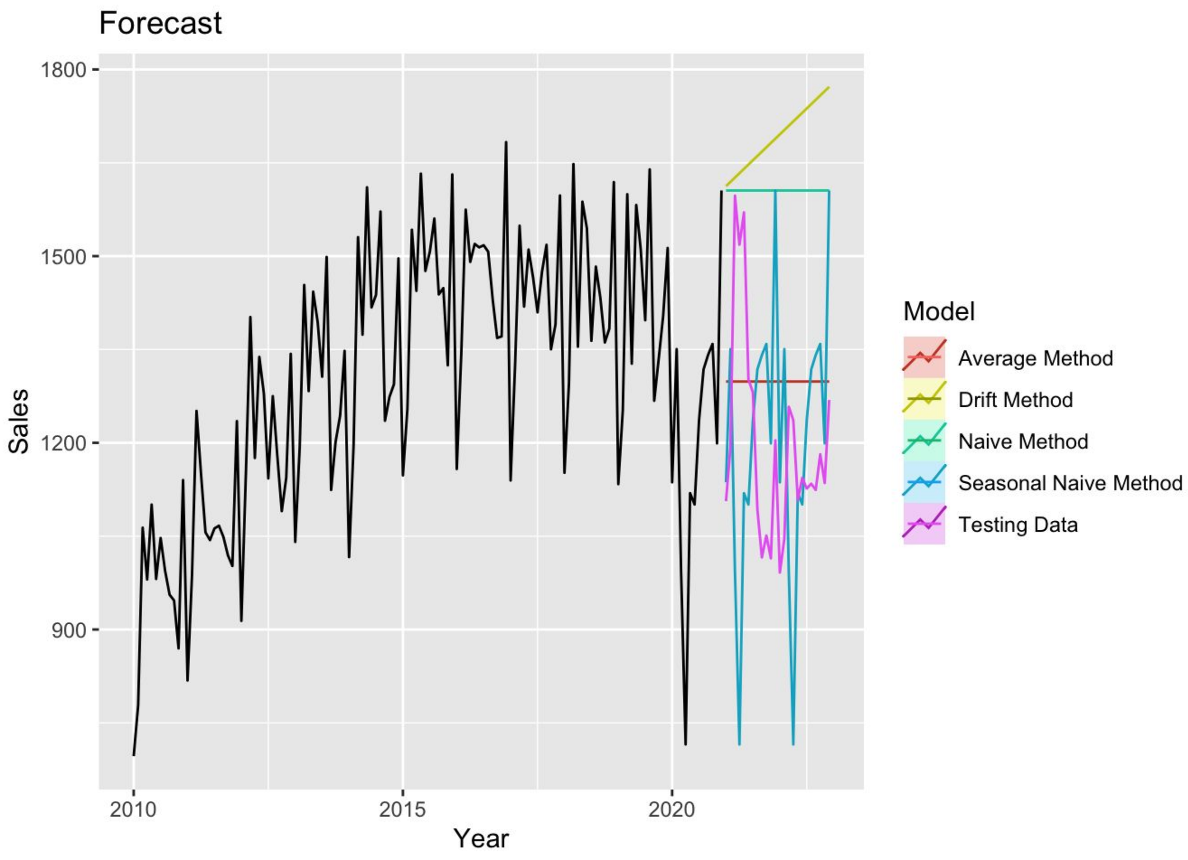
Time Series Decomposition





The remainder component for multiplicative decomposition looks like a white noise, so multiplicative decomposition makes more sense.

Baseline Approaches in Forecasting



	RMSE	MAE	Ljung-Box Test (p-value)
Average Method	192.0058	168.8056	< 0.05
Drift Method	529.8307	496.3407	< 0.05
Naive Method	440.5821	409.6111	< 0.05
Seasonal Naive Method	318.3891	254.9943	< 0.05

Linear Regression

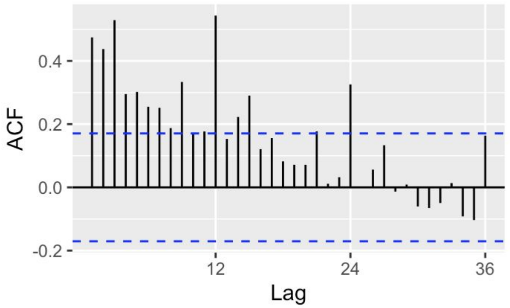
Linear Regression

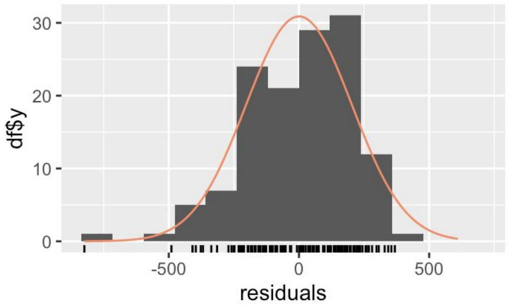
- **Model Fitting**

Light Weight Vehicle Sales ~ US Disposable Income

- **Model Diagnostics**

Residuals are autocorrelated



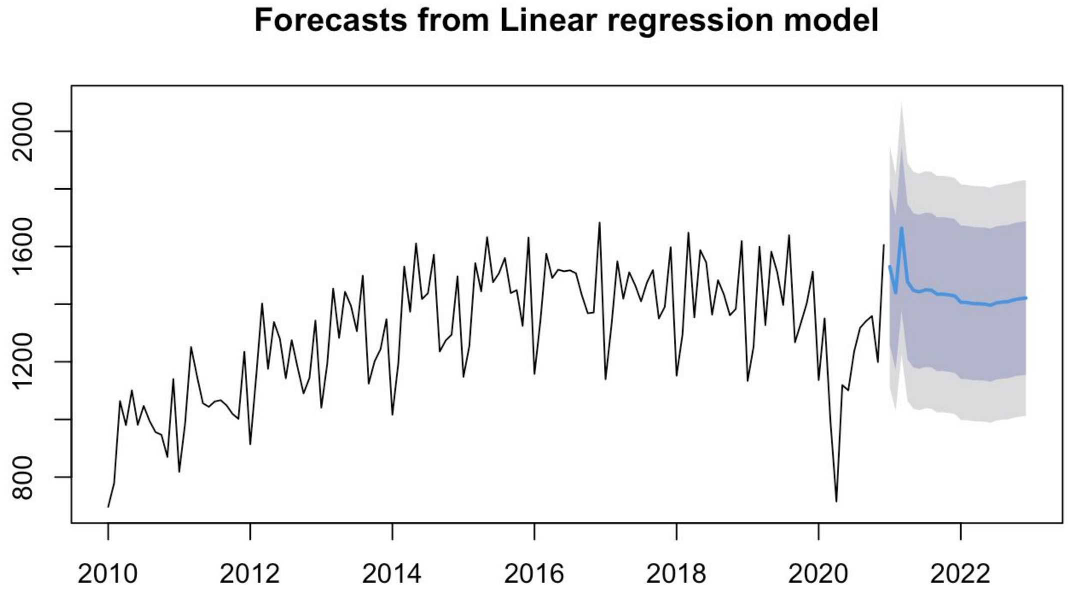


- **Model Evaluation**

Method: Train-test Split

RMSE: 279.3836

MAE: 255.9479




```
data:  Residuals from Linear regression model
LM test = 93.853, df = 24, p-value = 3.29e-10
```

Exponential Smoothing

Holt-Winters Seasonal Method

```r
## Linear trend with additive seasonality
m_add <- hw(ts_train, h=24, seasonal="additive", damped=FALSE)
m_add$model$aicc
```

```
## [1] 1887.252
```

```r
## Linear trend with additive seasonality and damping
m_damped_add <- hw(ts_train, h=24, seasonal="additive", damped=TRUE)
m_damped_add$model$aicc
```

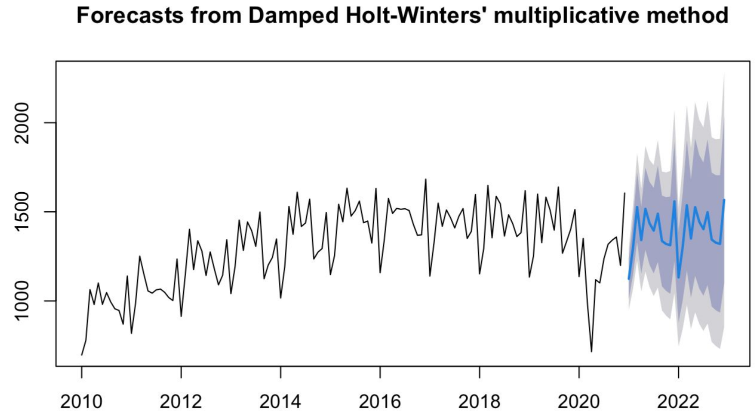
```
## [1] 1887.81
```

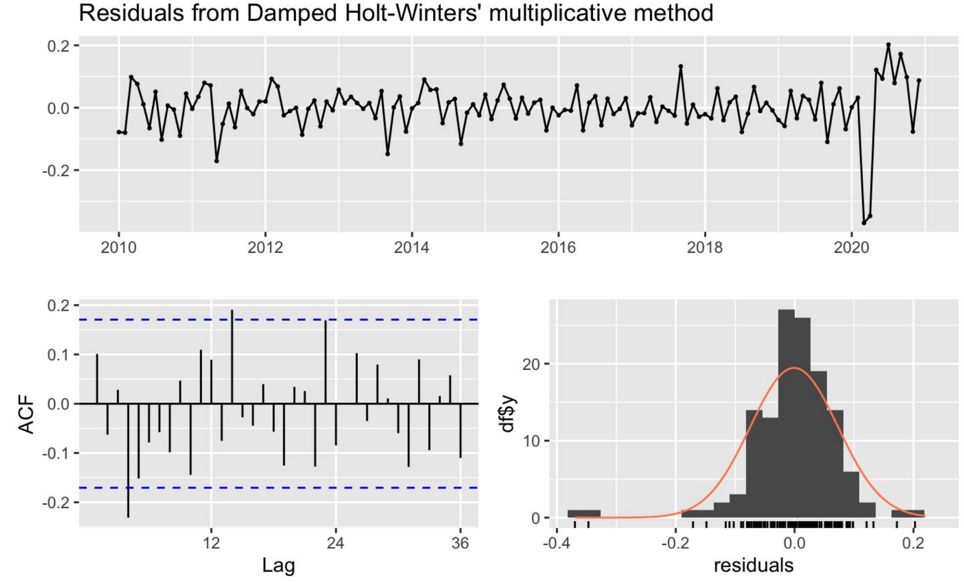
```r
## Linear trend with multiplicative seasonality
m_multi <- hw(ts_train, h=24, seasonal="multiplicative", damped=FALSE)
m_multi$model$aicc
```

```
## [1] 1899.315
```

```r
## Linear trend with multiplicative seasonality and damping
m_damped_multi <- hw(ts_train, h=24, seasonal="multiplicative", damped=TRUE)
m_damped_multi$model$aicc ## best model
```

```
## [1] 1886.972
```

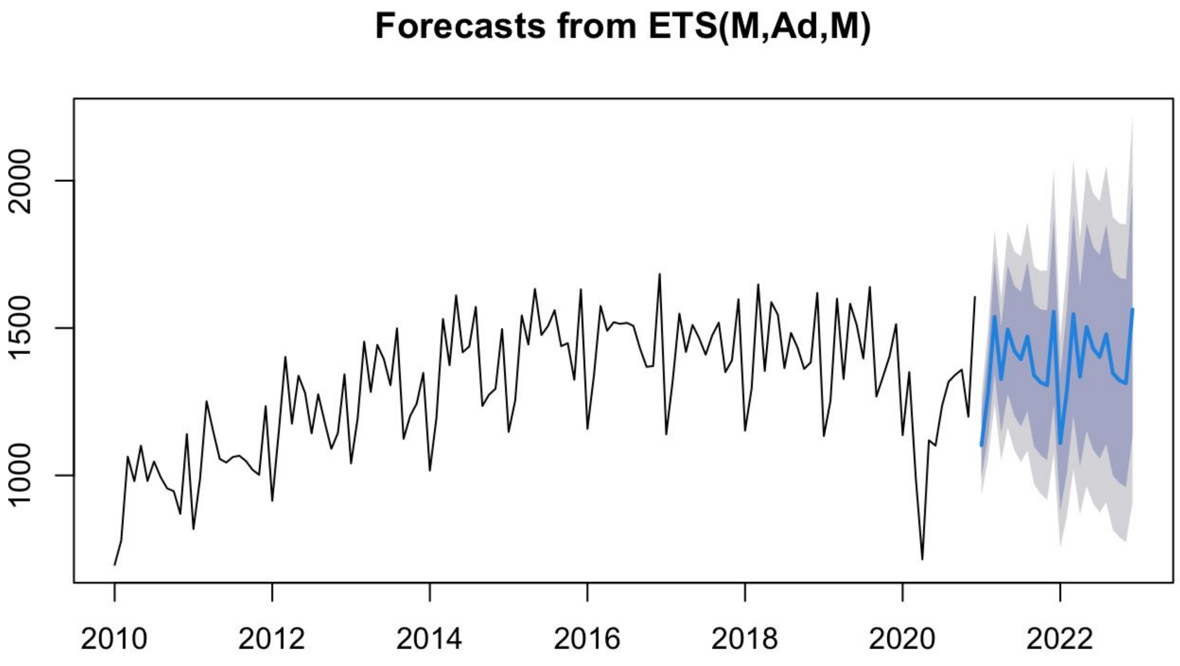


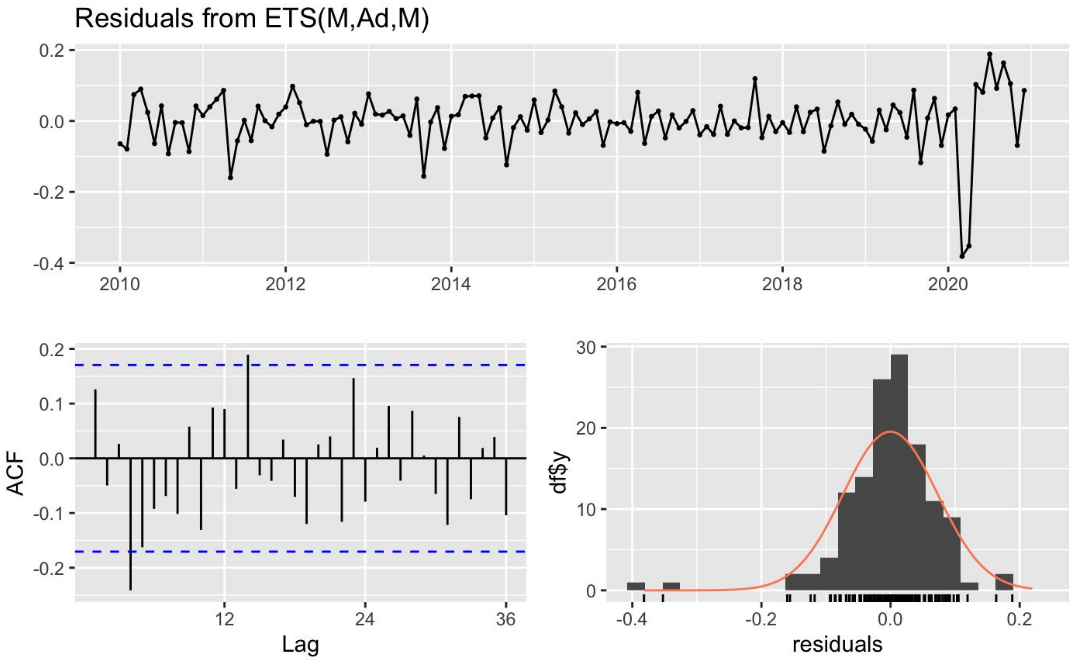



```
##
##  Ljung-Box test
##
## data:  Residuals from Damped Holt-Winters' multiplicative method
## Q* = 40.243, df = 24, p-value = 0.02013
##
## Model df: 0.   Total lags used: 24
```

Exponential Smoothing State Space Model

ETS (Error, Trend, Seasonal)

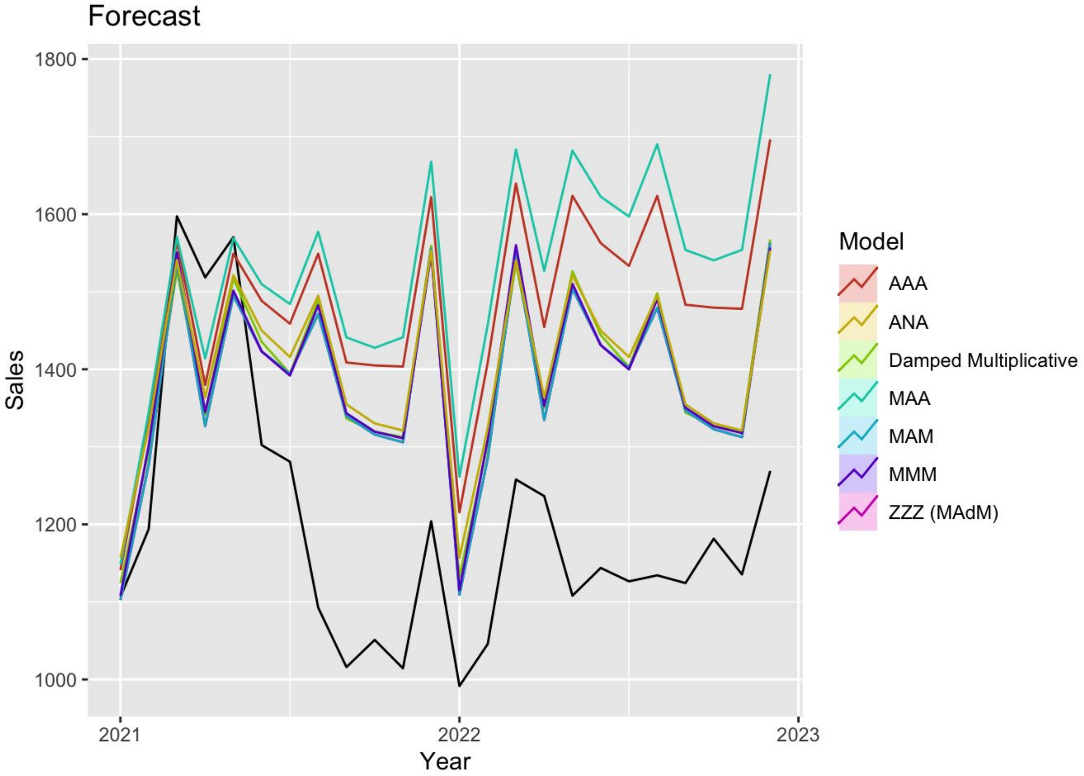




Model with *multiplicative errors, additive damped trend, and multiplicative seasonality*.

```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(M,Ad,M)
## Q* = 39.554, df = 24, p-value = 0.02387
##
## Model df: 0.   Total lags used: 24
```

Comparing Exponential Smoothing Models



The ETS (M, Ad, M) model has the best accuracy.

```
##                  ME     RMSE      MAE       MPE     MAPE      ACF1 Theil's U
## Test set -186.6456 240.8748 214.1647 -16.87944 18.67013 0.5718254  1.840553
```

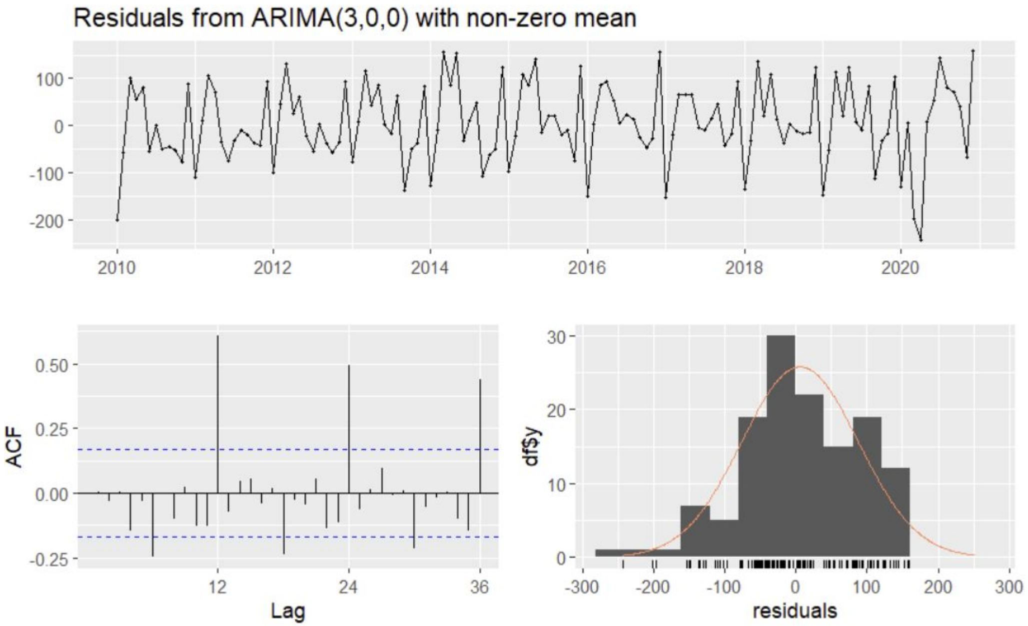
```
## ETS(M,Ad,M)
##
## Call:
##  ets(y = ts_train, model = "ZZZ")
##
##   Smoothing parameters:
##     alpha = 0.5253
##     beta  = 1e-04
##     gamma = 2e-04
##     phi   = 0.9789
##
##   Initial states:
##     l = 914.3187
##     b = 14.1728
##     s = 1.125 0.945 0.9527 0.9705 1.0663 1.0109
##            1.0325 1.0852 0.9637 1.1176 0.9285 0.8021
##
##   sigma:  0.0782
##
##      AIC     AICc      BIC
## 1878.922 1884.975 1930.813
```

ARIMA

ARIMA

ARIMA(3,0,3) with no seasonality

```
Series: train_ts
ARIMA(3,0,0) with non-zero mean
Box Cox transformation: lambda= 0.910537

Coefficients:
         ar1     ar2     ar3      mean
      0.2451  0.2065  0.4112  731.5146
s.e.  0.0812  0.0832  0.0826   48.1274

sigma^2 = 6927:  log likelihood = -769.6
AIC=1549.2   AICc=1549.68   BIC=1563.62

Training set error measures:
                      ME     RMSE      MAE        MPE     MAPE     MASE         ACF1
Training set 11.9926 155.0954 122.8004 -0.5851951 10.01626 1.277135 0.0004511321
```



AICc = 1549.68 BIC = 1563.62

Significant spikes at lag 12, 24… indicates the residual is not white noise and has strong seasonality

THE UNIVERSITY OF CHICAGO

SARIMA

Model1: ARIMA(0,1,1)(2,1,2)[12]

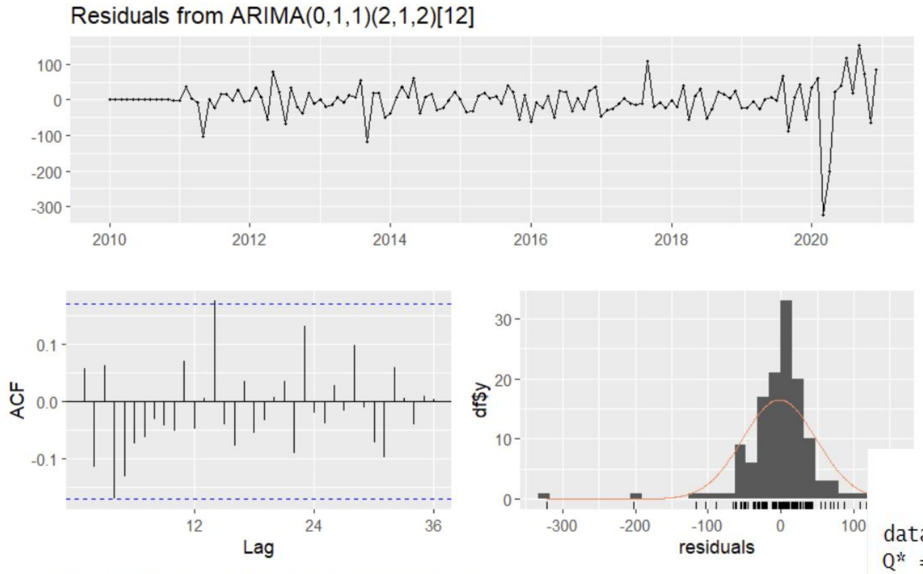
```
Series: train_ts
ARIMA(0,1,1)(2,1,2)[12]
Box Cox transformation: lambda= 0.9105539

Coefficients:
          ma1     sar1      sar2      sma1   sma2
      -0.4653  0.7983  -0.6690  -1.1568  0.659
s.e.   0.1139  0.2330   0.2045   0.2650  0.301

sigma^2 = 2947:  log likelihood = -646.64
AIC=1305.27   AICc=1306.02   BIC=1321.95

Training set error measures:
                    ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
Training set -4.134603 95.38499 57.09142 -0.7680598 4.632912 0.5937556 0.04946321
```

AICc = 1306.02 BIC = 1321.95

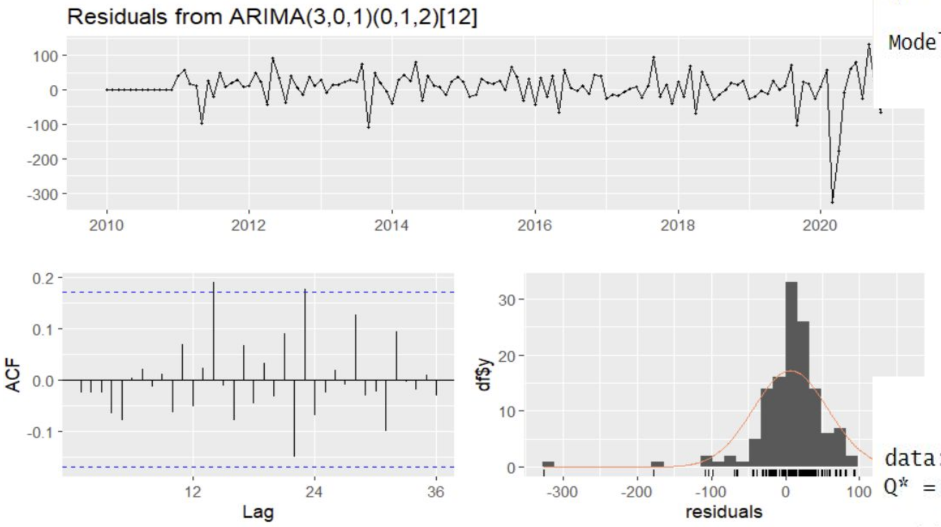
Model2: ARIMA(3,0,1)(0,1,2)[12] (with d =0)

```
Series: train_ts
ARIMA(3,0,1)(0,1,2)[12]
Box Cox transformation: lambda= 0.9105539

Coefficients:
          ar1     ar2     ar3     ma1     sma1     sma2
      -0.1314  0.5076  0.3503  0.7102  -0.3332  -0.2898
s.e.   0.3431  0.2625  0.0897  0.3726   0.1633   0.1453

sigma^2 = 2934:  log likelihood = -649.32
AIC=1312.63   AICc=1313.63   BIC=1332.14

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set 12.80432 95.27845 61.44751 0.580297 4.963525 0.6390592 -0.03469308
```

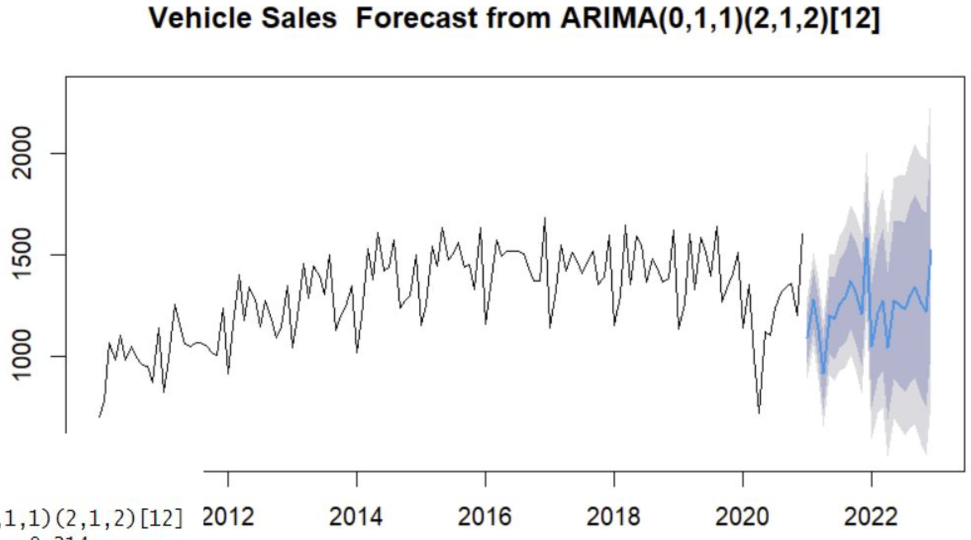
AICc = 1313.63 BIC = 1332.14

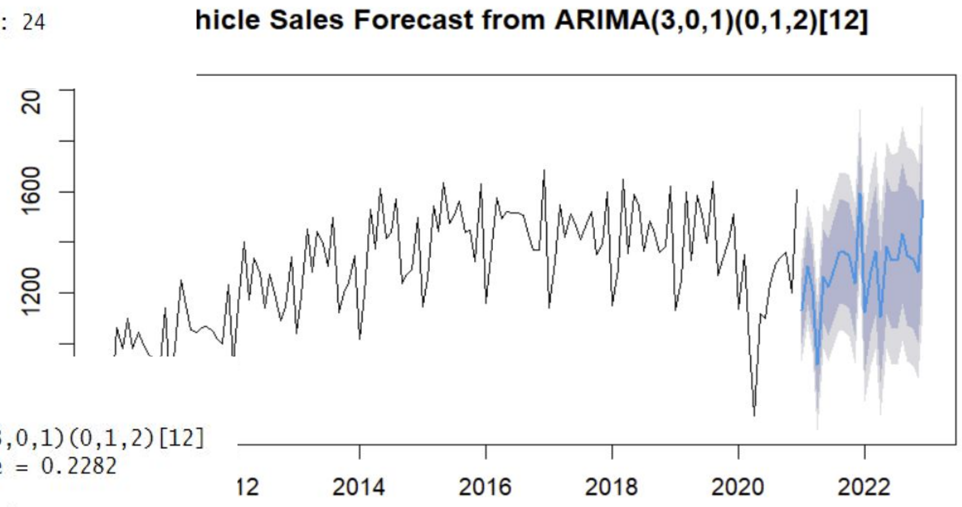



```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(2,1,2)[12]
Q* = 23.548, df = 19, p-value = 0.214

Model df: 5.   Total lags used: 24
```



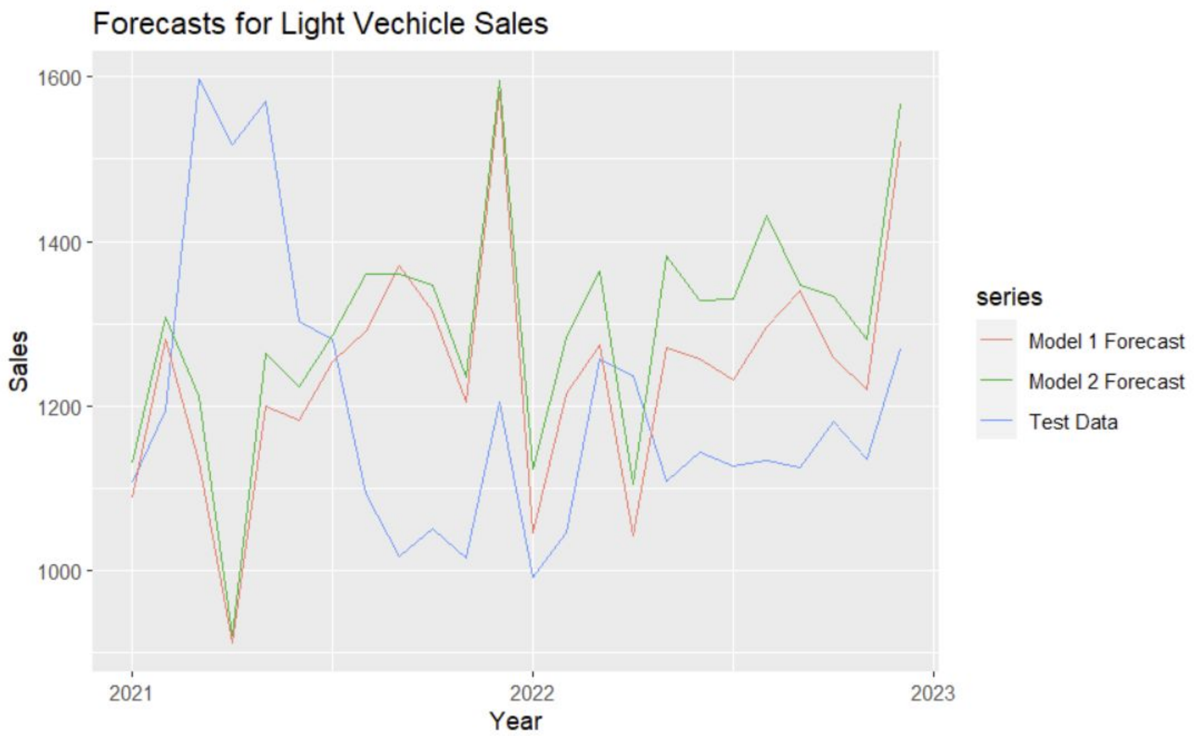

```
        Ljung-Box test

data:  Residuals from ARIMA(3,0,1)(0,1,2)[12]
Q* = 22.087, df = 18, p-value = 0.2282

Model df: 6.   Total lags used: 24
```





THE UNIVERSITY OF CHICAGO

SARIMA



Final model:

ARIMA(0,1,1)(2,1,2)[12]

Model <chr>	MAPE <dbl>	RMSE <dbl>
Model 1	15.88123	243.9537
Model 2	18.79237	260.0702

2 rows

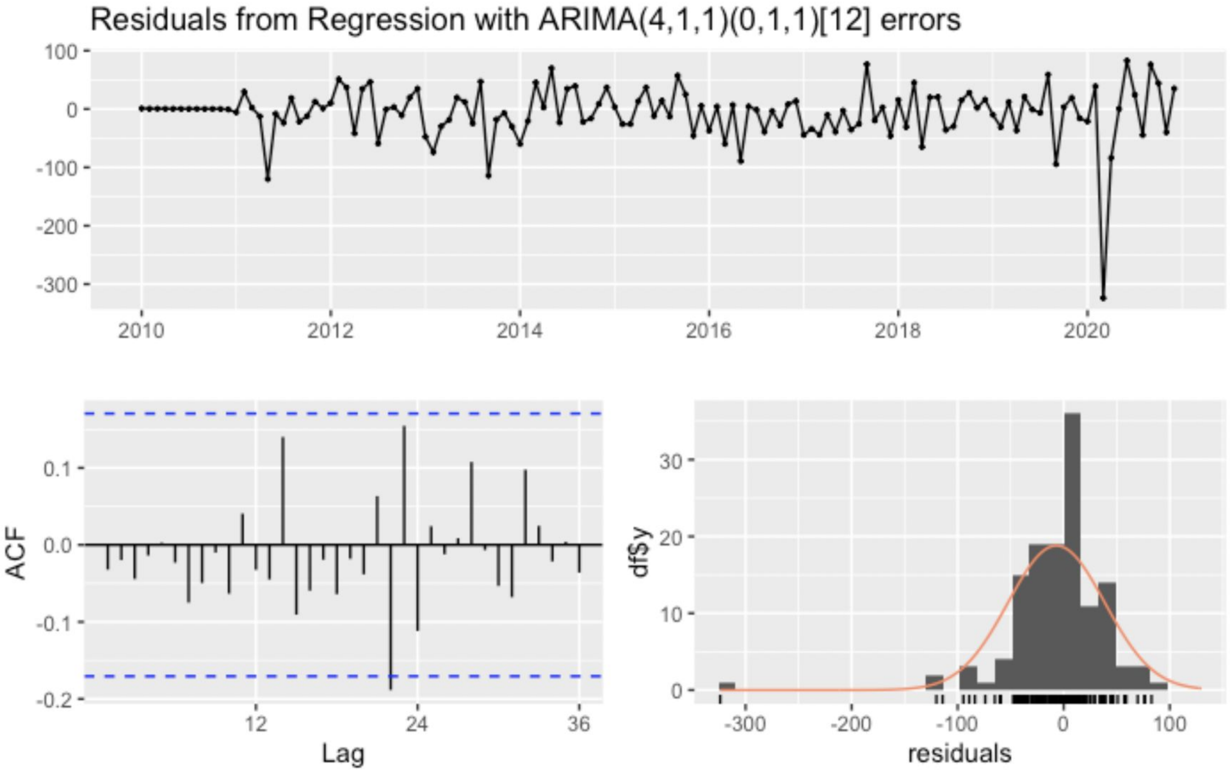
Regression with ARMA Errors

Regression with ARMA Errors

- **Model specification and Estimation: ARIMA(4,1,1)(0,1,1)[12] errors**

```
Series: train_vehicle_sales
Regression with ARIMA(4,1,1)(0,1,1)[12] errors
Box Cox transformation: lambda= 0.910537

Coefficients:
         ar1     ar2     ar3      ar4      ma1     sma1     xreg
      0.1426  0.1300  0.2447  -0.3146  -0.8344  -0.6794  -0.0978
s.e.  0.1199  0.1048  0.0978   0.0920   0.0924   0.1198   0.0219

sigma^2 = 2486:  log likelihood = -634.95
AIC=1285.91   AICc=1287.22   BIC=1308.14

Training set error measures:
                    ME     RMSE      MAE       MPE   MAPE      MASE
Training set -13.00368 87.19715 55.40375 -1.345904 4.4077 0.5762037 -0
```

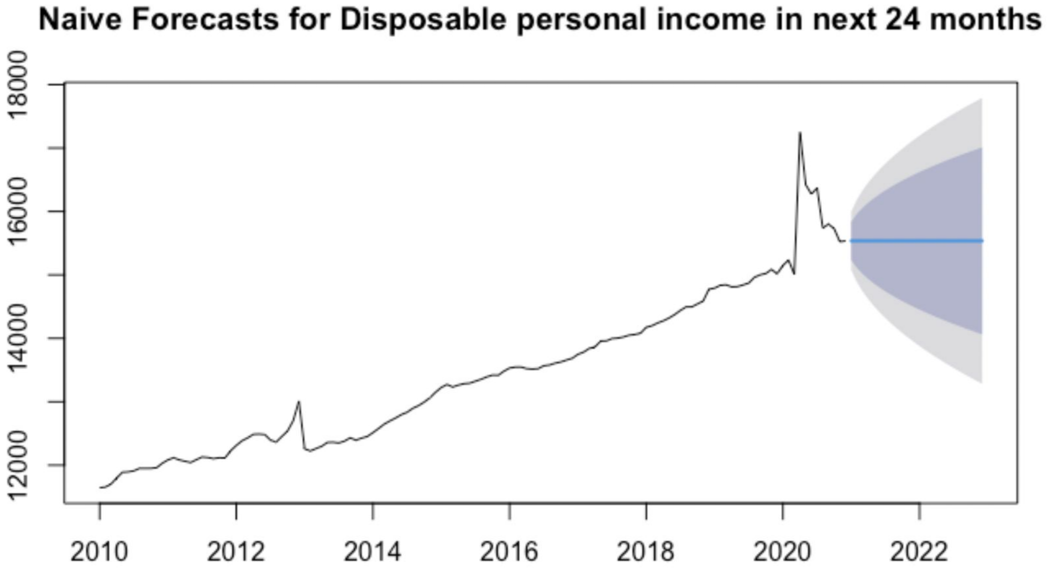
- **Model Diagnostics: checkresiduals()**




```
        Ljung-Box test

data:  Residuals from Regression with ARIMA(4,1,1)(0,1,1)[12] errors
Q* = 20.993, df = 18, p-value = 0.2797

Model df: 6.   Total lags used: 24
```

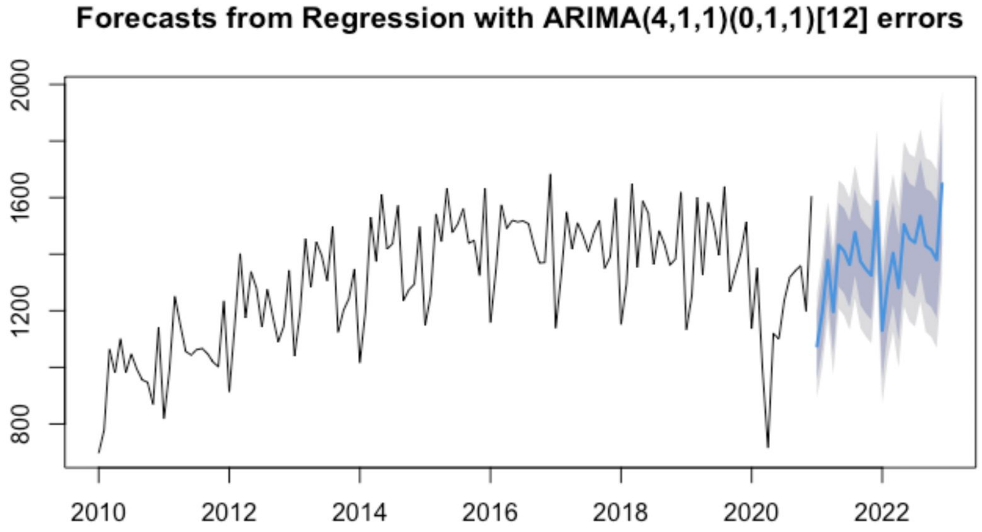
Regression with ARMA Errors

- **Forecast**



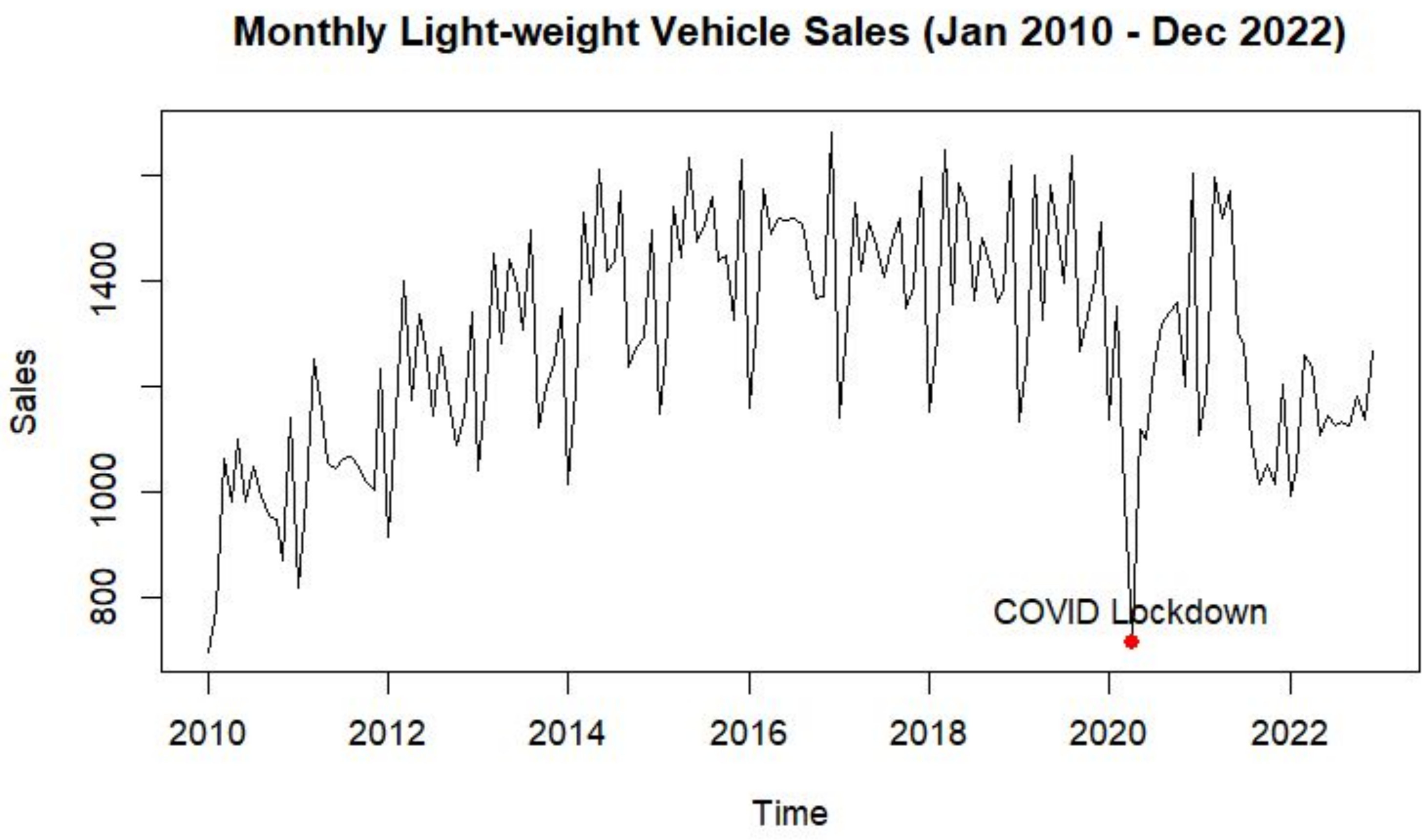
- **Model Evaluation:**
  - Train: 132 (2010.1-2020.12)
  - Test: 24 (2021.1-2022.12)
  - Around 80%, 20%

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-13.00368	87.19715	55.40375	-1.345904	4.40770	0.5762037	-0.03867808	NA
Test set	-182.76525	271.45624	241.92899	-16.988012	20.86397	2.5160819	0.68527567	2.049608

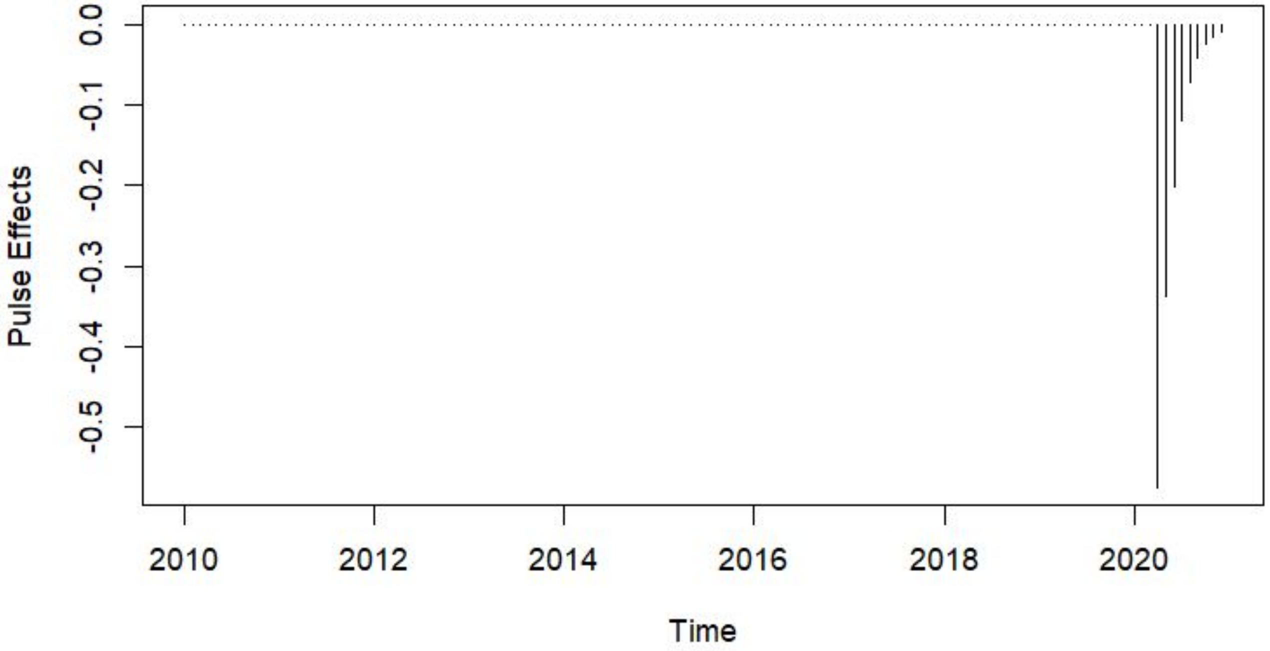


Intervention Analysis

Intervention Specification - Pulse Function



Intervention Analysis

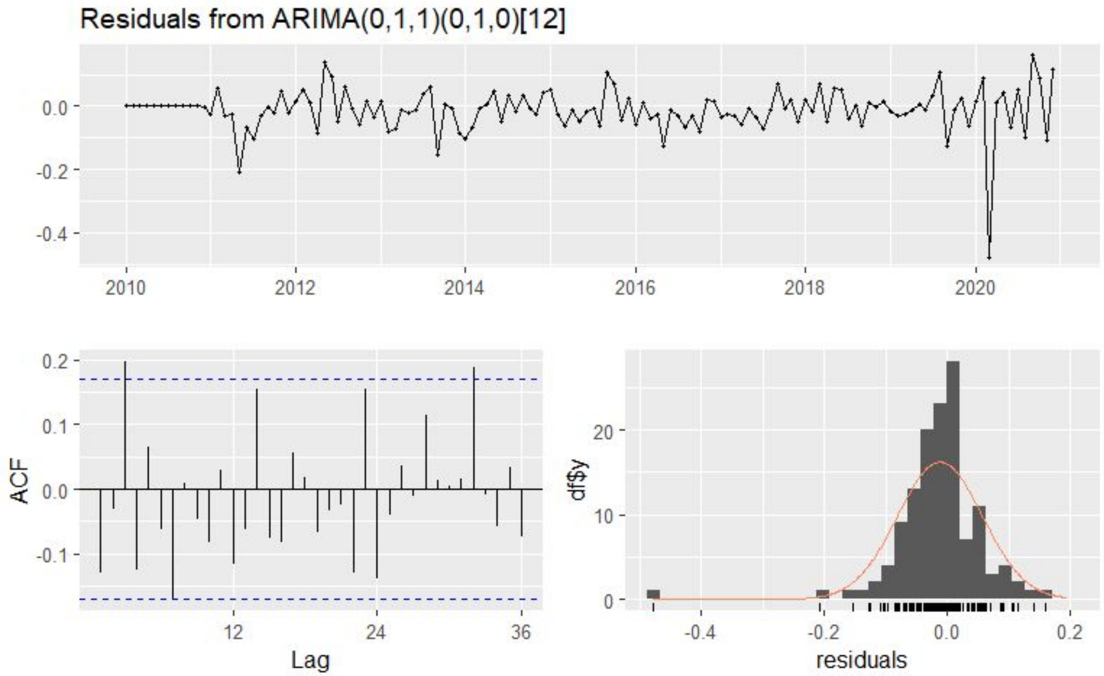


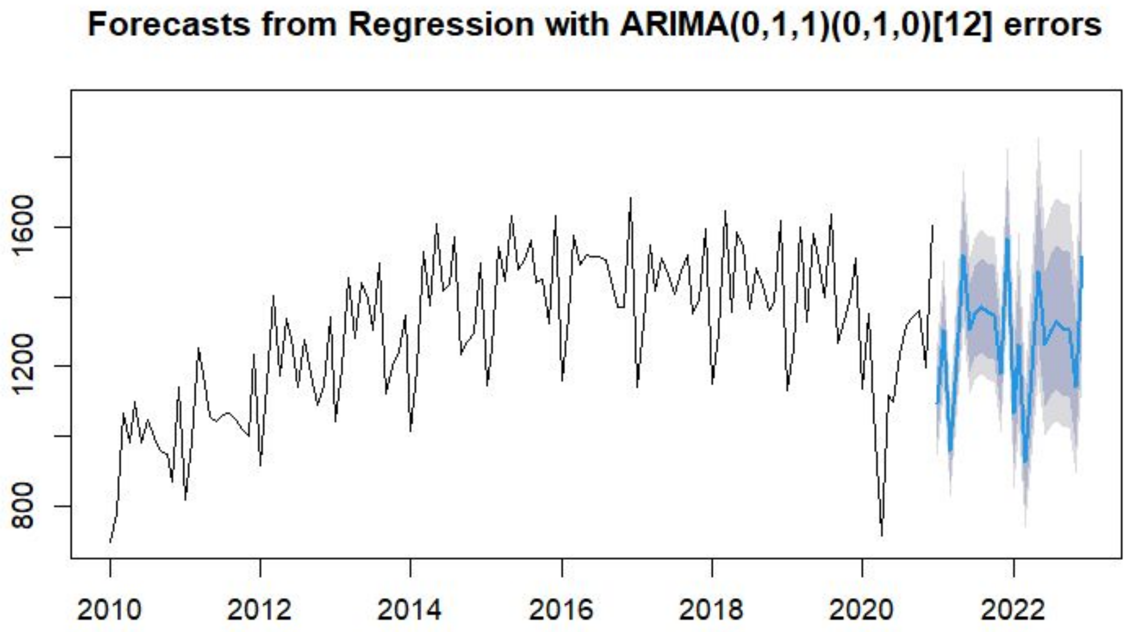

```
Call:
arimax(x = log(train), order = c(0, 1, 1), seasonal = list(order = c(0, 1, 0),
    period = 12), method = "ML", xtransf = data.frame(I2020, I2020), transfer = list(c(0,
    0), c(1, 0)))

Coefficients:
          ma1  I2020-MA0  I2020.1-AR1  I2020.1-MA0
      -0.8897    -0.0086       0.5953      -0.5658
s.e.   0.0391     0.1810       0.1035       0.1663

sigma^2 estimated as 0.005411:  log likelihood = 140.91,  aic = -273.82

Training set error measures:
                      ME       RMSE        MAE        MPE     MAPE      MASE       ACF1
Training set -0.01220034 0.06987645 0.04486797 -0.1744702 0.627302 0.3579166 -0.1304569
```

- Based on the pre intervention data, an ARIMA(0, 1, 1) × (0, 1, 0)[12] model was tentatively specified for the unperturbed process.
- The fitted model estimates that the the COVID-related activity restrictions reduced light-weight vehicle sales by 57.4% and sales k months later was lowered by $(1 - \exp(-0.5658 * 0.5953^k)) * 100\%$.

THE UNIVERSITY OF CHICAGO

Intervention Analysis - Residuals and Forecasts






```
        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(0,1,0)[12]
Q* = 36.216, df = 23, p-value = 0.03921

Model df: 1.   Total lags used: 24
```

Conclusion and Future Work

Model Selection: Average Method

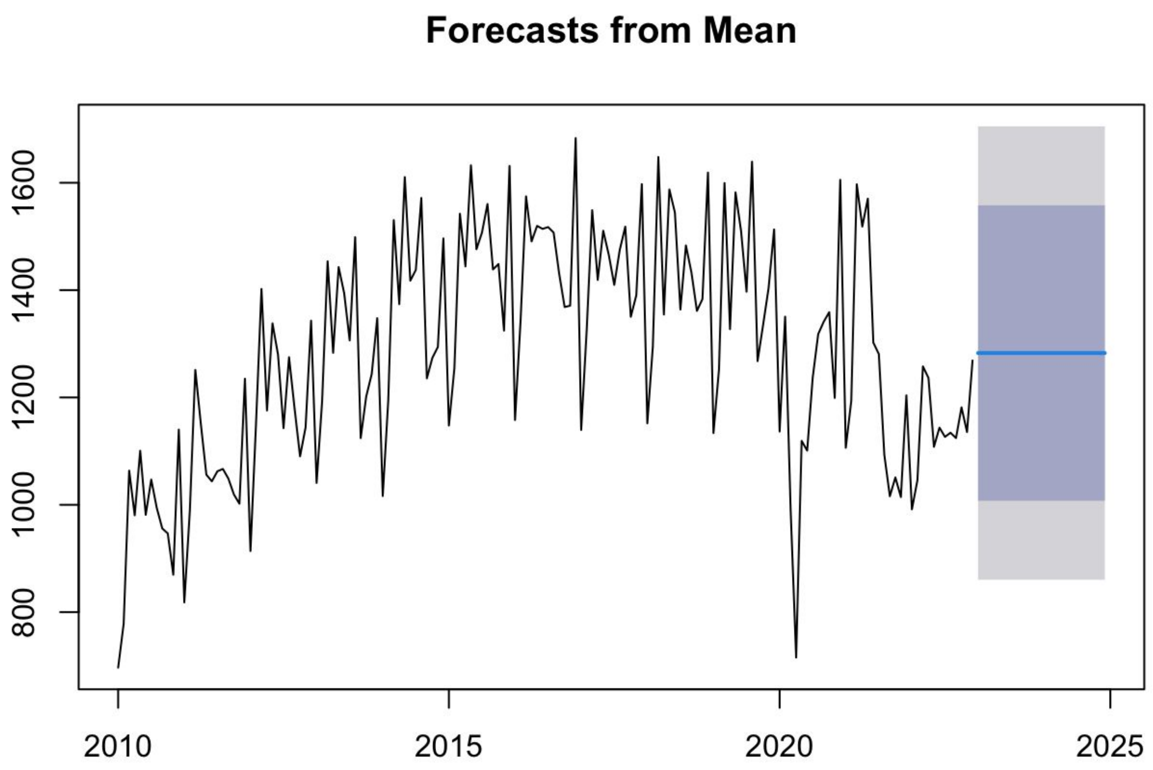
Baseline

	RMSE	MAE	Ljung-Box Test (p-value)
Average Method	192.0058	168.8056	< 0.05
Drift Method	529.8307	496.3407	< 0.05
Naive Method	440.5821	409.6111	< 0.05
Seasonal Naive Method	318.3891	254.9943	< 0.05

Our Models

	RMSE	MAE	Ljung-Box Test (p-value)
Linear Regression	279.3836	255.9479	< 0.05
Exponential Smoothing	240.8748	214.1647	< 0.05
ARIMA	243.95	195.2654	> 0.05
Regression with ARMA Errors	271.4562	241.9290	> 0.05
Intervention Analysis	245.6292	196.0529	< 0.05

Prediction with Average Method



- Forecasting Lightweight Vehicle Sales
- Forecast period: Jan 2023 to Dec 2024

##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Jan 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Feb 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Mar 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Apr 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## May 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Jun 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Jul 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Aug 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Sep 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Oct 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Nov 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Dec 2023	1282.735	1007.528	1557.942	860.3377	1705.132
## Jan 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Feb 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Mar 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Apr 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## May 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Jun 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Jul 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Aug 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Sep 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Oct 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Nov 2024	1282.735	1007.528	1557.942	860.3377	1705.132
## Dec 2024	1282.735	1007.528	1557.942	860.3377	1705.132