STAT 429 Project: US Housing Market

Kunal Bhardwaj

2024-02-23

```
library(readr)
MSPUS <- read_csv("Data/MSPUS.csv",
                  col_types = cols(DATE = col_date(format = "%Y-%m-%d")))

MSACSR <- read_csv("Data/MSACSR.csv",
                   col_types = cols(DATE = col_date(format = "%Y-%m-%d")))

HOUST <- read_csv("Data/HOUST.csv",
                  col_types = cols(DATE = col_date(format = "%Y-%m-%d"),
                                   HOUST = col_integer()))

RHORUSQ156N <- read_csv("Data/RHORUSQ156N.csv",
    col_types = cols(DATE = col_date(format = "%Y-%m-%d")))

MORTGAGE30US <- read_csv("Data/MORTGAGE30US.csv",
                         col_types = cols(DATE = col_date(format = "%Y-%m-%d"),
                                          MORTGAGE30US = col_number()))
```

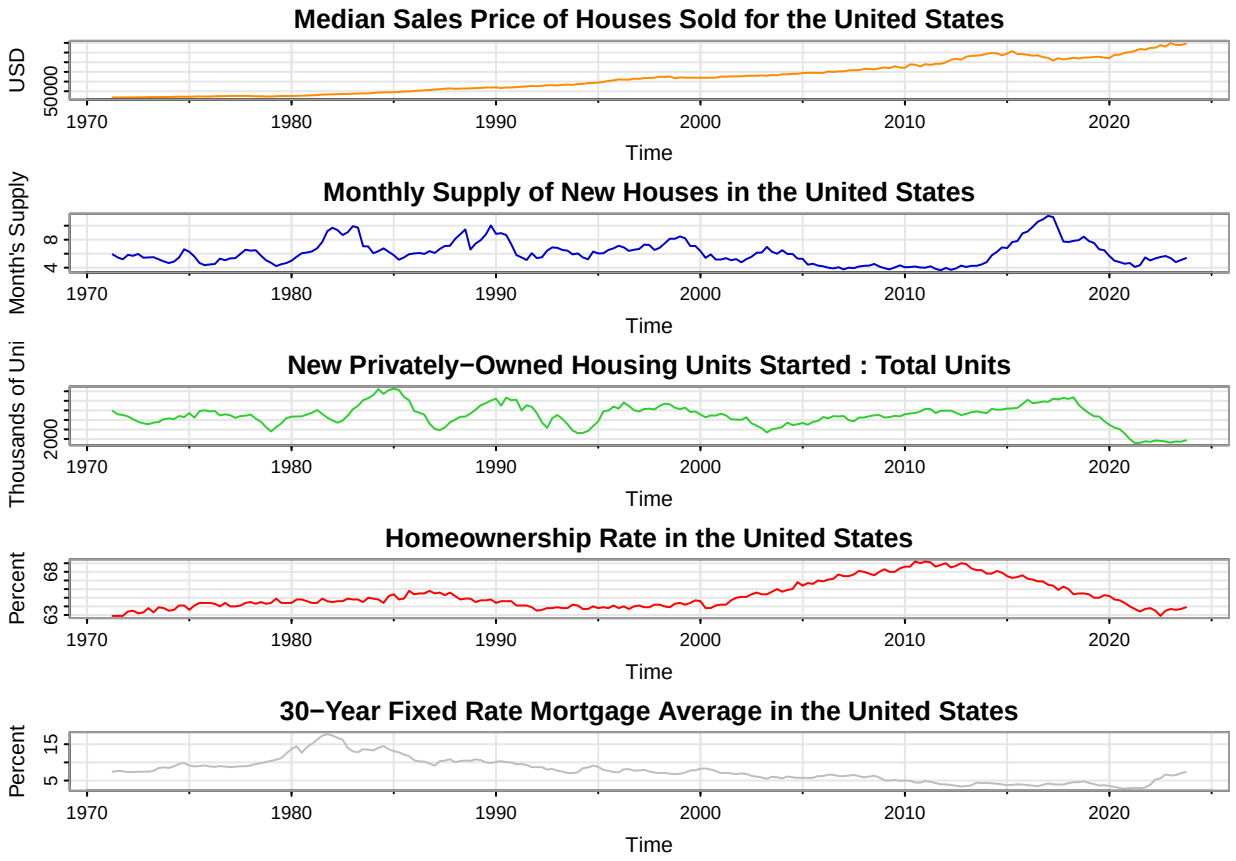
```
library(astsa)
library(ggplot2)
library(ggfortify)
library(fpp2)
library(imputeTS) # For handling missing values by imputation
library(tseries) # To carry out ADF & KPSS Tests
library(lmtest) # To carry out BP-Test

mspus.ts <- ts(MSPUS$MSPUS, start = c(1971, 2), frequency = 4, end = c(2023, 4))
msacsr.ts <- ts(MSACSR$MSACSR, start = c(1971, 2), frequency = 4, end = c(2023, 4))
houst.ts <- ts(HOUST$HOUST, start = c(1971, 2), frequency = 4, end = c(2023, 4))
rhorusq156n.ts <- ts(RHORUSQ156N$RHORUSQ156N, start = c(1971, 2), frequency = 4, end = c(2023, 4))


# MORTGAGE30US contains missing values. They are occupied by "."
# They will be imputed by replacing with NA follwed by linear interpolation
#MORTGAGE30US$MORTGAGE30US <- ifelse(MORTGAGE30US$MORTGAGE30US == ".",
#                                    NA, MORTGAGE30US$MORTGAGE30US)
#MORTGAGE30US$MORTGAGE30US <- na_interpolation(as.numeric(MORTGAGE30US$MORTGAGE30US))
# Select 2nd column from MORTGAGE30US. The first value on Apr 2 1971 falls on the 14th week
mortgage.ts <- ts(MORTGAGE30US$MORTGAGE30US, start = c(1971, 2) , frequency = 4, end = c(2023, 4))

par(mfrow=c(5,1))
tsplot(mspus.ts, col ="darkorange",
```

```
       main = "Median Sales Price of Houses Sold for the United States",
       ylab = "USD")
tsplot(msacsr.ts, col ="mediumblue",
       main = "Monthly Supply of New Houses in the United States",
       ylab = "Month's Supply")
tsplot(houst.ts, col ="limegreen",
       main = "New Privately-Owned Housing Units Started : Total Units",
       ylab = "Thousands of Units")
tsplot(rhorusq156n.ts, col ="red",
       main = "Homeownership Rate in the United States",
       ylab = "Percent")
tsplot(mortgage.ts, col ="grey",
       main = "30-Year Fixed Rate Mortgage Average in the United States", ylab = "Percent")
```

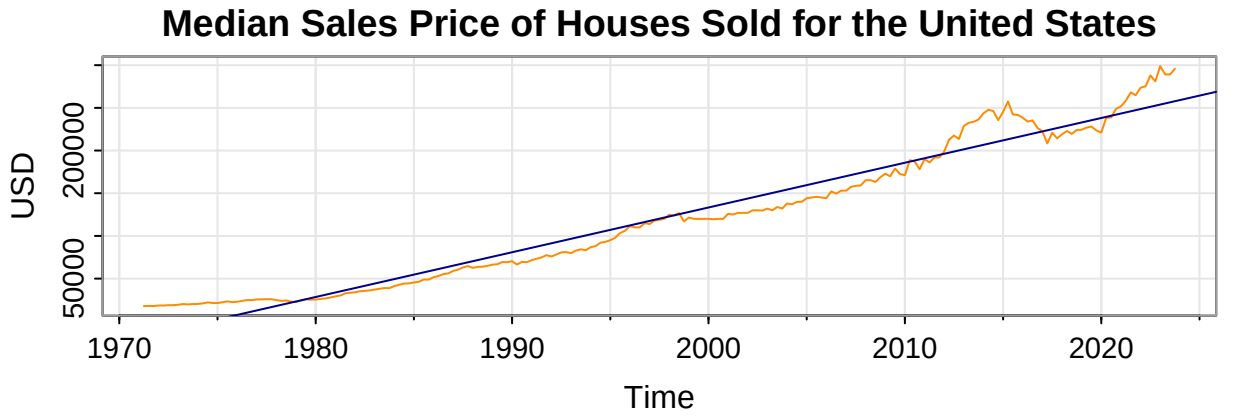


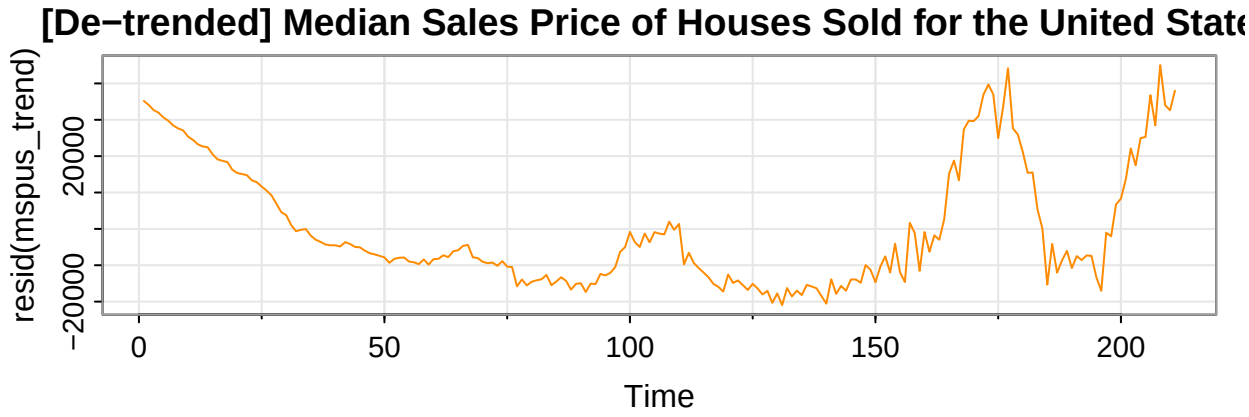
From the plot of Median Sales Price of Houses Sold, we can see that there exists an obvious trend in the data. Therefore de-trending needs to be carried out to achieve stationarity.

```
# De-trending MSPUS Time Series
mspus_trend <- lm(mspus.ts ~ time(mspus.ts))
#summary(mspus_trend)

par(mfrow=c(2,1))
tsplot(mspus.ts, col ="darkorange",
       main = "Median Sales Price of Houses Sold for the United States",
       ylab = "USD")
abline(mspus_trend, col = "darkblue")
```

```
tsplot(resid(mspus_trend), col ="darkorange",
       main = "[De-trended] Median Sales Price of Houses Sold for the United States")
```






```
# check for Stationarity using (Augmented) Dickey-Fuller Test and KPSS Test
adf.test(resid(mspus_trend))
```

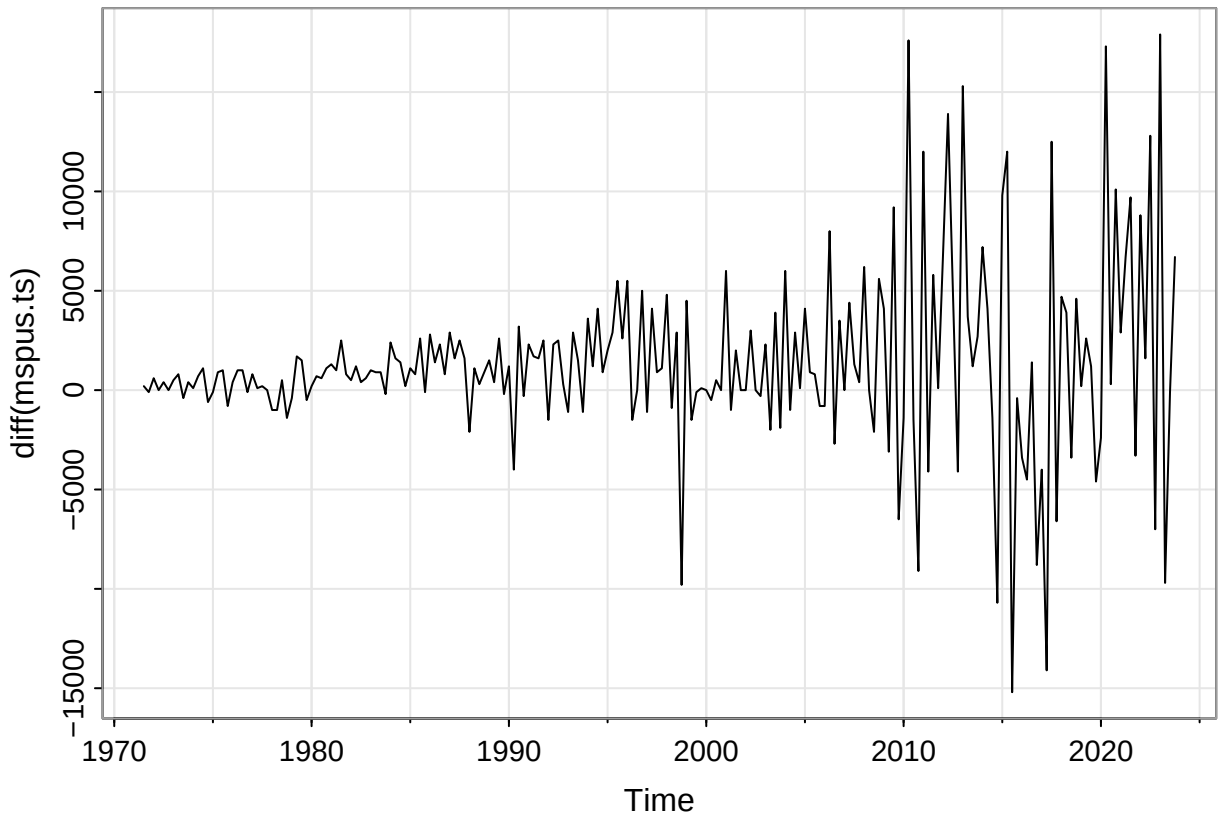
```
##
##  Augmented Dickey-Fuller Test
##
## data:  resid(mspus_trend)
## Dickey-Fuller = -2.8541, Lag order = 5, p-value = 0.2179
## alternative hypothesis: stationary
```

```
kpss.test(resid(mspus_trend))
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  resid(mspus_trend)
## KPSS Level = 0.62599, Truncation lag parameter = 4, p-value = 0.02027
```

Both ADF and KPSS tests conclude that the MSPUS de-trended series is non-stationary. We will try differencing to achieve stationarity.

```r
# Differencing MSPUS Time Series
tsplot(diff(mspus.ts))
```




```r
adf.test(diff(mspus.ts))
```

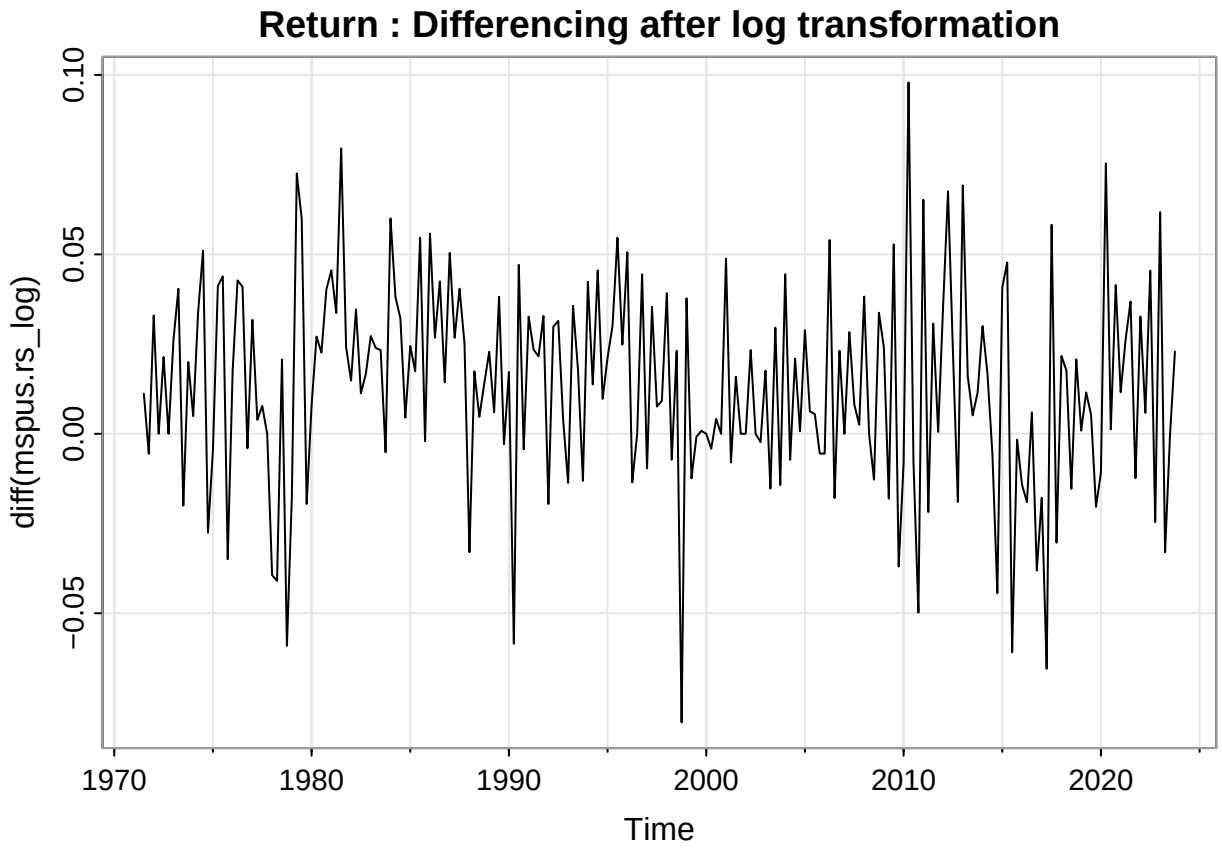
```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(mspus.ts)
## Dickey-Fuller = -3.6937, Lag order = 5, p-value = 0.02568
## alternative hypothesis: stationary
```

```r
kpss.test(diff(mspus.ts))
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  diff(mspus.ts)
## KPSS Level = 0.45015, Truncation lag parameter = 4, p-value = 0.05554
```

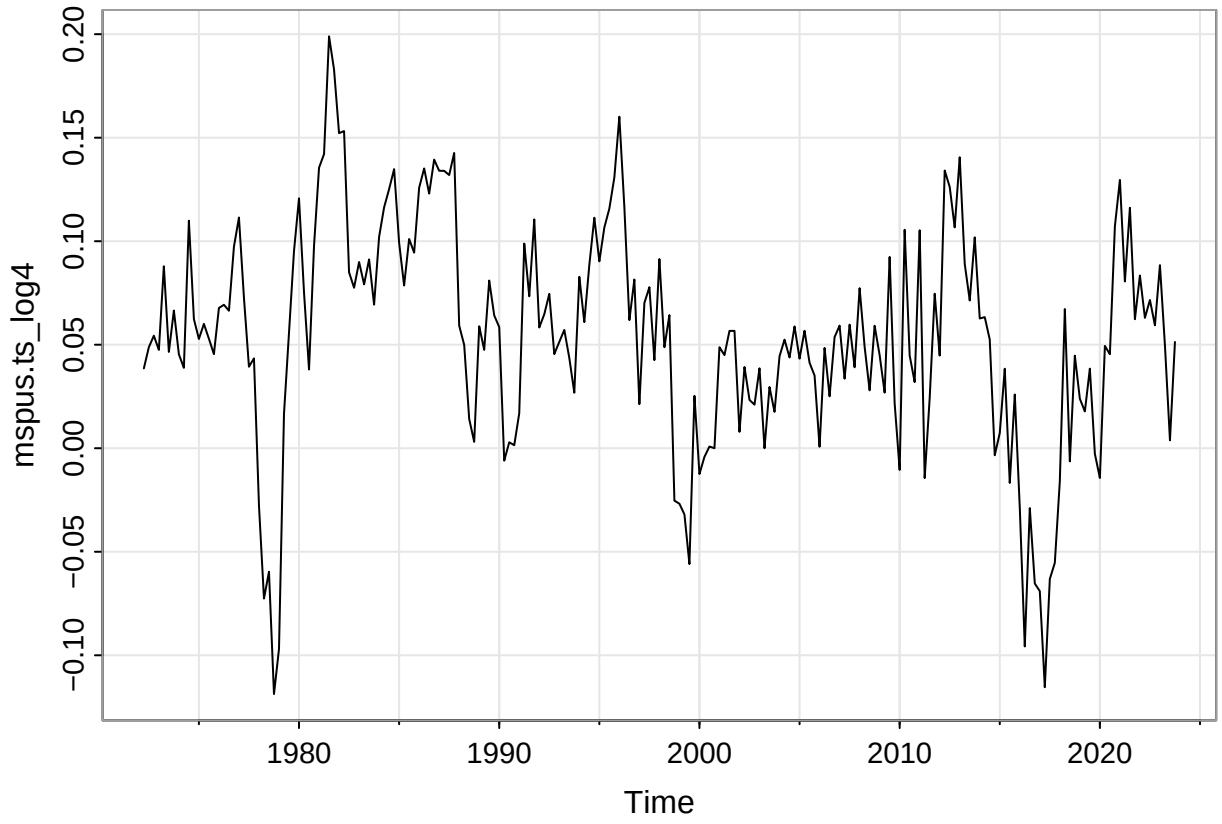
The series passes both these tests of stationarity. But the series exhibits an obvious heteroscedasticity where higher levels are associated with higher variation. A log-transformation is recommended.

```r
mspus.rs_log <- log(mspus.ts)
tsplot(diff(mspus.rs_log), main = "Return : Differencing after log transformation") # Return (Differenc
```



The BP Test for homoscedasticity fails for MSPUS, log(MSPUS), diff(MSPUS), diff(MSPUS, lag =4), diff(log(MSPUS)), and diff(resid(mspus_trend)). Will try differencing with a lag of 4 after log transformation.

```r
# Differencing with a lag of 4 after taking log
mspus.ts_log4 <- diff(log(mspus.ts), lag = 4)
tsplot(mspus.ts_log4)
```




```
adf.test(mspus.ts_log4)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  mspus.ts_log4
## Dickey-Fuller = -3.5134, Lag order = 5, p-value = 0.04271
## alternative hypothesis: stationary
```

```
kpss.test(mspus.ts_log4)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  mspus.ts_log4
## KPSS Level = 0.51376, Truncation lag parameter = 4, p-value = 0.03857
```

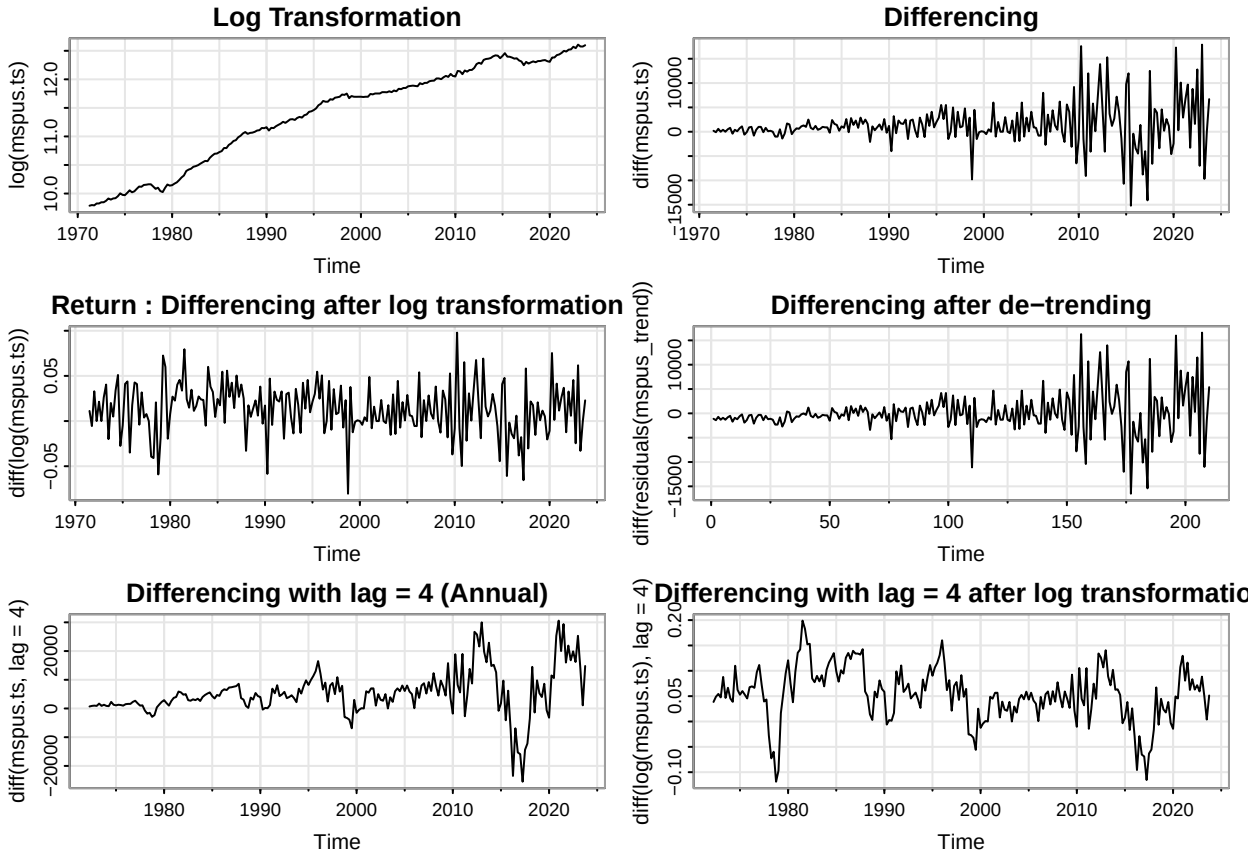
```
bptest(lm(mspus.ts_log4 ~ time(mspus.ts_log4)))
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lm(mspus.ts_log4 ~ time(mspus.ts_log4))
## BP = 0.41748, df = 1, p-value = 0.5182
```

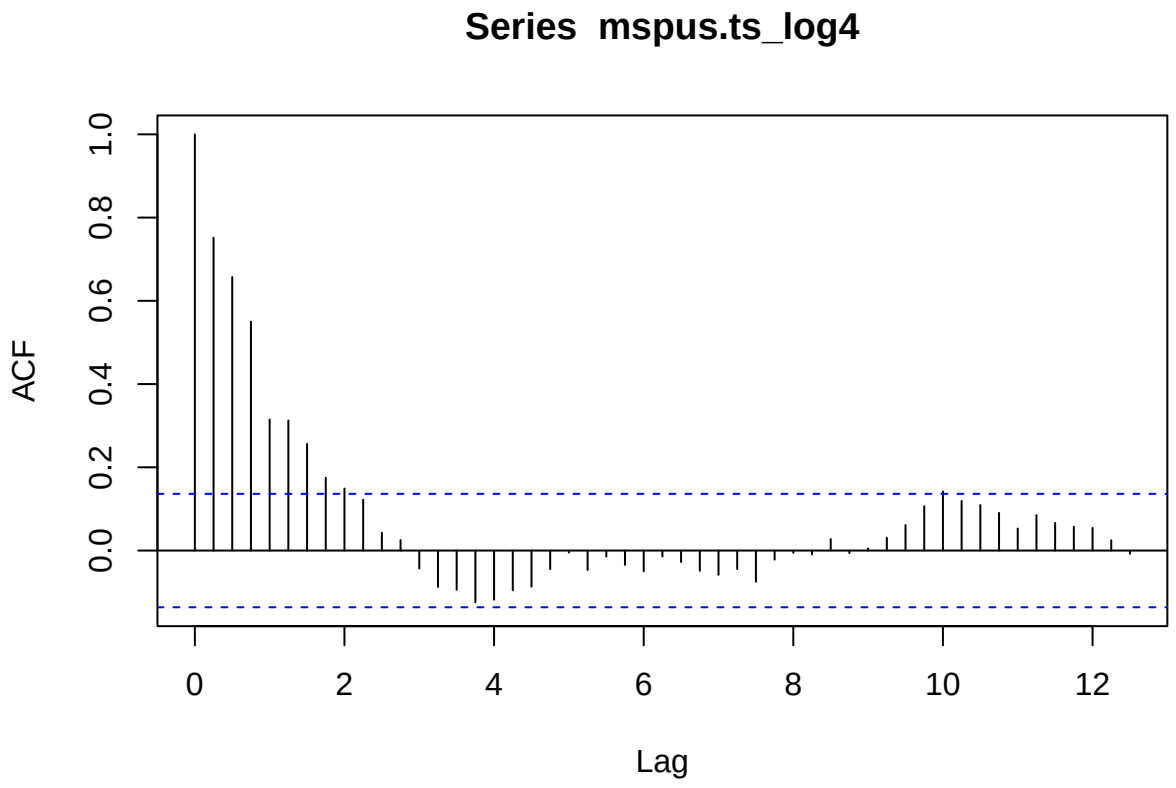
```
# Log transformation of the differencing at lag 4 is not carried out because it produces NaNs
```

Differencing with a lag of 4 after taking log passes ADF Test & KPSS tests of stationarity and BP-Test for homoscedasticity. Taking lag = 4 insted of the default lag = 1 while differencing seems to have helped in removing the seasonality associated with the MSPUS series.

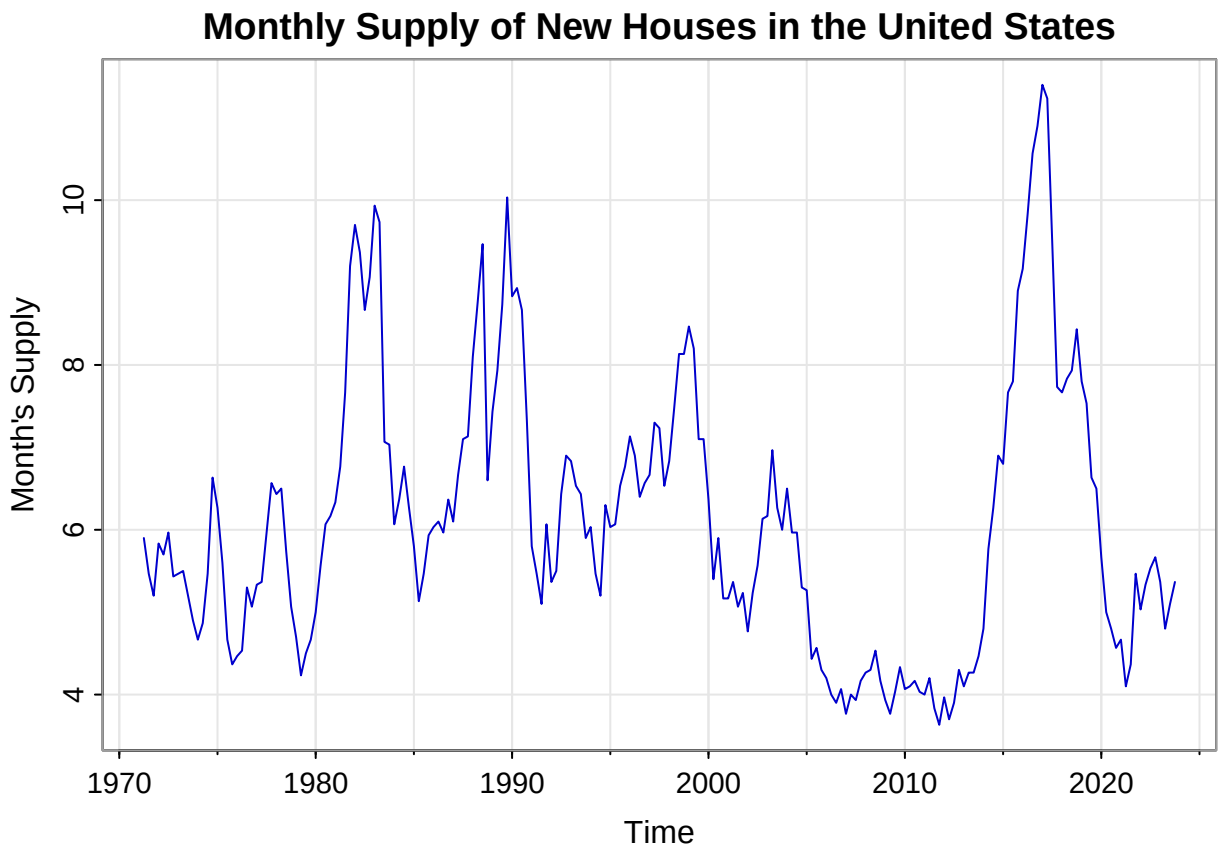
```
par(mfrow=c(3,2))
tsplot(log(mspus.ts), main = "Log Transformation") # Log Transformation
tsplot(diff(mspus.ts), main = "Differencing") # Differencing
tsplot(diff(log(mspus.ts)), main = "Return : Differencing after log transformation") # Return (Differen
tsplot(diff(residuals(mspus_trend)), main = "Differencing after de-trending") # Differencing after de-t
tsplot(diff(mspus.ts, lag = 4), main = "Differencing with lag = 4 (Annual)") # Differencing with a lag
tsplot(diff(log(mspus.ts), lag = 4), main = "Differencing with lag = 4 after log transformation")
```




```
# ACF of transformed outcome variable
acf(mspus.ts_log4, lag.max = 50)
```




```
#msacsr_trend <- lm(msacsr.ts ~ time(msacsr.ts))
#summary(msacsr_trend)
tsplot(msacsr.ts, col ="mediumblue",
       main = "Monthly Supply of New Houses in the United States",
       ylab = "Month's Supply")
```




```
#abline(msacsr_trend, col = "red")
#tsplot(resid(msacsr_trend), col = "darkblue",
#       main = "[De-trended] Monthly Supply of New Houses in the United States")
adf.test(msacsr.ts)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  msacsr.ts
## Dickey-Fuller = -3.6194, Lag order = 5, p-value = 0.03266
## alternative hypothesis: stationary
```

```
kpss.test(msacsr.ts)
```

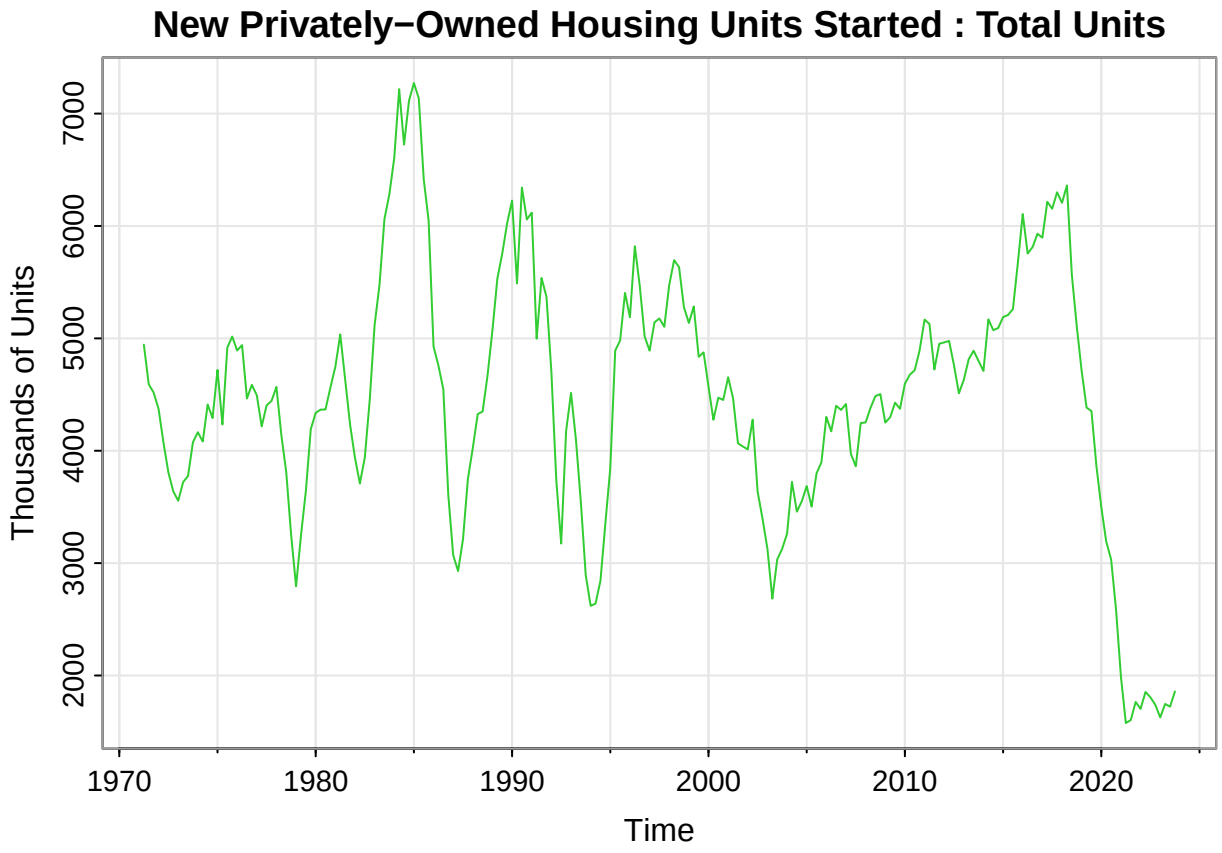
```
## Warning in kpss.test(msacsr.ts): p-value greater than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  msacsr.ts
## KPSS Level = 0.21796, Truncation lag parameter = 4, p-value = 0.1
```

```
#bptest(lm(msacsr.ts ~ time(msacsr.ts)))
#bptest(lm(log(msacsr.ts) ~ time(msacsr.ts)))
```

Thus the MSACSR series is stationary.

```
tsplot(houst.ts, col ="limegreen",
       main = "New Privately-Owned Housing Units Started : Total Units",
       ylab = "Thousands of Units")
```




```
adf.test(houst.ts)
```

```
## Warning in adf.test(houst.ts): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  houst.ts
## Dickey-Fuller = -4.0399, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

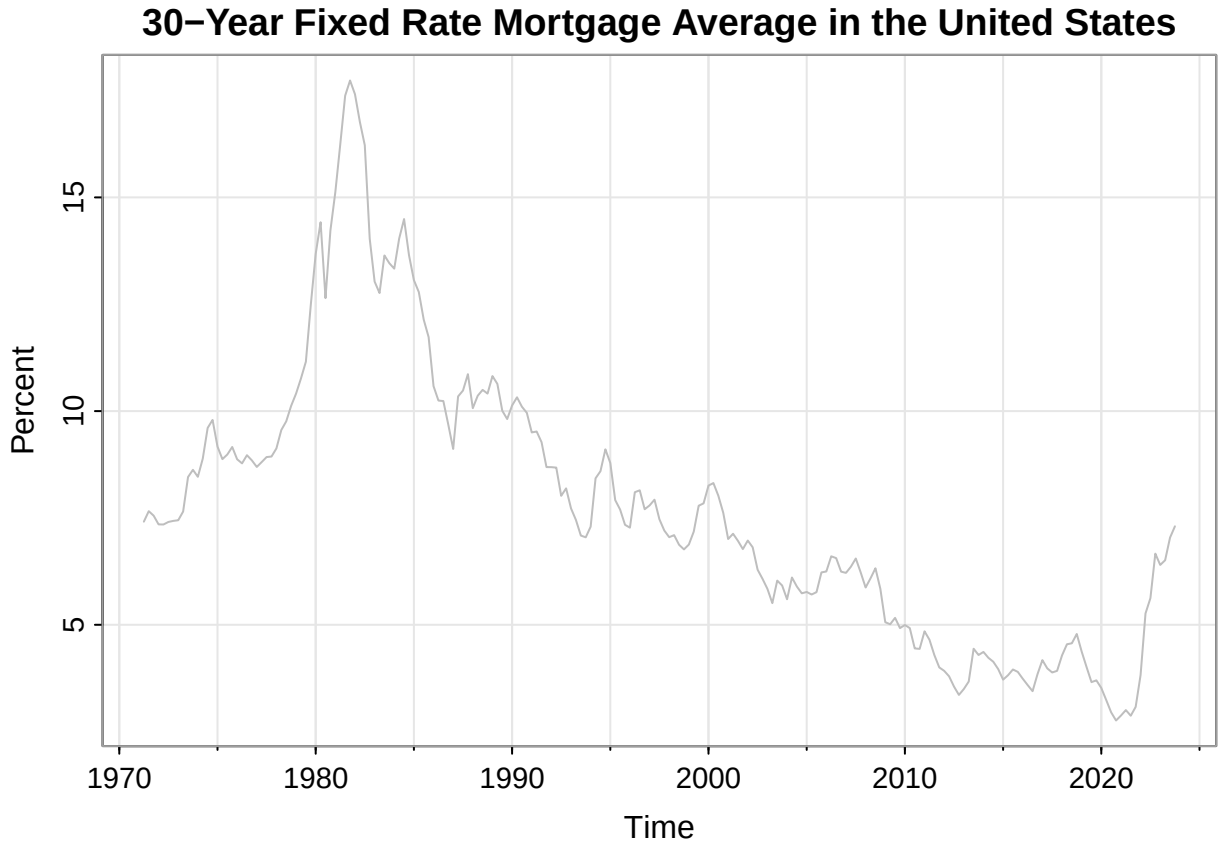
```
kpss.test(houst.ts)
```

```
## Warning in kpss.test(houst.ts): p-value greater than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  houst.ts
## KPSS Level = 0.25362, Truncation lag parameter = 4, p-value = 0.1
```

The ADF Test & KPSS tests confirm stationarity of HOUST.

```
tsplot(mortgage.ts, col ="grey",
       main = "30-Year Fixed Rate Mortgage Average in the United States", ylab = "Percent")
```




```
adf.test(mortgage.ts)
```

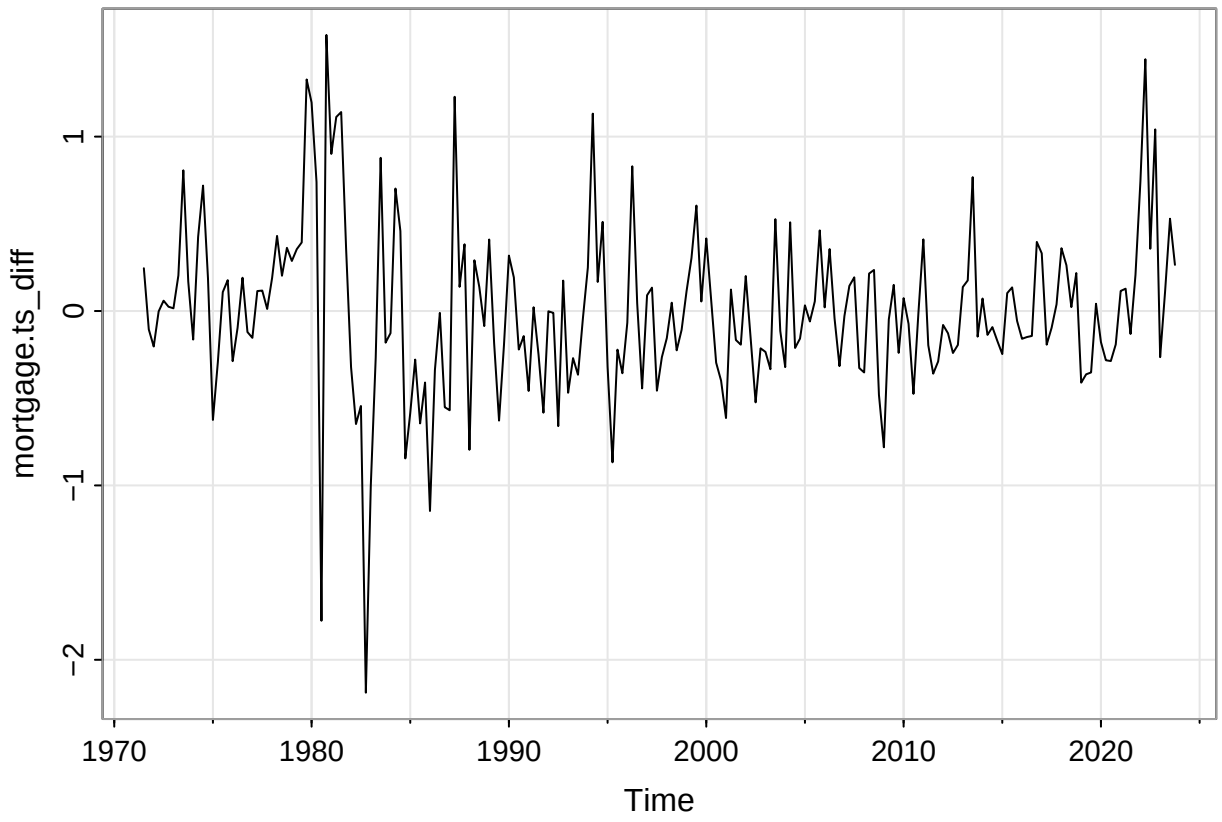
```
##
##  Augmented Dickey-Fuller Test
##
## data:  mortgage.ts
## Dickey-Fuller = -2.6749, Lag order = 5, p-value = 0.293
## alternative hypothesis: stationary
```

```
kpss.test(mortgage.ts)
```

```
## Warning in kpss.test(mortgage.ts): p-value smaller than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  mortgage.ts
## KPSS Level = 2.9416, Truncation lag parameter = 4, p-value = 0.01
```

```
mortgage.ts_diff <- diff(mortgage.ts)
tsplot(mortgage.ts_diff)
```




```
adf.test(mortgage.ts_diff)
```

```
## Warning in adf.test(mortgage.ts_diff): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  mortgage.ts_diff
## Dickey-Fuller = -5.5994, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

```
kpss.test(mortgage.ts_diff)
```

```
## Warning in kpss.test(mortgage.ts_diff): p-value greater than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  mortgage.ts_diff
## KPSS Level = 0.13192, Truncation lag parameter = 4, p-value = 0.1
```

```
bptest(lm(mortgage.ts_diff ~ time(mortgage.ts_diff)))
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lm(mortgage.ts_diff ~ time(mortgage.ts_diff))
## BP = 7.2256, df = 1, p-value = 0.007187
```

```
# Clipping the series to bring them to the same length
msacsr.ts <- ts(MSACSR$MSACSR, start = c(1972, 2), frequency = 4, end = c(2023, 4))
houst.ts <- ts(HOUST$HOUST, start = c(1972, 2), frequency = 4, end = c(2023, 4))
mortgage.ts <- ts(MORTGAGE30US$MORTGAGE30US, start = c(1972, 1) , frequency = 4, end = c(2023, 4))
mortgage.ts_diff <- diff(mortgage.ts)
```

```
trend <- time(mspus.ts_log4)
reg_model <- lm(mspus.ts_log4 ~ trend + msacsr.ts + houst.ts + mortgage.ts_diff)
summary(reg_model)
```

```
##
## Call:
## lm(formula = mspus.ts_log4 ~ trend + msacsr.ts + houst.ts + mortgage.ts_diff)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.185861 -0.031805 -0.001433  0.032478  0.140292
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.045e+00  4.734e-01   4.320 2.44e-05 ***
## trend            -9.651e-04  2.359e-04  -4.092 6.19e-05 ***
## msacsr.ts        -8.968e-03  2.282e-03  -3.931 0.000116 ***
## houst.ts         -1.819e-06  3.549e-06  -0.513 0.608816
## mortgage.ts_diff -7.465e-03  7.420e-03  -1.006 0.315619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05033 on 202 degrees of freedom
## Multiple R-squared:  0.1469, Adjusted R-squared:   0.13
## F-statistic: 8.697 on 4 and 202 DF,  p-value: 1.699e-06
```

```
summary(aov(reg_model))
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## trend              1 0.0369 0.03694  14.582 0.000178 ***
## msacsr.ts          1 0.0483 0.04833  19.079    2e-05 ***
```

```
## houst.ts           1 0.0003 0.00029   0.113 0.736649
## mortgage.ts_diff   1 0.0026 0.00256   1.012 0.315619
## Residuals        202 0.5117 0.00253
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```