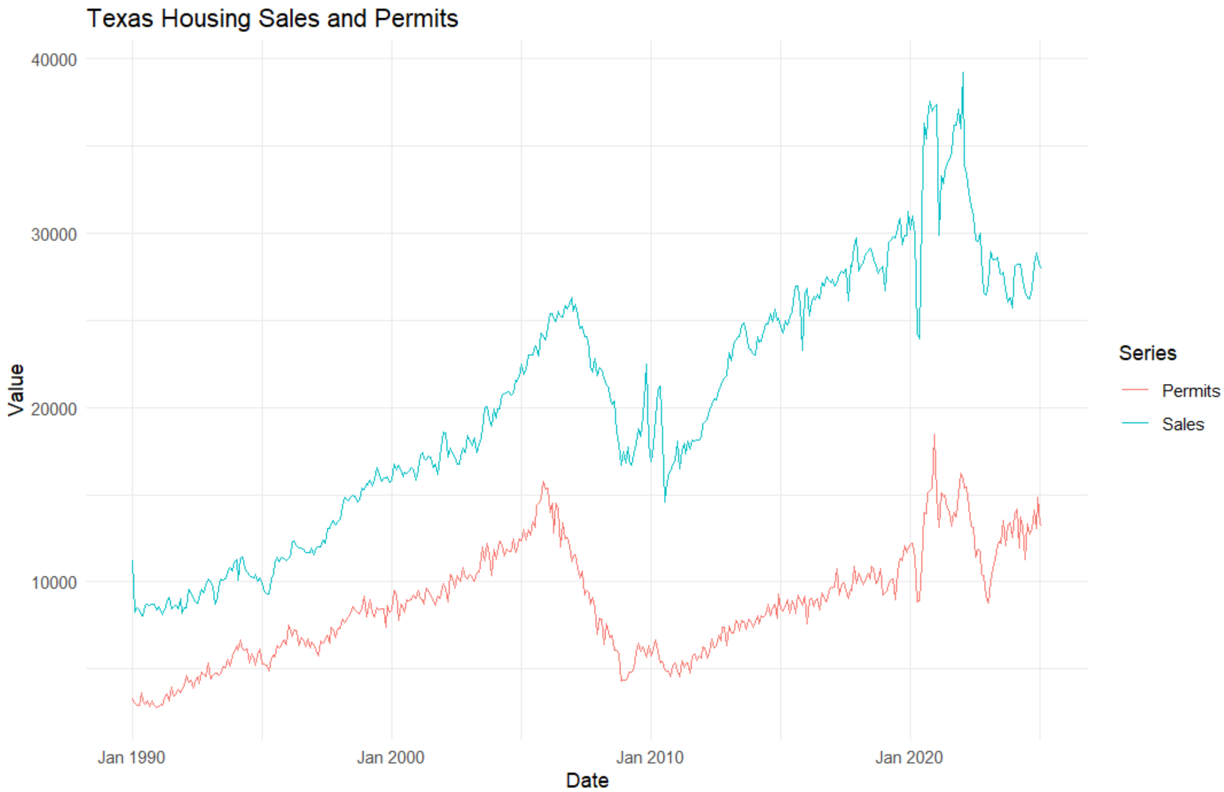
Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

Q1 :



After importing and plotting the data for Texas Housing Sales and Permits from January 1990 to February 2025, I observed :

- Both series exhibit strong upward trends, from the beginning of the series to the financial crisis and from 2009 to the COVID-19 shock in 2020.
- There are breaks in the trend, in 2008 and 2020, which met the financial and COVID-19 crisis.
- These are non-stationary series.

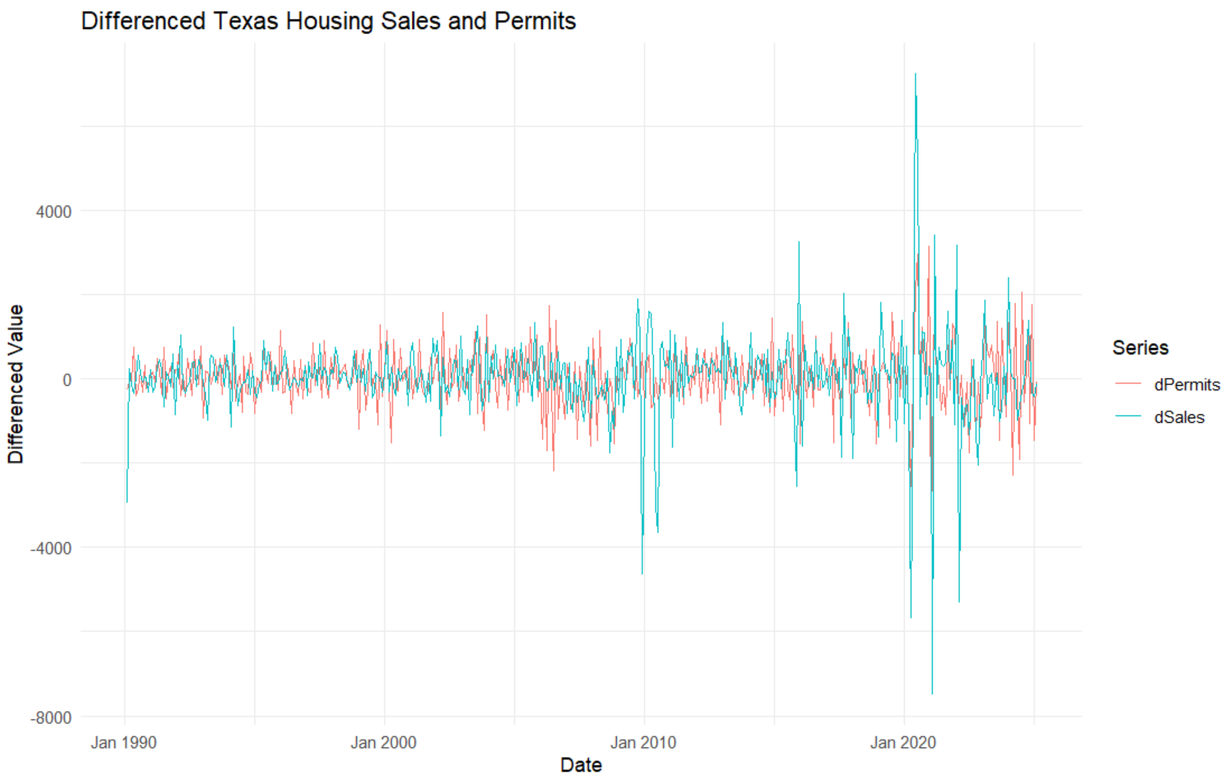
Q2: a)

```
> ndiffs_sales
[1] 1
> ndiffs_permits
[1] 1
```

Both the Sales and Permits series require 1 difference to achieve stationarity.

Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

Q2: b)



The differenced series now fluctuates around zero, and the upward trend seen in the level series is gone. However, some spikes remain.

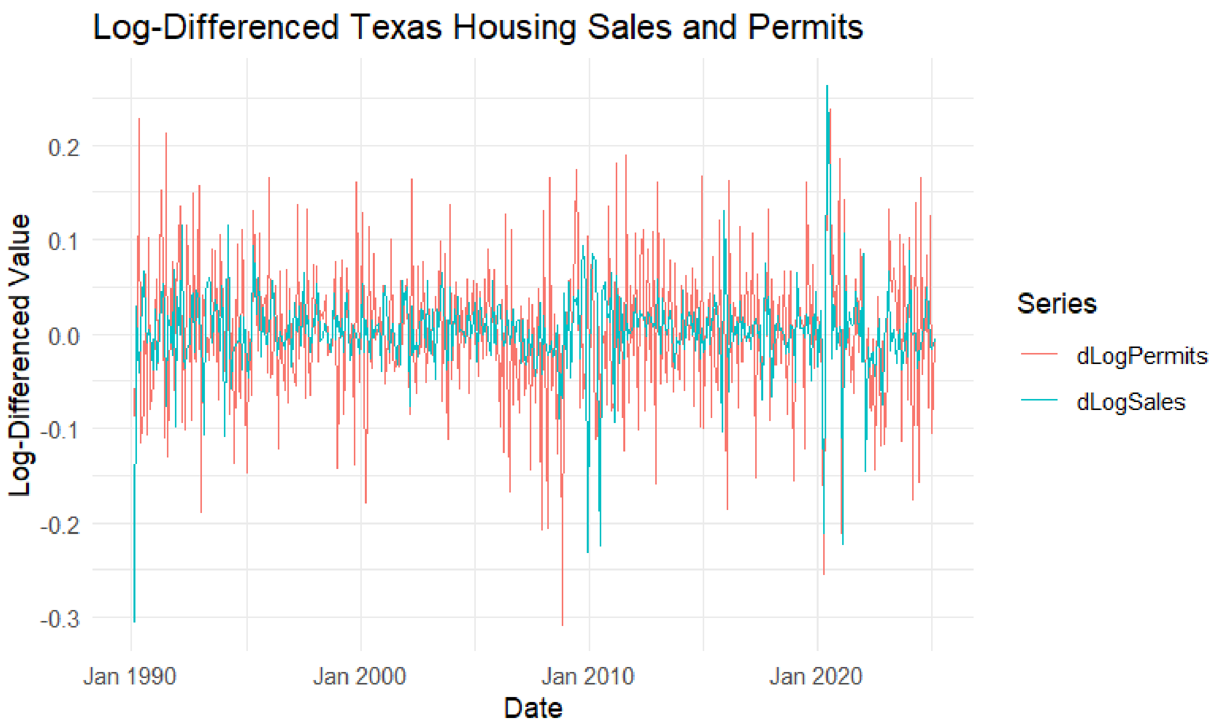
Q3 (a)

```
> ndiffs(log(df$Sales))
[1] 1
> ndiffs(log(df$Permits))
[1] 1
```

Still need 1 difference after logging → same as level series.

Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

Q3 (b)



Variance appears more stable _ we can look at the scale_ after log-differencing — especially for Sales.

Q3 (c)

Yes, compared to the plain differenced series, the log-differenced series shows reduced and more stable variance, particularly for Sales, which had large fluctuations in the raw differenced form.

Knowing the fact that logging helps normalize the growth rate _ minimizes the scale_ making the series better suited for time series modeling like VAR.

Q4

```
> lag_selection$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
    12      4      3     12
```

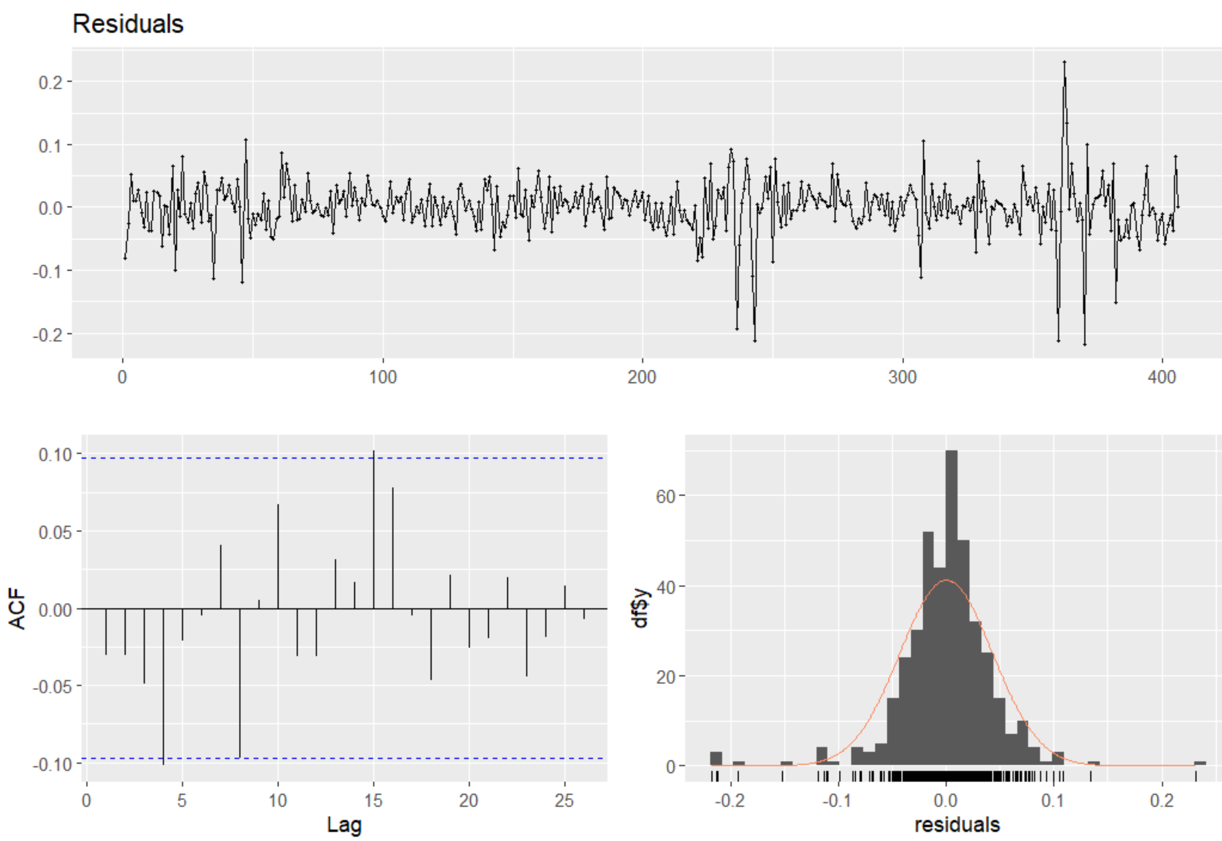
Based on the Schwarz Criterion, the optimal lag length is **3**. This is the number of lags we should use in estimating the VAR model later.

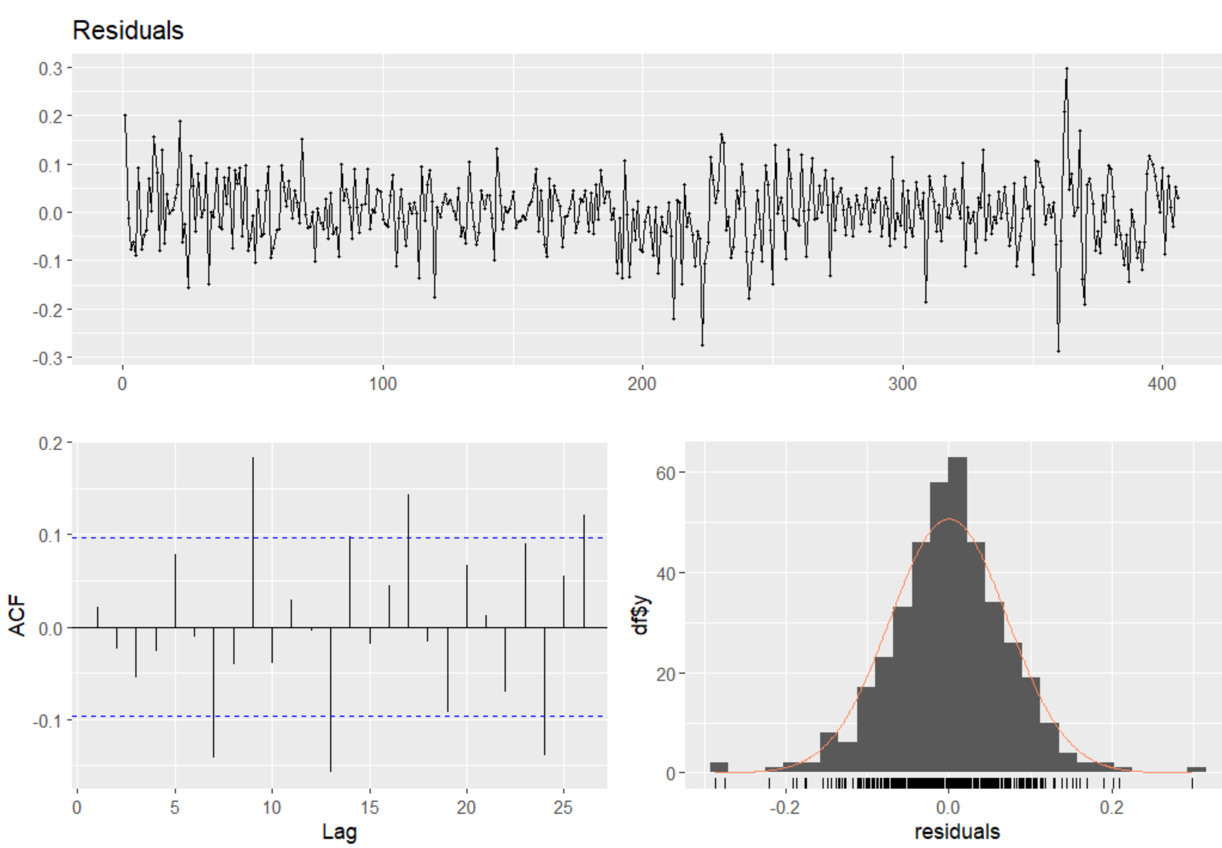
Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

Q5

```
> summary(var_model)

VAR Estimation Results:
=========================
Endogenous variables: dLogSales, dLogPermits
Deterministic variables: const
Sample size: 406
Log Likelihood: 1222.614
Roots of the characteristic polynomial:
0.7187 0.7187 0.5742 0.5742 0.4253 0.417
Call:
VAR(y = train_df, p = 3, type = "const")


Estimation results for equation dLogSales:
==========================================
dLogSales = dLogSales.l1 + dLogPermits.l1 + dLogSales.l2 + dLogPermits.l2 + dLogSales.l3 + dLogPermits.l3 + const

                Estimate Std. Error t value Pr(>|t|)
dLogSales.l1   -0.105004   0.052814  -1.988 0.047475 *
dLogPermits.l1  0.114168   0.031508   3.623 0.000328 ***
dLogSales.l2   -0.211276   0.052549  -4.021 6.94e-05 ***
dLogPermits.l2  0.072550   0.034036   2.132 0.033652 *
dLogSales.l3   -0.166853   0.049517  -3.370 0.000826 ***
dLogPermits.l3  0.074372   0.031557   2.357 0.018918 *
const           0.003305   0.002169   1.524 0.128324
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 0.04336 on 399 degrees of freedom
Multiple R-Squared: 0.0724,     Adjusted R-squared: 0.05845
F-statistic: 5.191 on 6 and 399 DF,  p-value: 3.702e-05



Estimation results for equation dLogPermits:
============================================
dLogPermits = dLogSales.l1 + dLogPermits.l1 + dLogSales.l2 + dLogPermits.l2 + dLogSales.l3 + dLogPermits.l3 + const

                Estimate Std. Error t value Pr(>|t|)
dLogSales.l1    0.221704   0.088247   2.512  0.01239 *
dLogPermits.l1 -0.394532   0.052647  -7.494 4.35e-13 ***
dLogSales.l2    0.151305   0.087804   1.723  0.08562 .
dLogPermits.l2 -0.183536   0.056871  -3.227  0.00135 **
dLogSales.l3   -0.091118   0.082738  -1.101  0.27144
dLogPermits.l3  0.221608   0.052729   4.203 3.26e-05 ***
const           0.004429   0.003624   1.222  0.22235
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 0.07246 on 399 degrees of freedom
Multiple R-Squared:  0.22,      Adjusted R-squared: 0.2083
F-statistic: 18.76 on 6 and 399 DF,  p-value: < 2.2e-16



Covariance matrix of residuals:
            dLogSales dLogPermits
dLogSales    0.001880    0.001127
dLogPermits  0.001127    0.005250

Correlation matrix of residuals:
            dLogSales dLogPermits
dLogSales      1.0000      0.3588
dLogPermits    0.3588      1.0000
```

Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3





Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

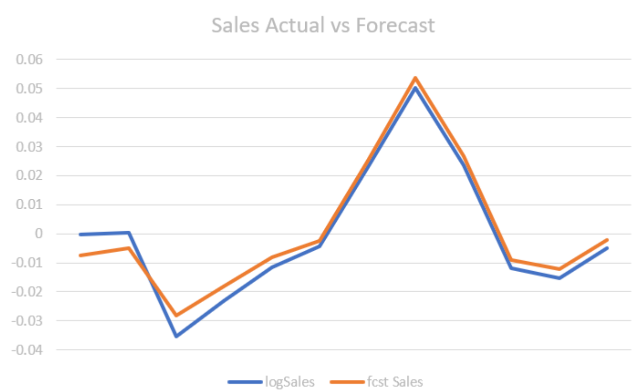
Significant lags were found for both variables. Adjusted R^2 is higher for Permits than for Sales, and Residuals appear reasonable. No major autocorrelation and the residuals behave like white noise. This confirms that the VAR(3) model is well specified.

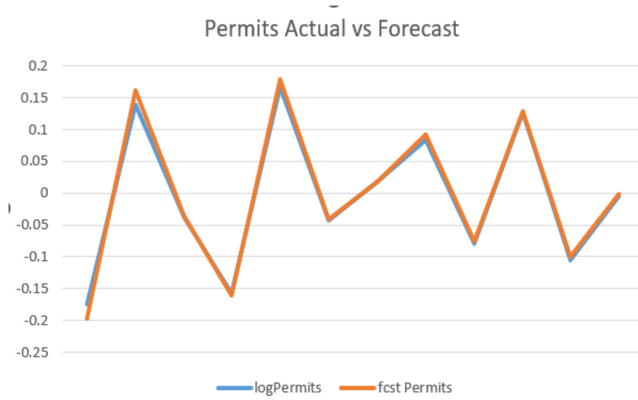
Q6:

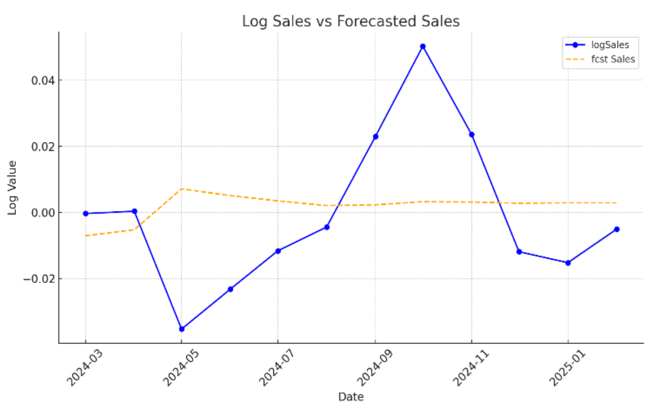
```
              fcst       lower      upper
 [1,] -0.007061349 -0.09205113 0.07792843
 [2,] -0.005246507 -0.09162897 0.08113596
 [3,]  0.007158671 -0.08038749 0.09470483
 [4,]  0.005078947 -0.08286799 0.09302589
 [5,]  0.003506754 -0.08457950 0.09159301
 [6,]  0.002065282 -0.08608645 0.09021701
 [7,]  0.002310052 -0.08585562 0.09047572
 [8,]  0.003249932 -0.08492584 0.09142571
 [9,]  0.003133102 -0.08504380 0.09131001
[10,]  0.002798765 -0.08537913 0.09097666
[11,]  0.002906896 -0.08527133 0.09108512
[12,]  0.002893125 -0.08528511 0.09107136
              CI
 [1,] 0.08498978
 [2,] 0.08638246
 [3,] 0.08754616
 [4,] 0.08794694
 [5,] 0.08808625
 [6,] 0.08815173
 [7,] 0.08816567
 [8,] 0.08817578
 [9,] 0.08817690
[10,] 0.08817789
[11,] 0.08817823
[12,] 0.08817823

$dLogPermits
              fcst      lower     upper        CI
 [1,] -0.022497986 -0.1645079 0.1195119 0.1420099
 [2,]  0.022569845 -0.1287686 0.1739082 0.1513384
 [3,]  0.002408050 -0.1489671 0.1537832 0.1513751
 [4,] -0.004212651 -0.1612711 0.1528458 0.1570585
 [5,]  0.013337733 -0.1463422 0.1730177 0.1596799
 [6,]  0.001367101 -0.1583847 0.1611189 0.1597518
 [7,]  0.001033619 -0.1592079 0.1612752 0.1602416
 [8,]  0.007230946 -0.1534265 0.1678884 0.1606575
 [9,]  0.002571076 -0.1581204 0.1632626 0.1606915
[10,]  0.003292216 -0.1574344 0.1640188 0.1607266
[11,]  0.005058891 -0.1557293 0.1658471 0.1607882
[12,]  0.003180894 -0.1576189 0.1639806 0.1607998
```

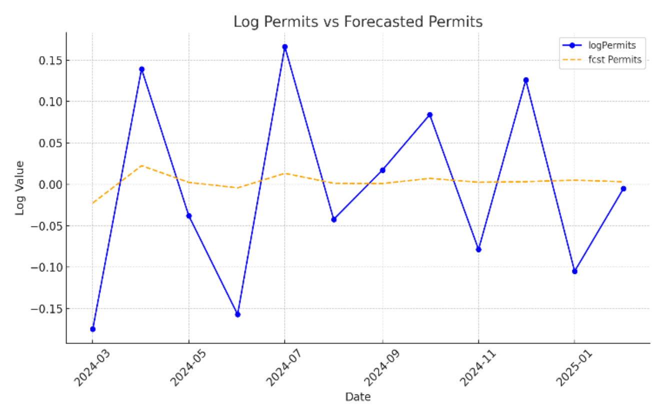
Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

In Excel, I found several methods to display the results. I chose some of them to ensure clarity. (stacked line and line with markets)



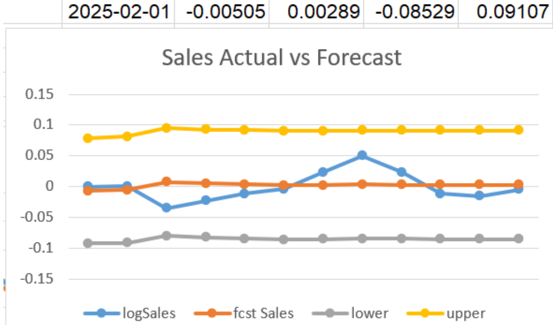


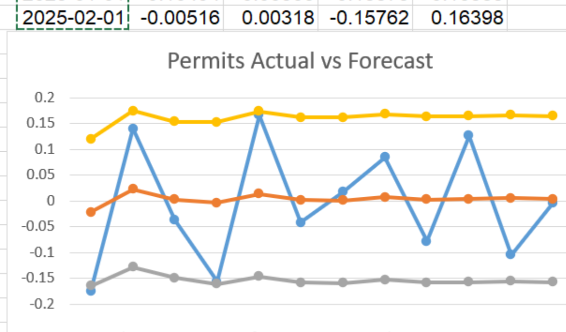




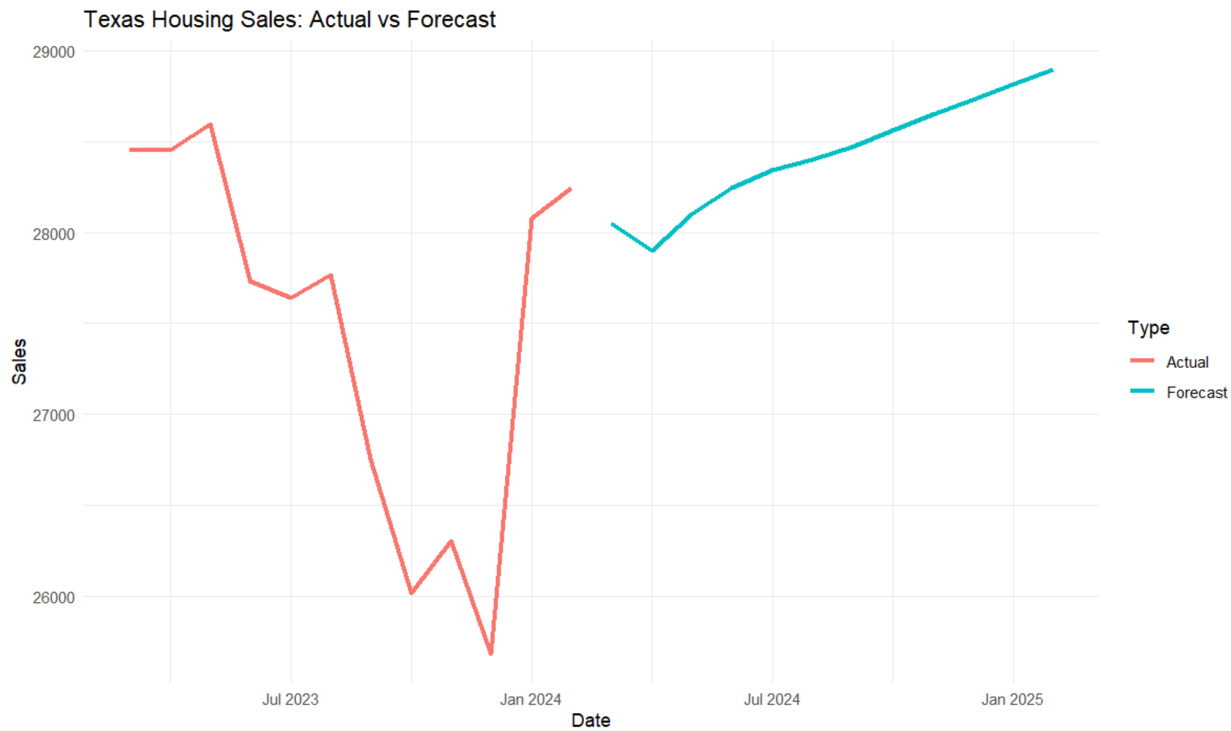
Date	logSales	fcst Sales	lower	upper
2024-03-01	-0.00035	-0.00706	-0.09205	0.07793
2024-04-01	0.00038	-0.00525	-0.09163	0.08114
2024-05-01	-0.03532	0.00716	-0.08039	0.0947
2024-06-01	-0.0232	0.00508	-0.08287	0.09303
2024-07-01	-0.01163	0.00351	-0.08458	0.09159
2024-08-01	-0.00442	0.00207	-0.08609	0.09022
2024-09-01	0.02302	0.00231	-0.08586	0.09048
2024-10-01	0.05032	0.00325	-0.08493	0.09143
2024-11-01	0.02358	0.00313	-0.08504	0.09131
2024-12-01	-0.01189	0.0028	-0.08538	0.09098
2025-01-01	-0.01522	0.00291	-0.08527	0.09109
2025-02-01	-0.00505	0.00289	-0.08529	0.09107

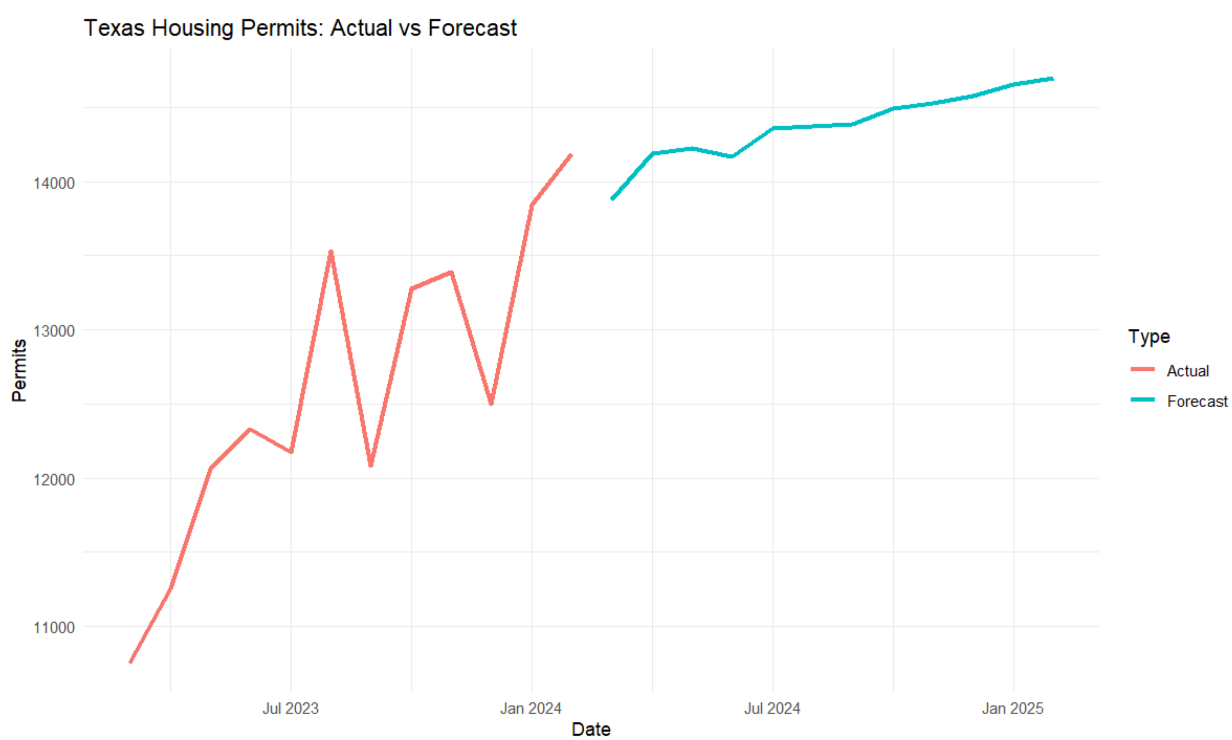
Date	logPermits	fcst Permi	lower	upper
2024-03-01	-0.17485	-0.0225	-0.16451	0.11951
2024-04-01	0.13937	0.02257	-0.12877	0.17391
2024-05-01	-0.03794	0.00241	-0.14897	0.15378
2024-06-01	-0.15712	-0.00421	-0.16127	0.15285
2024-07-01	0.16626	0.01334	-0.14634	0.17302
2024-08-01	-0.04227	0.00137	-0.15838	0.16112
2024-09-01	0.01737	0.00103	-0.15921	0.16128
2024-10-01	0.08417	0.00723	-0.15343	0.16789
2024-11-01	-0.07859	0.00257	-0.15812	0.16326
2024-12-01	0.12607	0.00329	-0.15743	0.16402
2025-01-01	-0.10494	0.00506	-0.15573	0.16585
2025-02-01	-0.00516	0.00318	-0.15762	0.16398





Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3





Name : Ammar Yasir A Alshehre
UIN : 636000447
ECMT 674 #HW3

As seen, I did not expect the results. the forecast model used for both sales and permits does not account for the high volatility in the actual data, making the forecasting less representative of the actual trend on the opposite of the log, which suggests more variability in both series.

However, the first plot for both series (stacked line), shows a very strong behavior for both sales and permits of following the fluctuations of the actual data.

Summary of the Outputs

QUESTION	DESCRIPTION
1	Plot of raw series
2	ndiffs() result for levels
2	Plot of differenced series
3	ndiffs() result for log-series
3	Plot of log-differenced series
4	Optimal lags from VARselect()
5	VAR model summary
5	Residual diagnostics
6	12-month log-diff forecast
6	Actual vs forecast plots