

Modeling and Forecasting House Price Trends

Kuanyi Chen

2025-01-19

```
# import necessary packages
library(tidyverse)
library(ggplot2)
library(forecast)
library(rsq)
```

I. Introduction

```
# import data
data <- read.csv("house_property_sales.csv")

# filter data to only include houses with 2 bedrooms
data <- data %>%
  filter(bedrooms == 2 & type == "house")
```

This time series data is the house property sales data for 2 bedroom houses for a specific region (region is not indicated in the source), recorded quarterly from 9/30/2007 to 9/30/2019. The data is the moving average of median house price, which is calculated by taking the average of median house prices over a consecutive period of time, updated as new data points are added. This is a useful time series data, as house prices are relevant to sellers and buyers in the market. It will give people an idea of the trends and changes in the market.

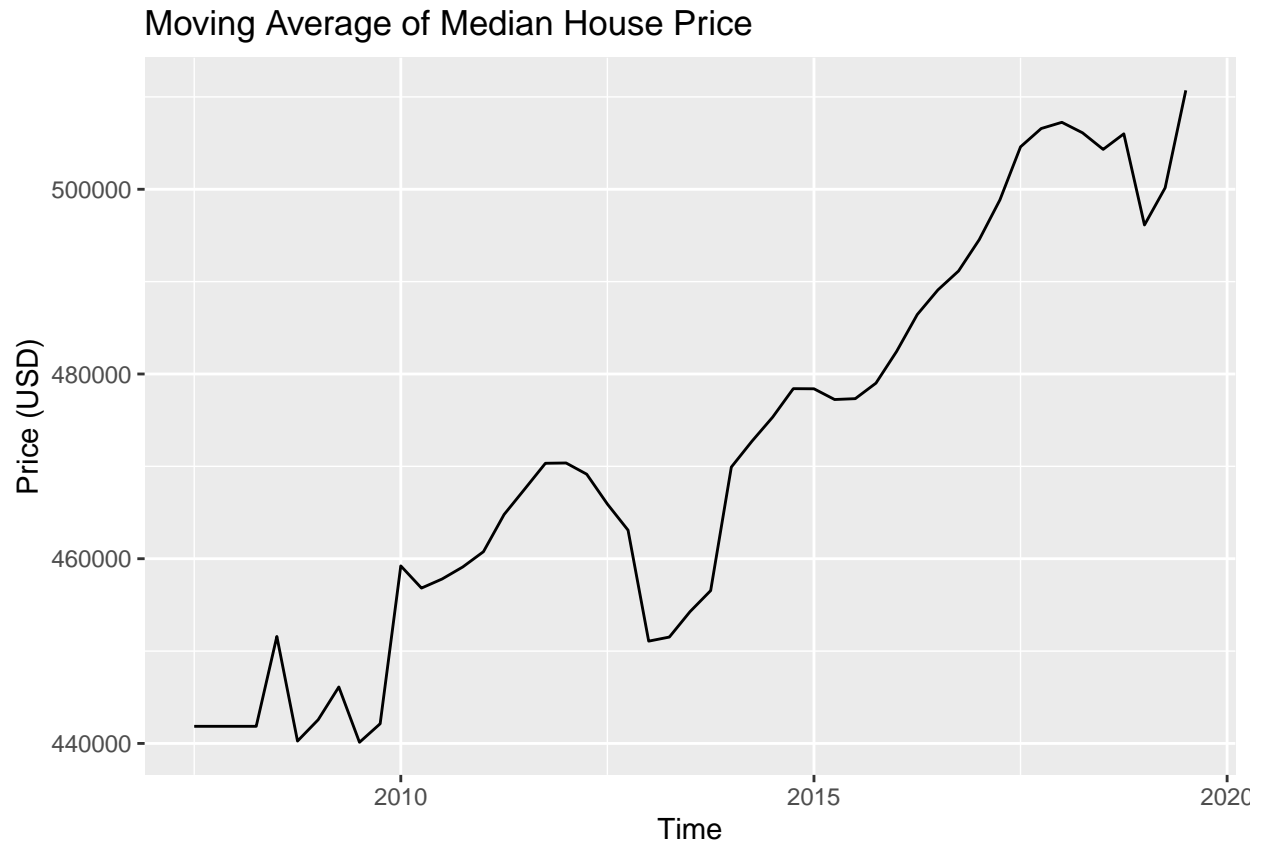
II. Results

1. Modeling and Forecasting Trend

(a)

```
# turn data into a time series
data_ts <- ts(data$MA, start = c(2007, 3), frequency = 4)

# plot the time series
autoplot(data_ts) +
  labs(title = "Moving Average of Median House Price",
       y = "Price (USD)")
```

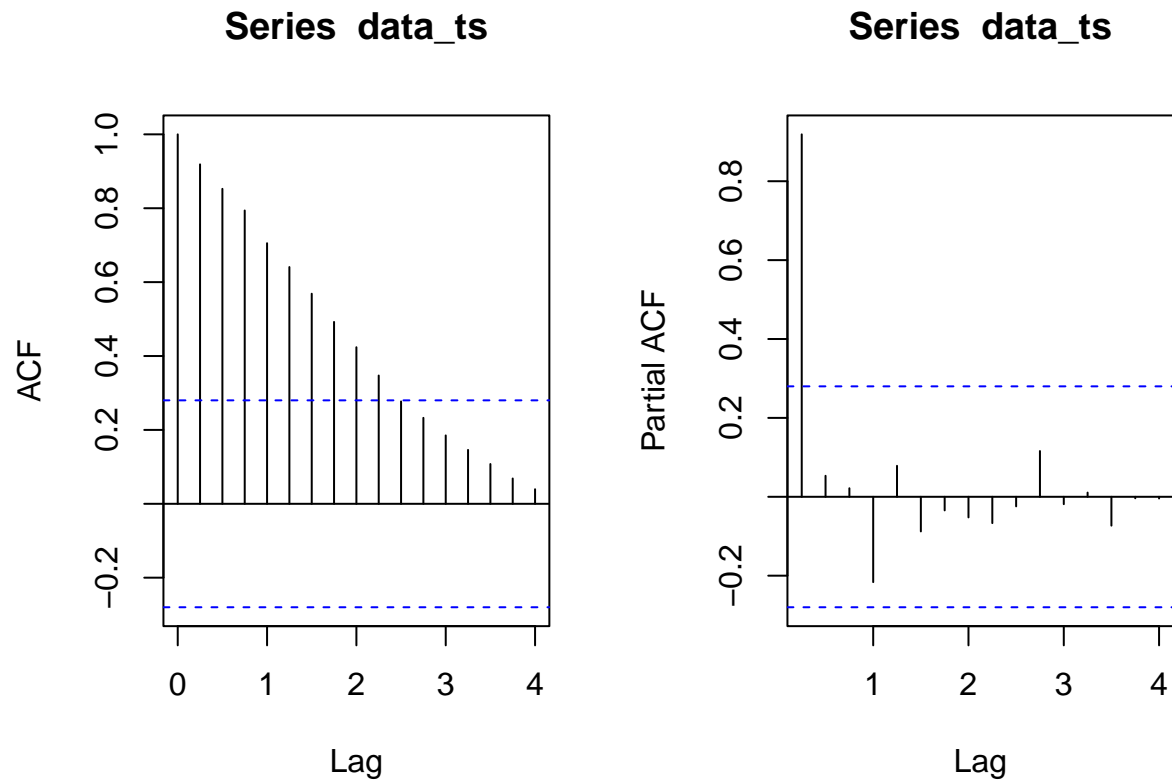


(b)

The plot in (a) does not suggest the data is covariance stationary. In order to be covariance stationary, a time series must satisfy three conditions: constant mean, constant variance, and time independent covariances. Since the plot has an overall increasing trend, it does not have constant mean, thus not covariance stationary.

(c)

```
# plot the ACF and PACF plot next to each other
par(mfrow=c(1,2))
acf(data_ts)
pacf(data_ts)
```



The ACF plot showcases a gradual decline in the bars, with the autocorrelation being significant until a little bit past the second lag. This gradual decline in autocorrelation means the autocorrelation between the time series and its lagged values gets weaker as the lag increases. This also means a long-term dependency in the data.

The PACF does not showcase significant autocorrelation at any lags except the zeroth lag, so it is not necessary to include any lags to address autocorrelation.

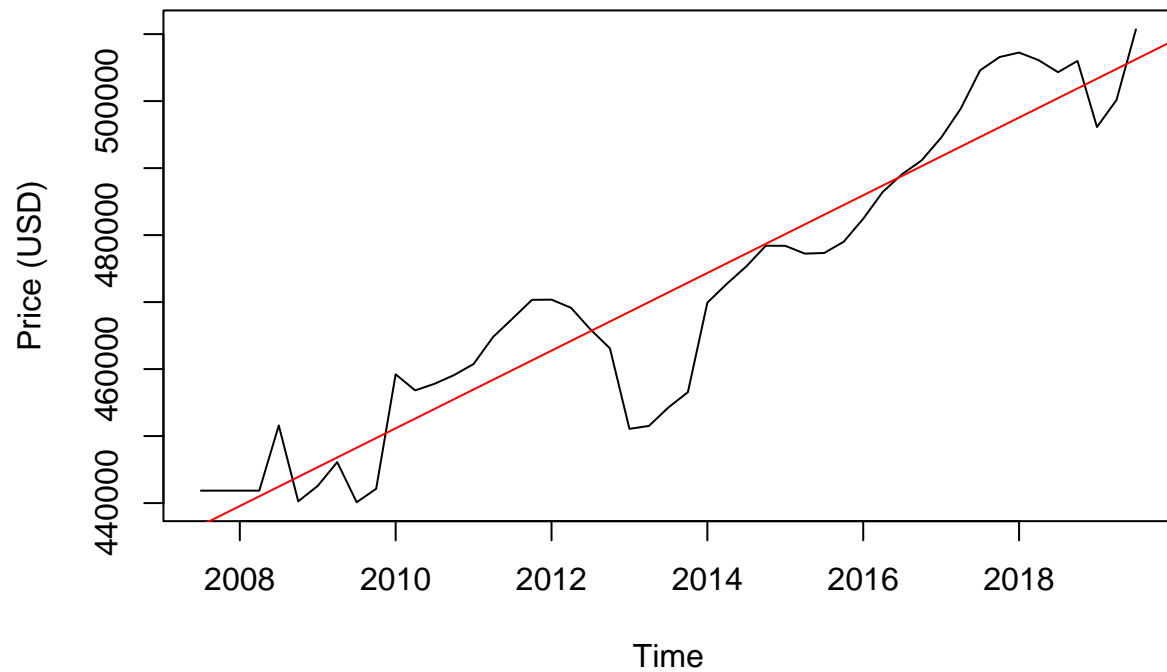
(d) (e) (f) (g)

```
# extract the time of the time series and convert into numeric type
t <- as.numeric(time(data_ts))
```

```
# fit and plot a linear model
linear_m <- lm(data_ts ~ t)
plot(data_ts, ylab = "Price (USD)", main = "Moving Average of Median House Price")
abline(linear_m, col = "red")
```

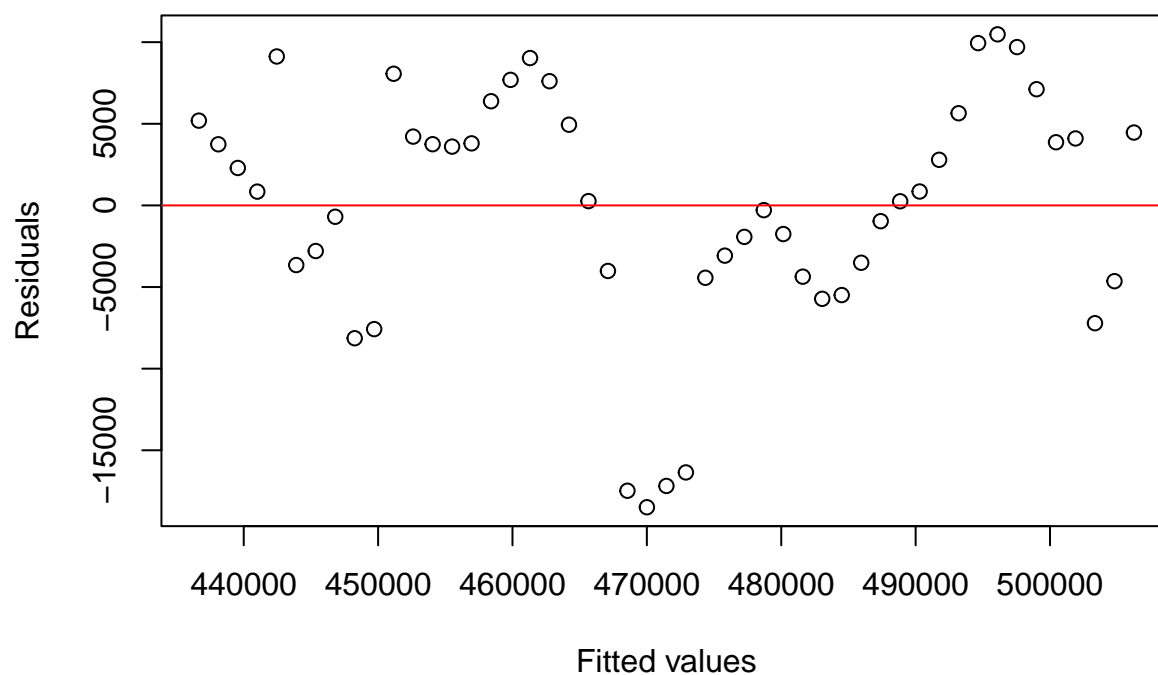
Linear Model

Moving Average of Median House Price



```
# residuals vs fitted values plot
plot(linear_m$fitted.values, linear_m$residuals,
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```

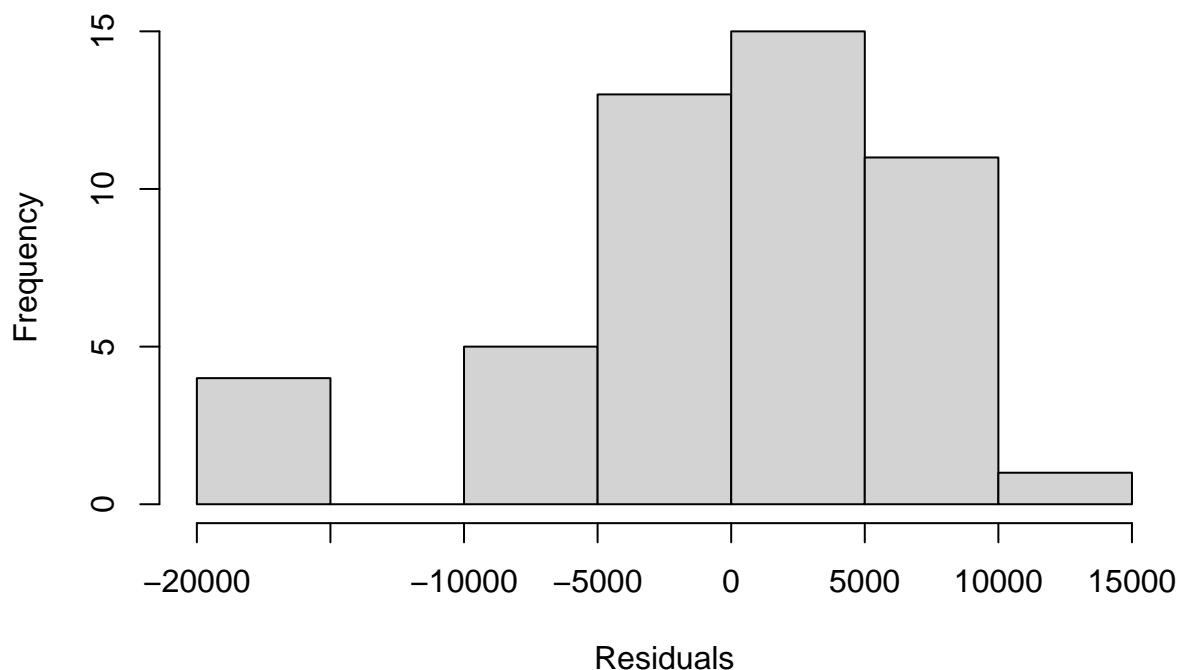
Residuals vs Fitted Values



The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals
hist(linear_m$residuals,
     main = "Histogram of Residuals",
     xlab = "Residuals")
```

Histogram of Residuals



The residuals histogram roughly follows a normal distribution with some values present on the further left negative side, which may be an outlier.

```
# diagnostics
summary(linear_m)
```

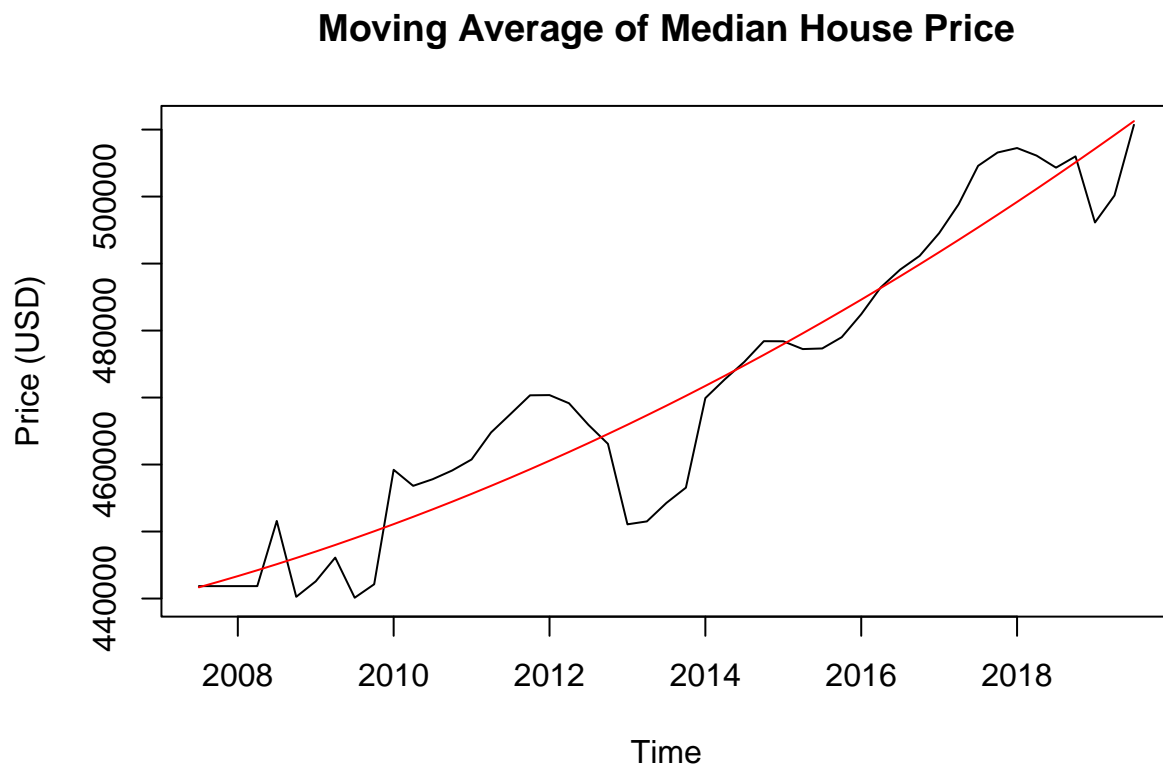
```
##
## Call:
## lm(formula = data_ts ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18488.2  -4014.8   844.5   4943.7  10478.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11204731    598094  -18.73  <2e-16 ***
## t              5799         297   19.52  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7351 on 47 degrees of freedom
## Multiple R-squared:  0.8902, Adjusted R-squared:  0.8879
## F-statistic: 381.1 on 1 and 47 DF,  p-value: < 2.2e-16
```

The adjusted R^2 is 0.8879, meaning linear model explains the variability in the data well. According to

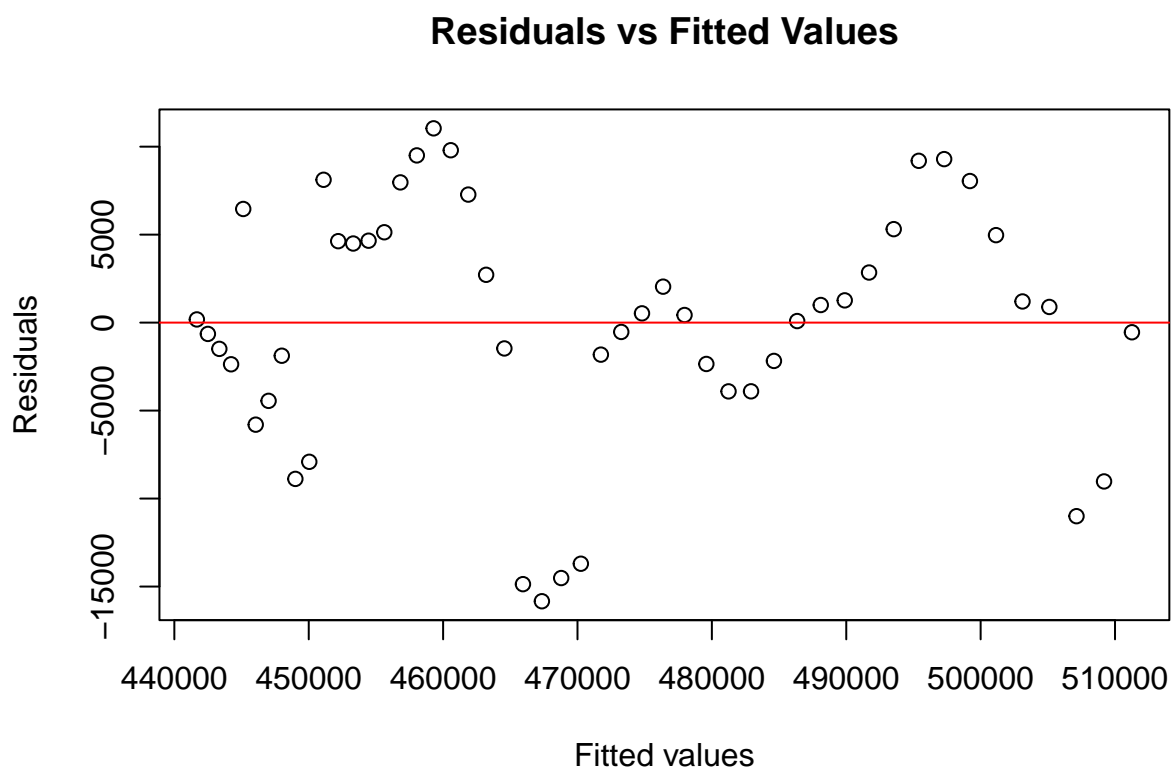
p-value of the t-distribution, both the coefficient and intercept are significant at the 5% significance level. According to the p-value of the F-distribution, the linear model is significant in fitting the data.

```
# fit and plot a quadratic model
quadratic_m <- lm(data_ts ~ t + I(t^2))
plot(data_ts, ylab = "Price (USD)", main = "Moving Average of Median House Price")
lines(t, quadratic_m$fitted.values, col = "red")
```

Quadratic Model



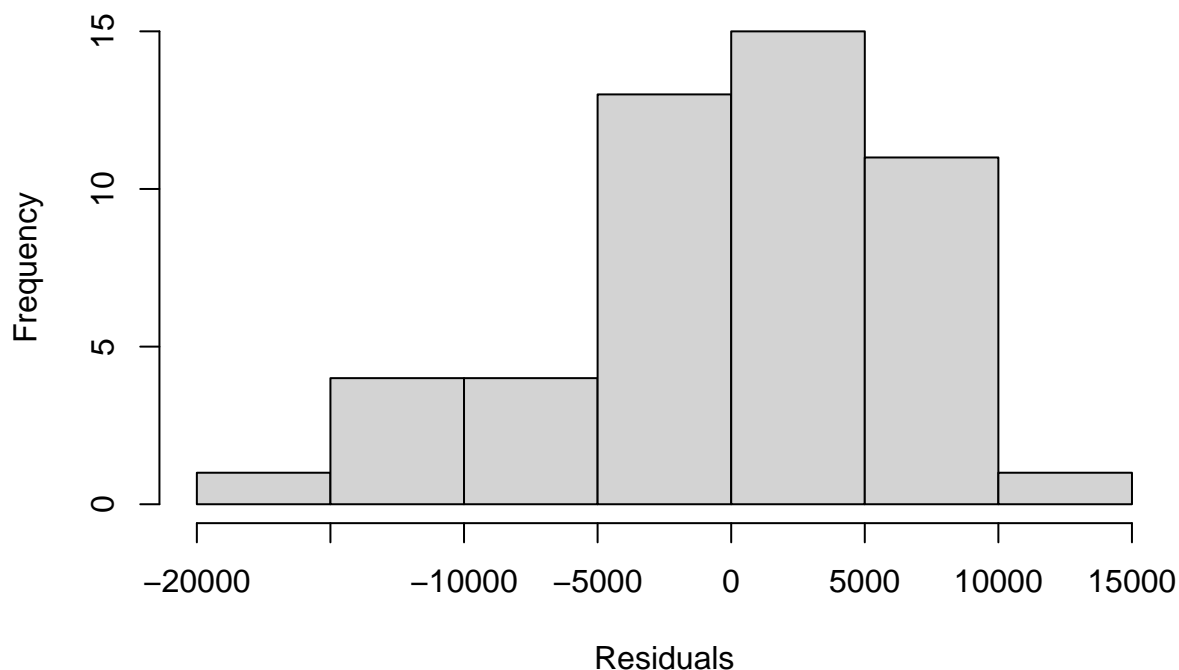
```
# residuals vs fitted values plot
plot(quadratic_m$fitted.values, quadratic_m$residuals,
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```



The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals  
hist(quadratic_m$residuals,  
      main = "Histogram of Residuals",  
      xlab = "Residuals")
```


Histogram of Residuals



The residuals histogram roughly follows a normal distribution and is left skewed. Some residuals on the further left may potentially be outliers.

```
# diagnostics
summary(quadratic_m)
```

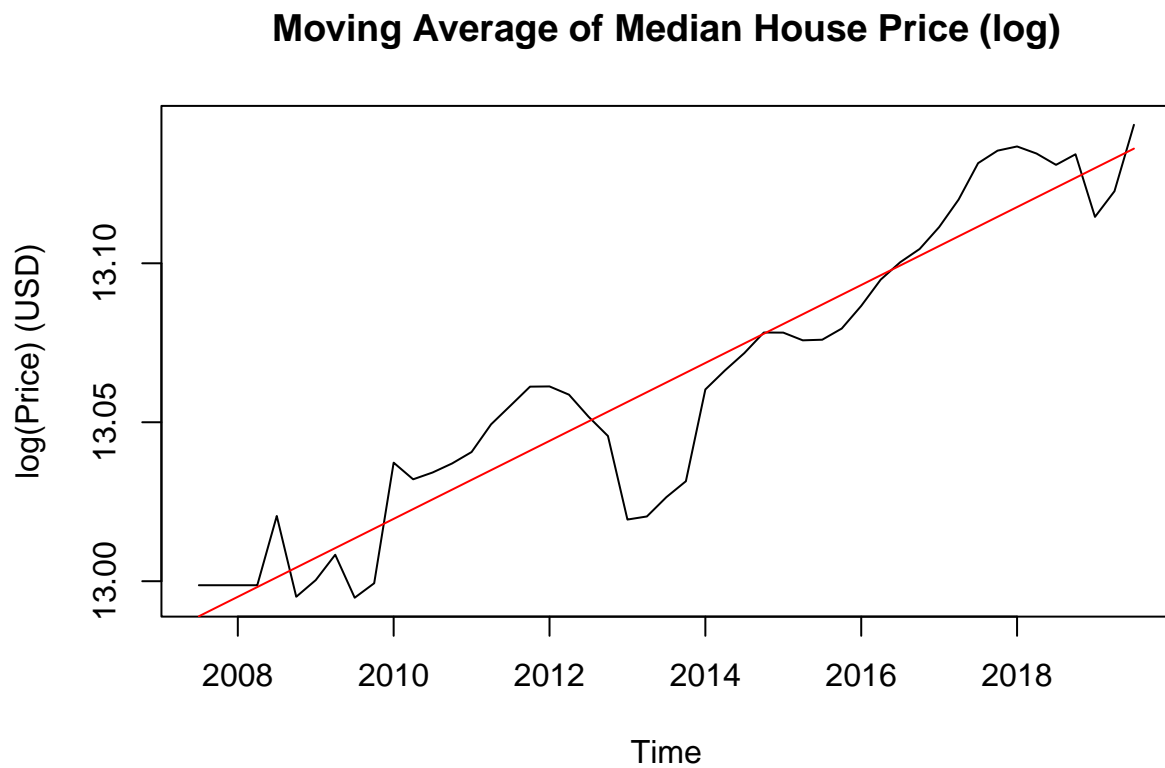
```
##
## Call:
## lm(formula = data_ts ~ t + I(t^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15835.1  -2368.6    432.1   4970.5  11040.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.536e+08  3.635e+08   2.349  0.0232 *
## t           -8.532e+05  3.610e+05  -2.363  0.0224 *
## I(t^2)        2.133e+02  8.965e+01   2.379  0.0215 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7012 on 46 degrees of freedom
## Multiple R-squared:  0.9022, Adjusted R-squared:  0.898
## F-statistic: 212.3 on 2 and 46 DF,  p-value: < 2.2e-16
```

The adjusted R^2 is 0.898, meaning quadratic model explains the variability in the data well. According to

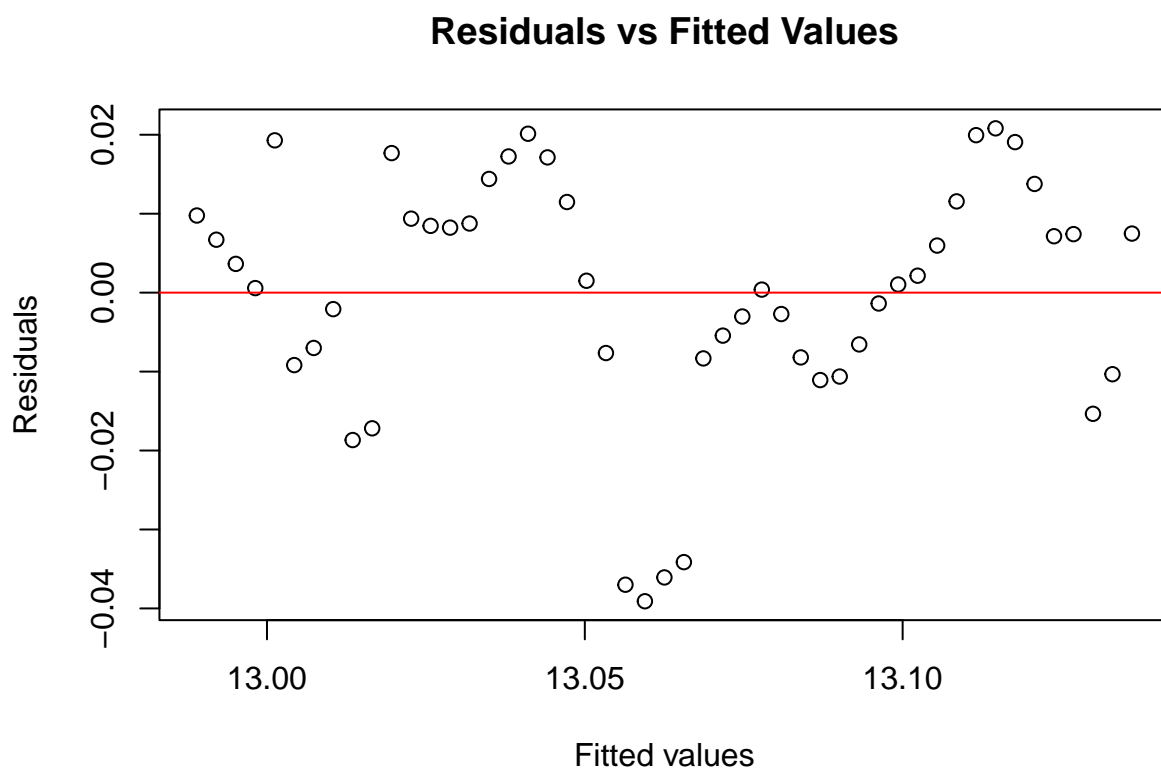
p-value of the t-distribution, both the coefficient and intercept are significant at the 5% significance level. According to the p-value of the F-distribution, the quadratic model is significant in fitting the data.

```
# fit and plot a log-linear model
log_linear_m <- lm(log(data_ts) ~ t)
plot(log(data_ts), ylab = "log(Price) (USD)", main = "Moving Average of Median House Price (log)")
lines(t, log_linear_m$fitted.values, col = "red")
```

Log-linear Model



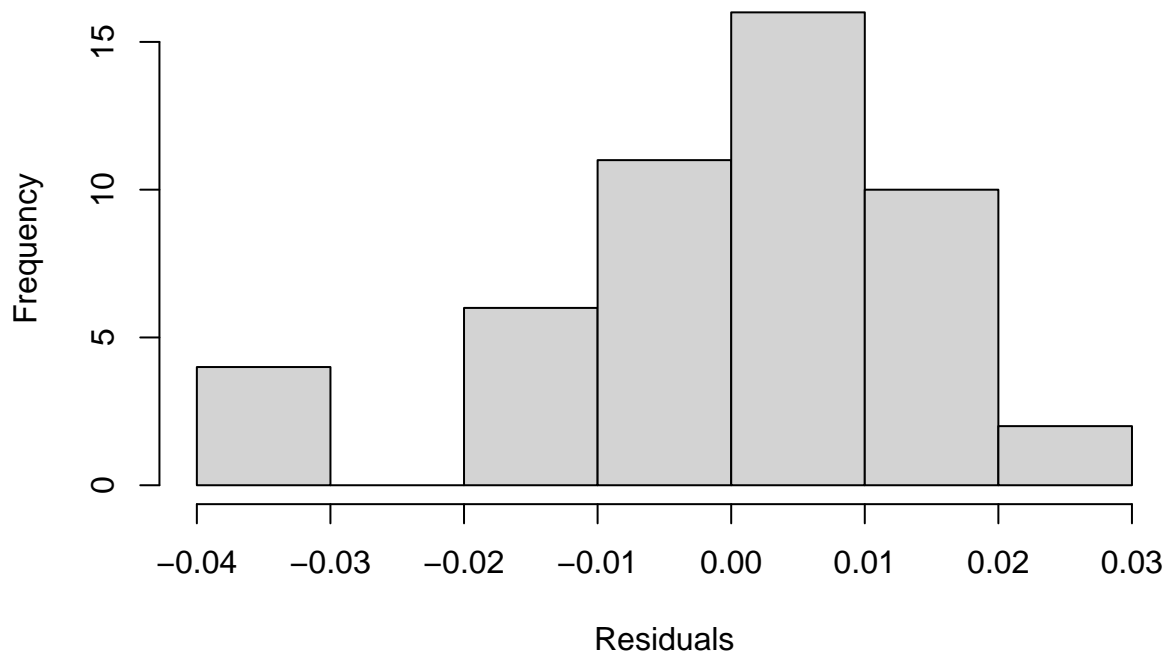
```
# residuals vs fitted values plot
plot(log_linear_m$fitted.values, log_linear_m$residuals,
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```



The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals
hist(log_linear_m$residuals,
     main = "Histogram of Residuals",
     xlab = "Residuals")
```

Histogram of Residuals



The residuals histogram roughly follows a normal distribution, with some residuals on the further left on the negative side, which may potentially be outliers

```
# diagnostics
summary(log_linear_m)
```

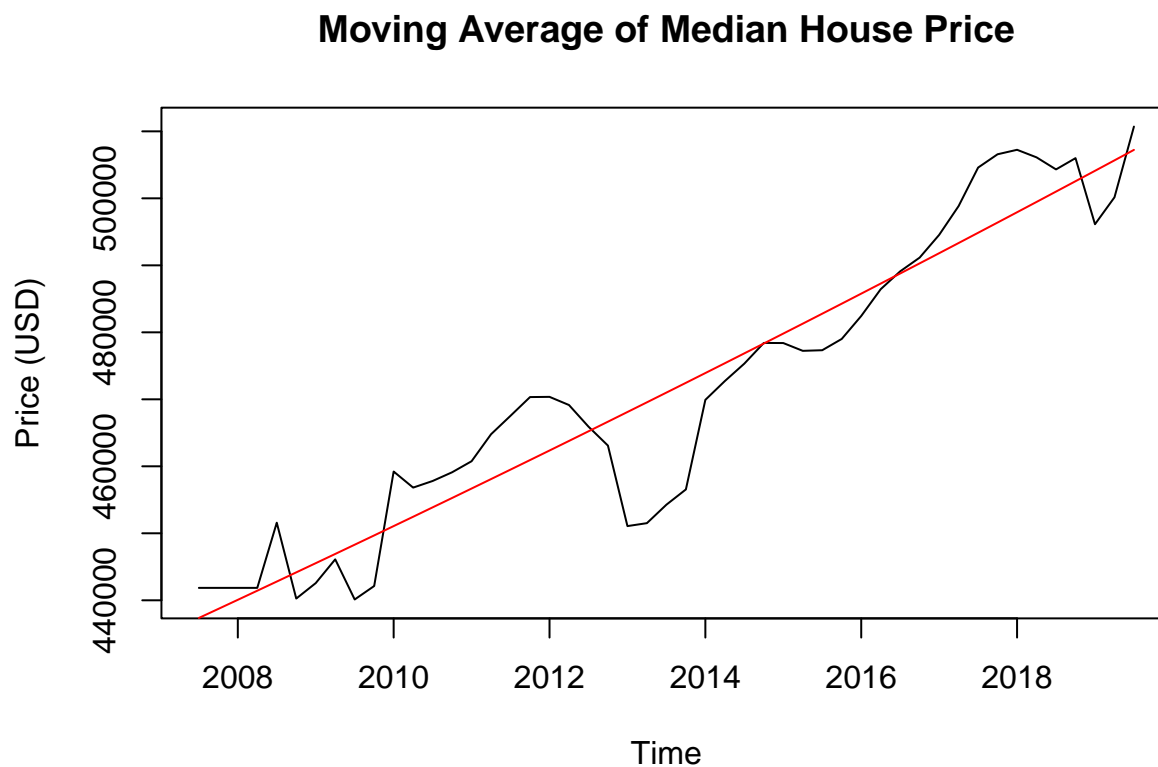
```
##
## Call:
## lm(formula = log(data_ts) ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.039089 -0.008204  0.001506  0.009771  0.020811
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11.621802   1.256512  -9.249 3.74e-12 ***
## t              0.012259   0.000624  19.645 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01544 on 47 degrees of freedom
## Multiple R-squared:  0.8914, Adjusted R-squared:  0.8891
## F-statistic: 385.9 on 1 and 47 DF,  p-value: < 2.2e-16
```

The adjusted R^2 is 0.8891, meaning log-linear model explains the variability in the data well. According to

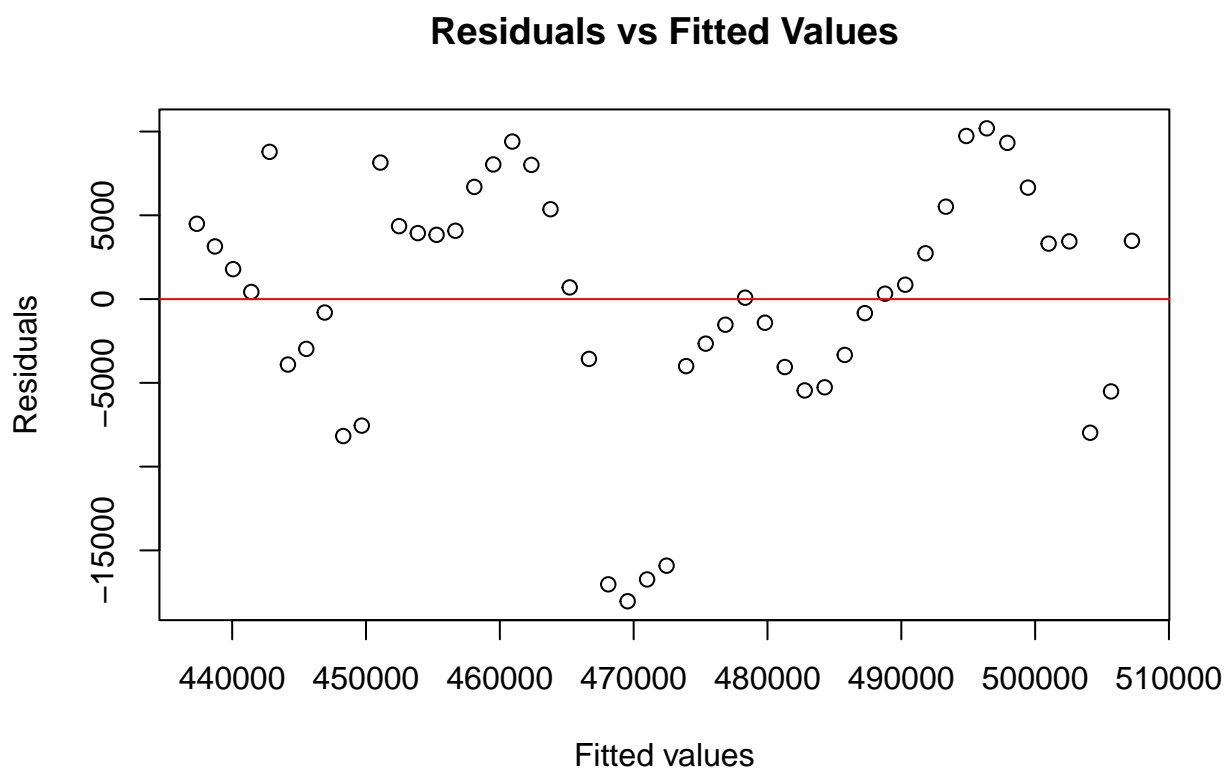
p-value of the t-distribution, both the coefficient and intercept are significant at the 5% significance level. According to the p-value of the F-distribution, the log-linear model is significant in fitting the data.

```
# fit and plot an exponential model
data_ts_df <- data.frame(x = t, y = data_ts)
exponential_m <- nls(y ~ exp(a + b * t), data = data_ts_df,
                     start = list(a = log(min(data_ts)), b = 0.01))
plot(data_ts, ylab = "Price (USD)", main = "Moving Average of Median House Price")
lines(t, predict(exponential_m, list(x = t)), col = "red")
```

Exponential Model



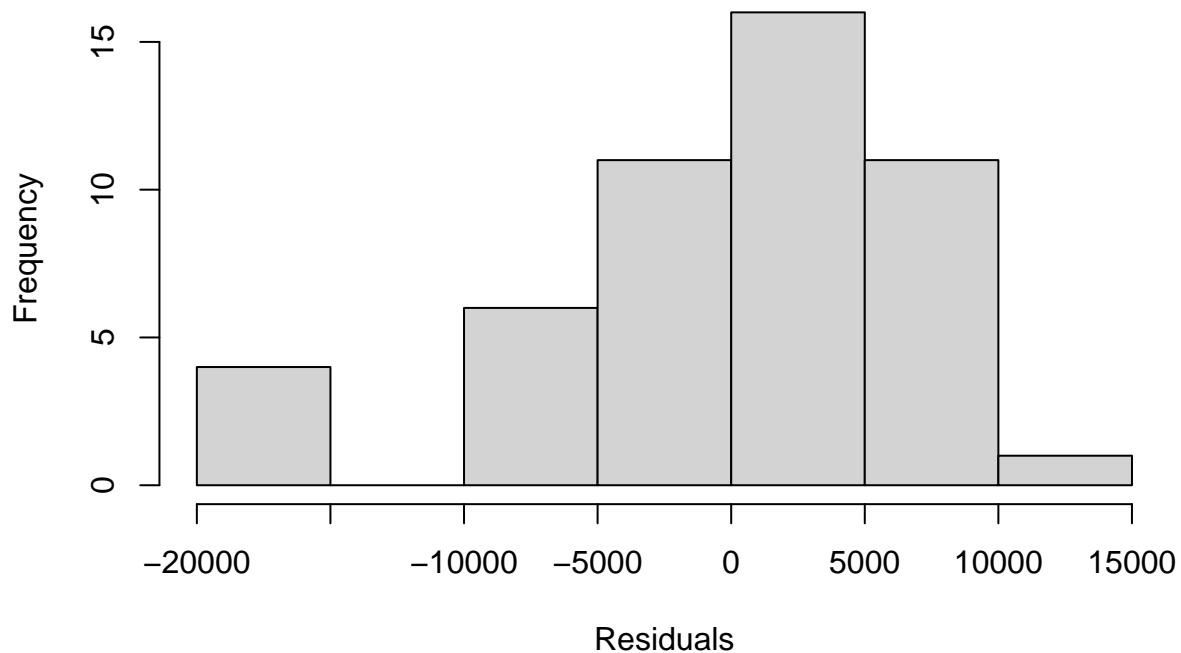
```
# residuals vs fitted values plot
plot(fitted.values(exponential_m), residuals(exponential_m),
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```



The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals
hist(residuals(exponential_m),
     main = "Histogram of Residuals",
     xlab = "Residuals")
```

Histogram of Residuals



The residuals histogram roughly follows a normal distribution with some values present on the further left negative side, which may be an outlier.

```
# diagnostics
```

```
summary(exponential_m)
```

```
##
## Formula: y ~ exp(a + b * t)
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## a -1.181e+01  1.249e+00  -9.454  1.9e-12 ***
## b  1.235e-02  6.202e-04  19.917  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7226 on 47 degrees of freedom
##
## Number of iterations to convergence: 24
## Achieved convergence tolerance: 3.037e-07
```

```
# MSE
```

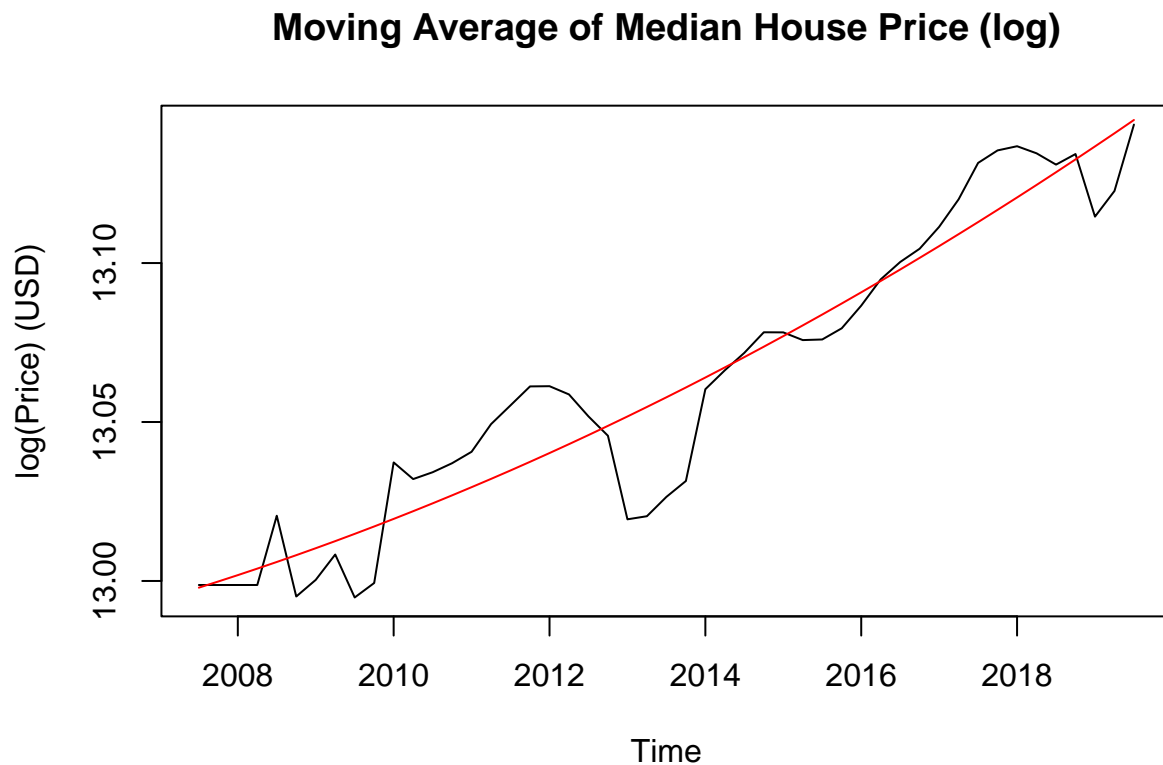
```
MSE <- sum((data_ts - fitted(exponential_m))^2) / length(data_ts)
MSE
```

```
## [1] 50087470
```

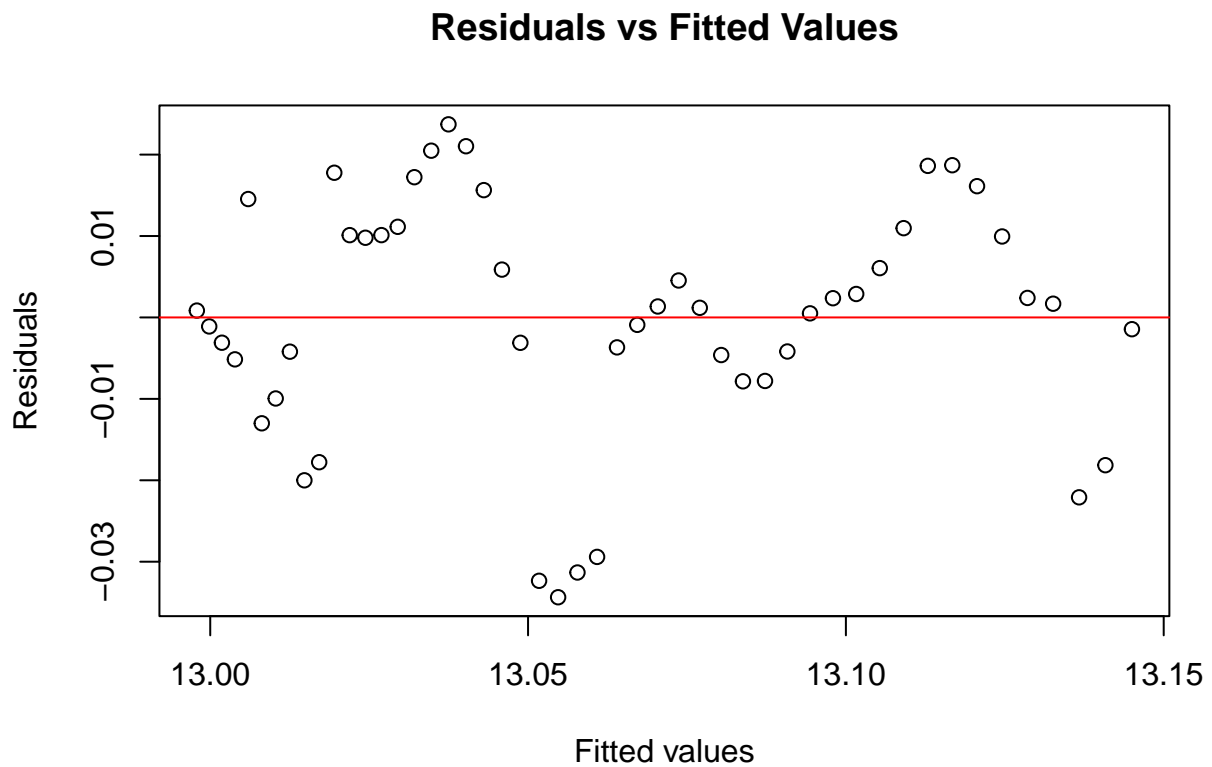
Since the exponential regression is not linear, R^2 is not appropriate here. Instead, we assess the model by its mean squared error (MSE). The MSE of the exponential fit is 50087470, which is a high value, indicating exponential may not be a good fit for the data. According to p-value of the t-distribution, both the **a** and **b** are significant at the 5% significance level.

```
# fit and plot a log-quadratic model
log_quadratic_m <- lm(log(data_ts) ~ t + I(t^2))
plot(log(data_ts), ylab = "log(Price) (USD)", main = "Moving Average of Median House Price (log)")
lines(t, log_quadratic_m$fitted.values, col = "red")
```

Log-quadratic



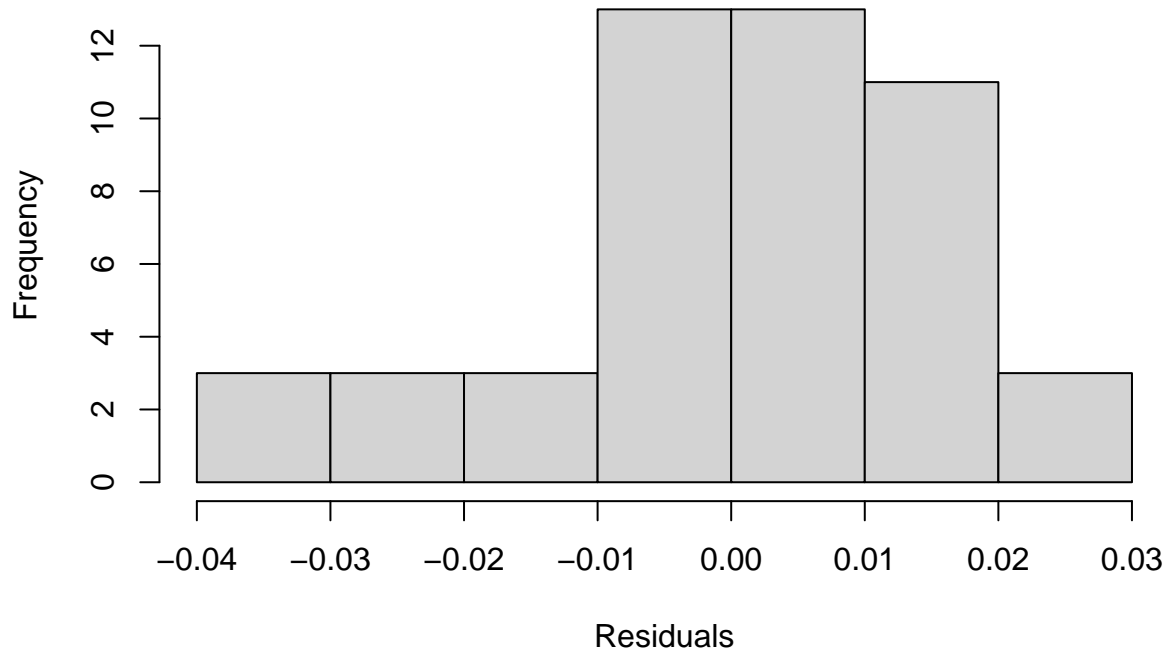
```
# residuals vs fitted values plot
plot(log_quadratic_m$fitted.values, log_quadratic_m$residuals,
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```

The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals  
hist(log_quadratic_m$residuals,  
      main = "Histogram of Residuals",  
      xlab = "Residuals")
```

Histogram of Residuals



The residuals histogram roughly follows a normal distribution and is left skewed. Some residuals on the further left may potentially be outliers.

```
# diagnostics
summary(log_quadratic_m)
```

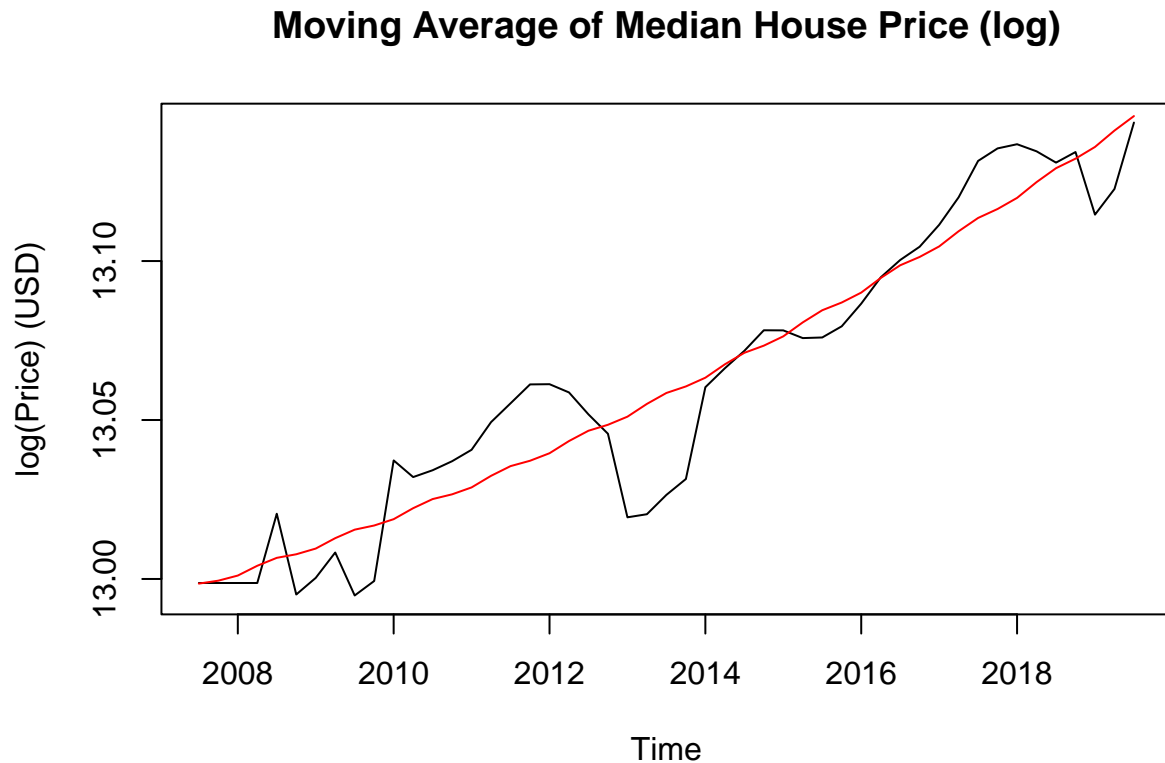
```
##
## Call:
## lm(formula = log(data_ts) ~ t + I(t^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.034364 -0.005147  0.001183  0.010112  0.023716
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.529e+03  7.767e+02   1.968  0.0551 .
## t            -1.518e+00  7.715e-01  -1.967  0.0552 .
## I(t^2)         3.800e-04  1.916e-04   1.983  0.0533 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01498 on 46 degrees of freedom
## Multiple R-squared:  0.9, Adjusted R-squared:  0.8956
## F-statistic: 207 on 2 and 46 DF, p-value: < 2.2e-16
```

The adjusted R^2 is 0.8956, meaning log quadratic model explains the variability in the data well. According

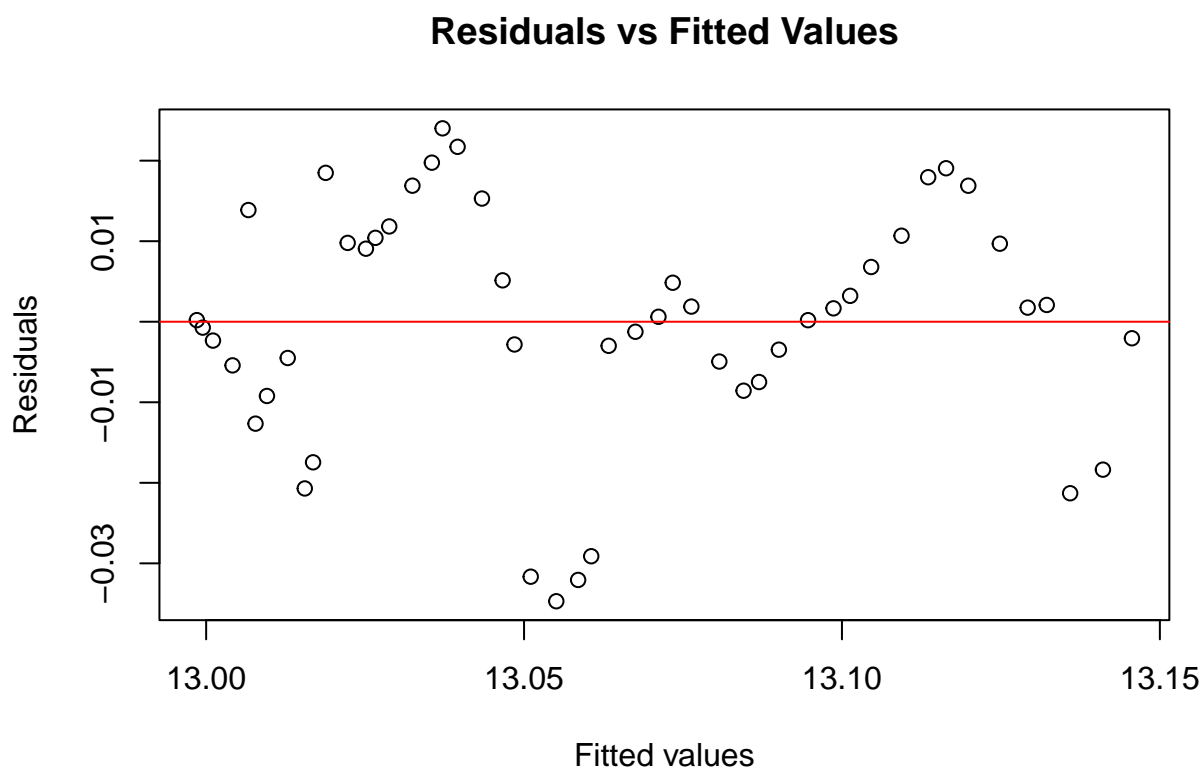
to p-value of the t-distribution, none of the coefficient and intercept are significant at the 5% significance level, but are significant at the 10% significance level. However, according to the p-value of the F-distribution, the log quadratic model is significant in fitting the data.

```
sin_t <- sin(2 * pi * t)
cos_t <- cos(2 * pi * t)
log_quadratic_periodic_m <- lm(log(data_ts) ~ t + I(t^2) + sin_t + cos_t)
plot(log(data_ts), ylab = "log(Price) (USD)", main = "Moving Average of Median House Price (log)")
lines(t, log_quadratic_periodic_m$fitted.values, col = "red")
```

Log-quadratic-periodic Model



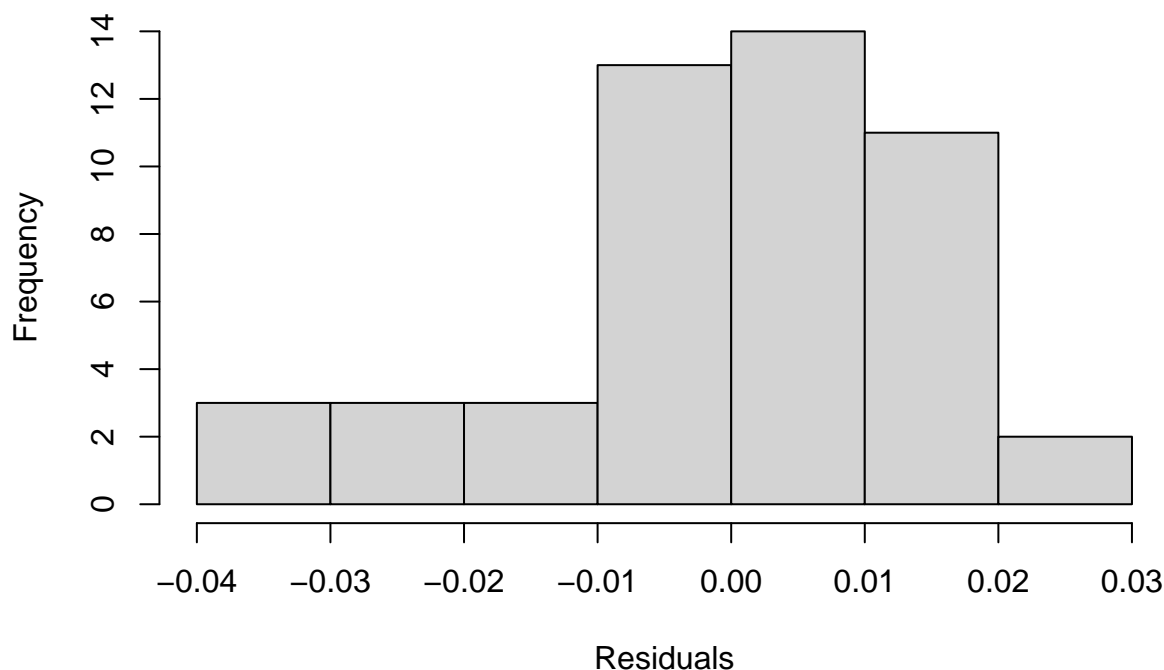
```
# residuals vs fitted values plot
plot(log_quadratic_periodic_m$fitted.values, log_quadratic_periodic_m$residuals,
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```



The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals  
hist(log_quadratic_periodic_m$residuals,  
      main = "Histogram of Residuals",  
      xlab = "Residuals")
```

Histogram of Residuals



The residuals histogram roughly follows a normal distribution and is left skewed. Some residuals on the further left may potentially be outliers.

```
# diagnostics
summary(log_quadratic_periodic_m)

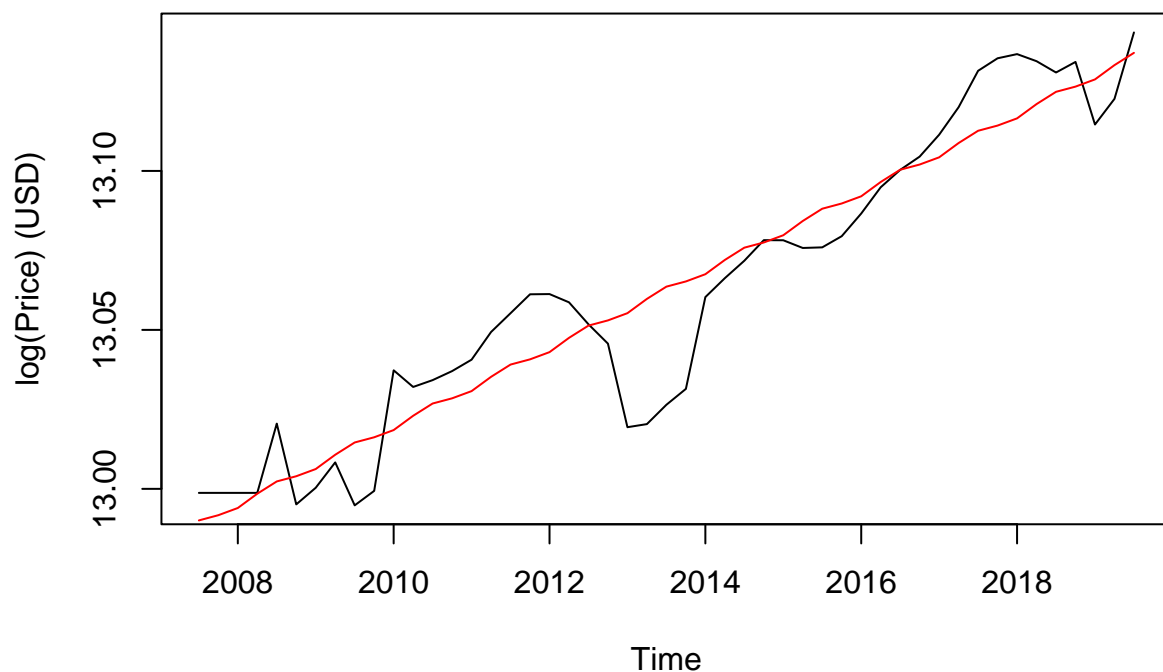
##
## Call:
## lm(formula = log(data_ts) ~ t + I(t^2) + sin_t + cos_t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.034705 -0.005418  0.000617  0.010419  0.024012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.516e+03  7.954e+02   1.906   0.0632 .
## t            -1.505e+00  7.901e-01  -1.905   0.0633 .
## I(t^2)        3.769e-04  1.962e-04   1.921   0.0612 .
## sin_t         3.163e-04  3.129e-03   0.101   0.9199
## cos_t        -7.114e-04  3.070e-03  -0.232   0.8178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01531 on 44 degrees of freedom
## Multiple R-squared:  0.9001, Adjusted R-squared:  0.8911
## F-statistic: 99.15 on 4 and 44 DF,  p-value: < 2.2e-16
```

The adjusted R^2 is 0.8911, meaning log quadratic periodic model explains the variability in the data well. According to p-value of the t-distribution, none of the coefficient and intercept are significant at the 5% significance level, but the intercept, t , and t^2 are significant at the 10% significance level. According to the p-value of the F-distribution, the log quadratic periodic model is significant in fitting the data.

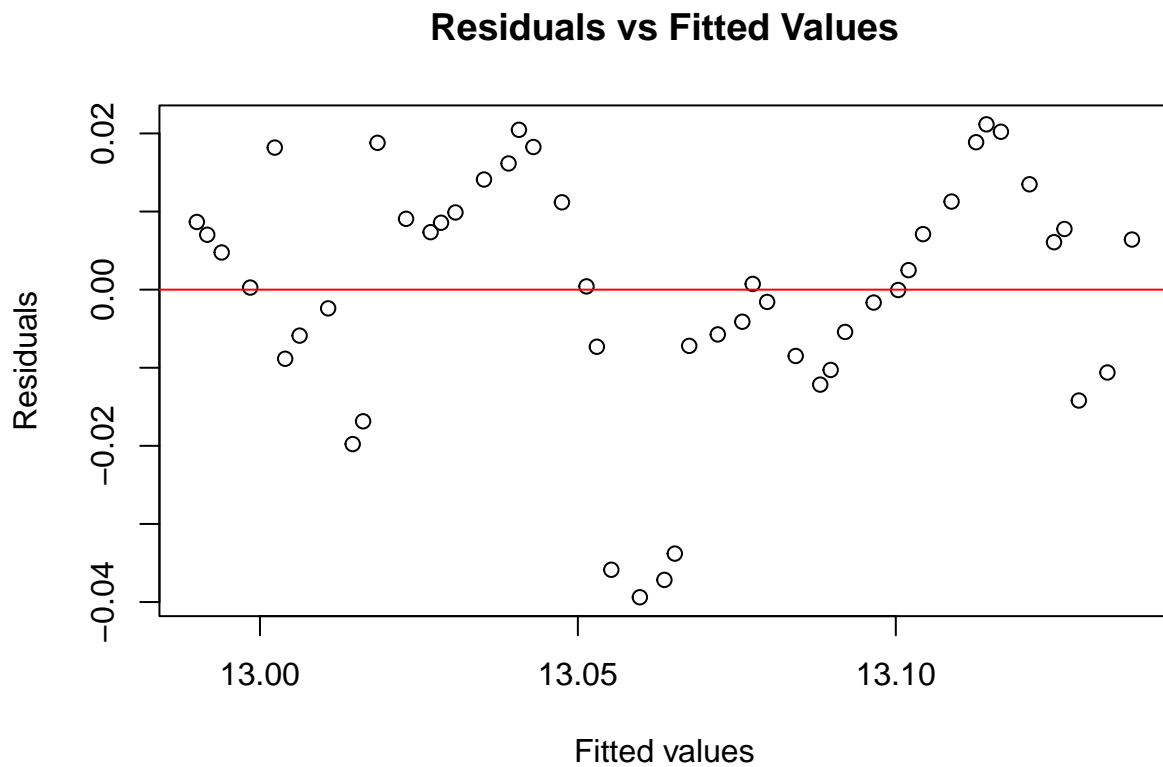
```
sin_t <- sin(2 * pi * t)
cos_t <- cos(2 * pi * t)
log_linear_periodic_m <- lm(log(data_ts) ~ t + sin_t + cos_t)
plot(log(data_ts), ylab = "log(Price) (USD)", main = "Moving Average of Median House Price (log)")
lines(t, log_linear_periodic_m$fitted.values, col = "red")
```

Log-linear-periodic Model

Moving Average of Median House Price (log)



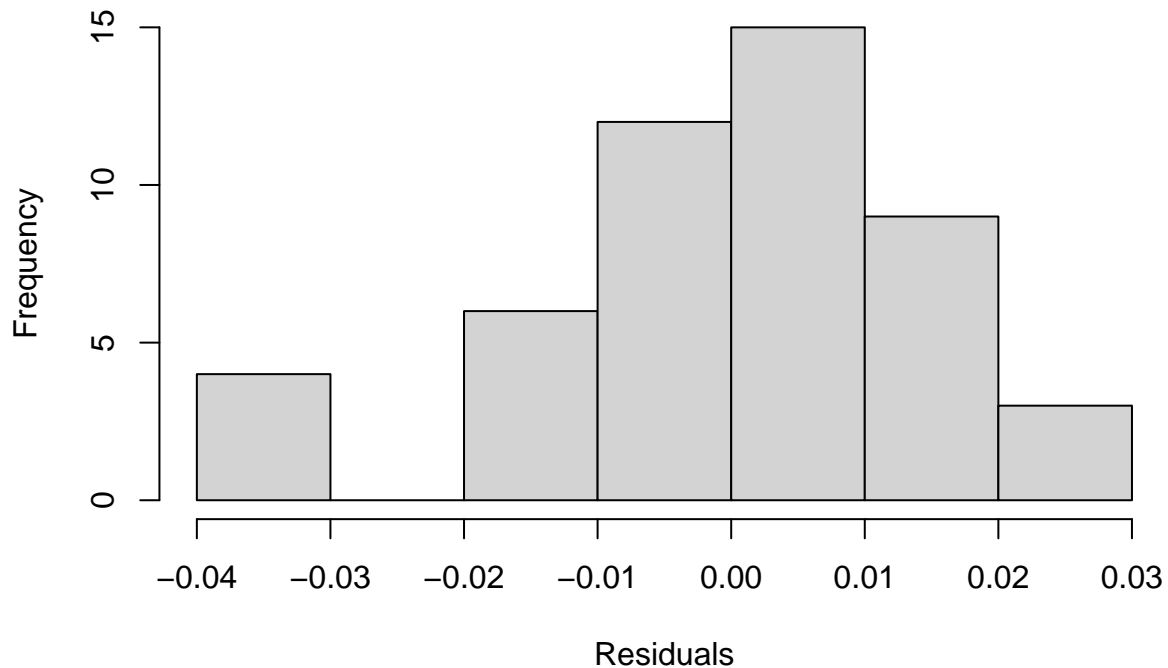
```
# residuals vs fitted values plot
plot(log_linear_periodic_m$fitted.values, log_linear_periodic_m$residuals,
     main = "Residuals vs Fitted Values",
     xlab = "Fitted values",
     ylab = "Residuals")
abline(h = 0, col = "red")
```



The residuals showcase an up and down pattern, similar to a sine and cosine wave. This indicates some seasonality in the data, as the data fluctuates along the growing trend.

```
# histogram of residuals  
hist(log_linear_periodic_m$residuals,  
      main = "Histogram of Residuals",  
      xlab = "Residuals")
```

Histogram of Residuals



The residuals histogram roughly follows a normal distribution with some values present on the further left negative side, which may be an outlier.

```
# diagnostics
summary(log_linear_periodic_m)
```

```
##
## Call:
## lm(formula = log(data_ts) ~ t + sin_t + cos_t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.039384 -0.007312  0.000728  0.009883  0.021163
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.162e+01  1.284e+00  -9.048 1.08e-11 ***
## t            1.226e-02  6.376e-04  19.223 < 2e-16 ***
## sin_t        3.163e-04  3.221e-03   0.098  0.922
## cos_t       -1.111e-03  3.153e-03  -0.352  0.726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01576 on 45 degrees of freedom
## Multiple R-squared:  0.8918, Adjusted R-squared:  0.8845
## F-statistic: 123.6 on 3 and 45 DF,  p-value: < 2.2e-16
```


The adjusted R^2 is 0.8845, meaning log quadratic periodic model explains the variability in the data well. According to p-value of the t-distribution, the coefficient of t and intercept are significant at the 5% significance level. According to the p-value of the F-distribution, the log linear periodic model is significant in fitting the data.

(h)

```
AIC(linear_m, log_linear_m, quadratic_m, log_quadratic_m, exponential_m,
    log_quadratic_periodic_m, log_linear_periodic_m)
```

```
##                df        AIC
## linear_m        3 1015.4733
## log_linear_m     3 -265.6967
## quadratic_m      4 1011.7855
## log_quadratic_m  4 -267.7172
## exponential_m    3 1013.7908
## log_quadratic_periodic_m 6 -263.7883
## log_linear_periodic_m 5 -261.8422
```

```
BIC(linear_m, log_linear_m, quadratic_m, log_quadratic_m, exponential_m,
    log_quadratic_periodic_m, log_linear_periodic_m)
```

```
##                df        BIC
## linear_m        3 1021.1488
## log_linear_m     3 -260.0212
## quadratic_m      4 1019.3528
## log_quadratic_m  4 -260.1499
## exponential_m    3 1019.4662
## log_quadratic_periodic_m 6 -252.4374
## log_linear_periodic_m 5 -252.3831
```

Both AIC and BIC indicate that the log quadratic model is the most suitable.

(i)

```
# plot forecast for 24 steps (quarters) ahead
tn <- data.frame(t = seq(from = 2019.75, length.out = 24, by = 0.25))
pred <- predict(lm(log(data_ts) ~ t + I(t^2)), tn, se.fit = TRUE)
plot(c(time(data_ts), tn$t), c(log(data_ts), pred$fit), type = 'l', xlim = c(2007, 2026),
     ylab = "Price (USD)", xlab = "Time", main = "Moving Average of Median House Price (log)")
lines(tn$t, pred$fit, col = "red")

# compute the 95% confidence and prediction intervals
pred.plim <- predict(lm(log(data_ts) ~ t + I(t^2)), tn, level = 0.95, interval = "prediction")
pred.clim <- predict(lm(log(data_ts) ~ t + I(t^2)), tn, level = 0.95, interval = "confidence")

# shade the confidence interval
polygon(c(tn$t, rev(tn$t)),
       c(pred.clim[,3], rev(pred.clim[,2])),
```

```

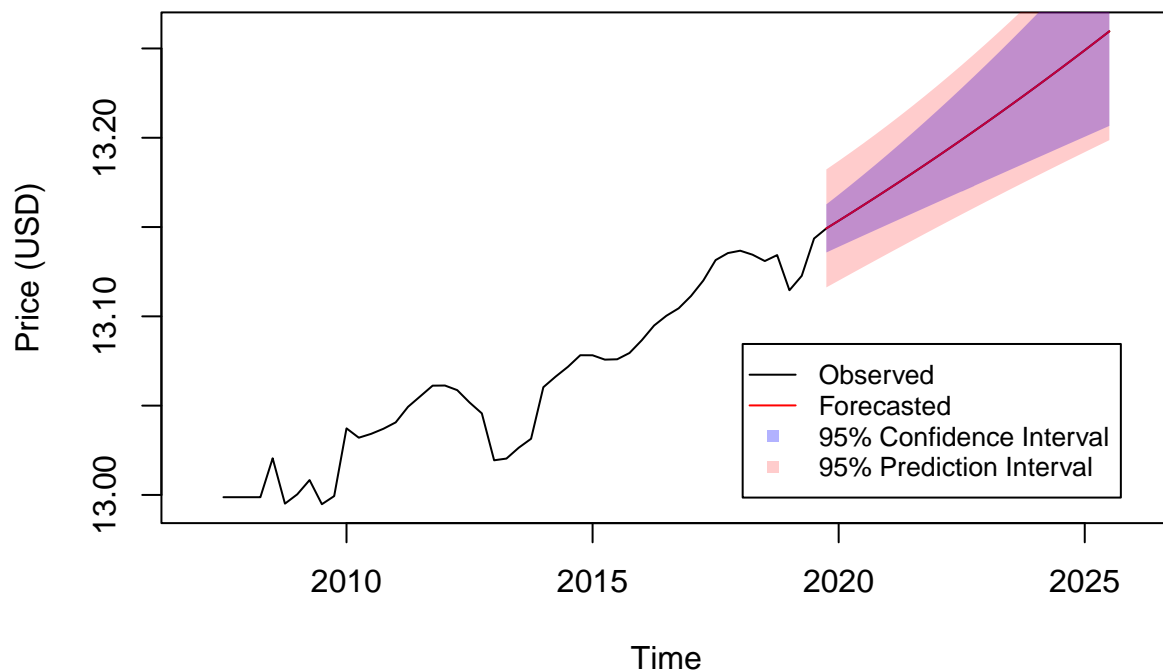
col = rgb(0, 0, 1, 0.3), border = NA)

# shade the prediction interval
polygon(c(tn$t, rev(tn$t)),
       c(pred.plim[,3], rev(pred.plim[,2])),
       col = rgb(1, 0, 0, 0.2), border = NA)

# add legend
legend("bottomright", legend = c("Observed", "Forecasted", "95% Confidence Interval",
                                "95% Prediction Interval"),
      col = c("black", "red", rgb(0, 0, 1, 0.3), rgb(1, 0, 0, 0.2)), lty = c(1, 1, NA, NA),
      pch = c(NA, NA, 15, 15), cex = 0.8, inset = c(0.05, 0.05))

```

Moving Average of Median House Price (log)



```

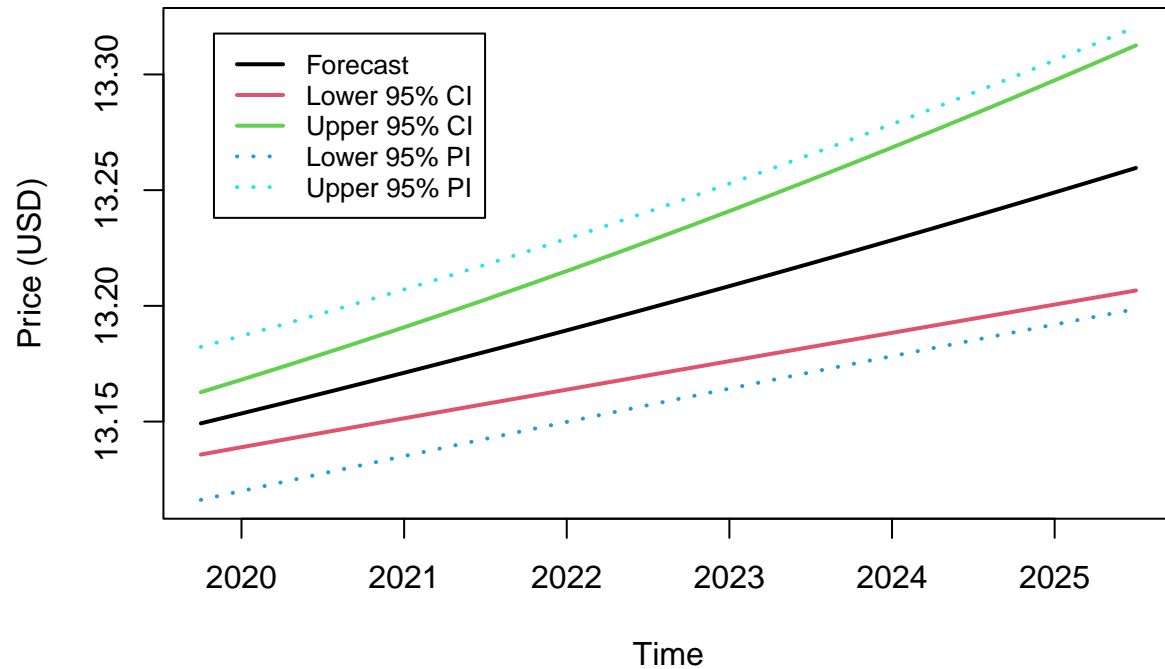
# plot confidence and prediction interval of forecast
matplot(tn$t, cbind(pred.clim, pred.plim[, -1]),
      lty = c(1, 1, 1, 3, 3), type = "l", lwd = 2, ylab = "Price (USD)", xlab = "Time",
      main = "Forecast of Moving Average of Median House Price")

# add legend
legend("topleft",
      legend = c("Forecast", "Lower 95% CI", "Upper 95% CI", "Lower 95% PI", "Upper 95% PI"),
      lty = c(1, 1, 1, 3, 3),
      lwd = 2,
      col = 1:5,
      cex = 0.8,

```

```
inset = c(0.05, 0.05))
```

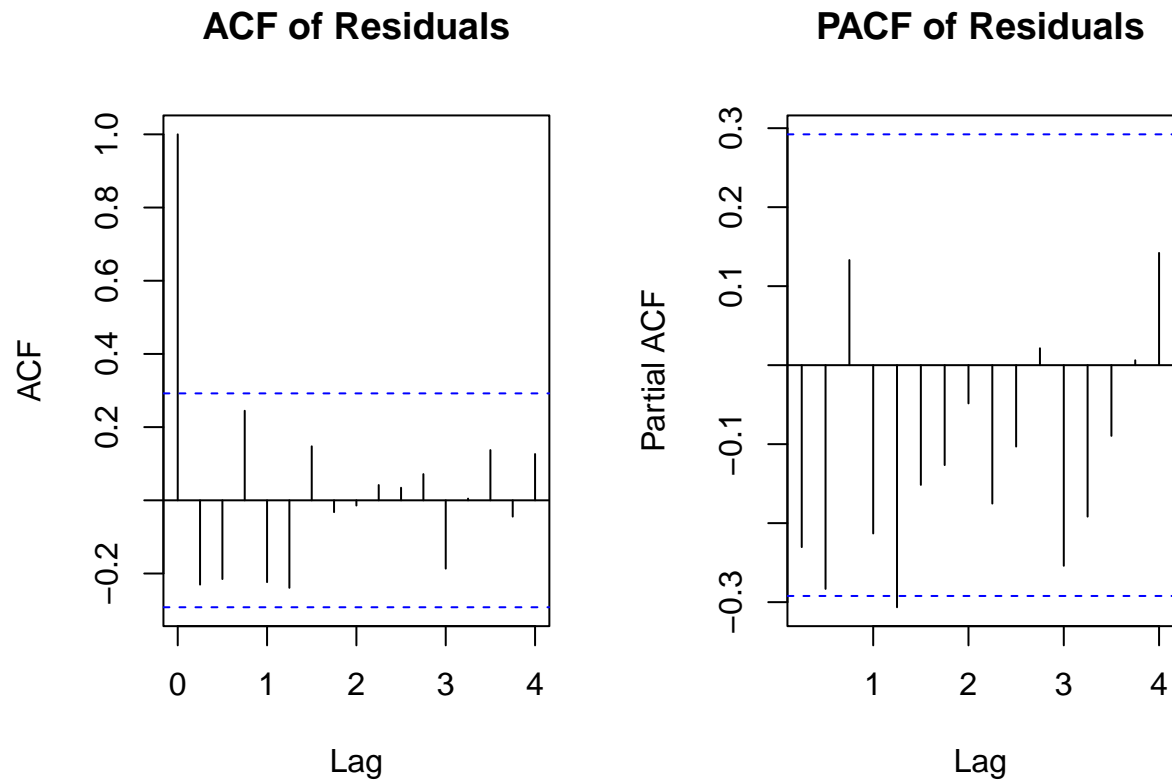
Forecast of Moving Average of Median House Price



2. Trend and Seasonal Adjustments

(a)

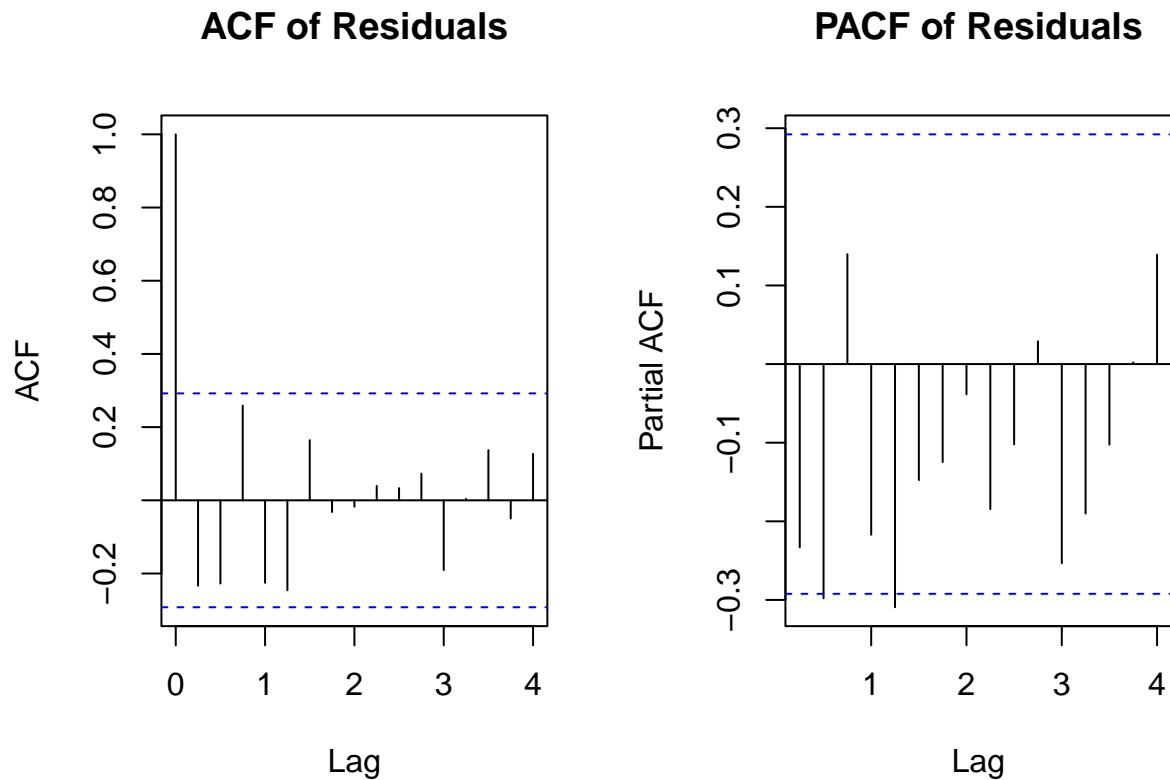
```
# additive decomposition
additive_decomp <- decompose(data_ts, type = "additive")
par(mfrow=c(1,2))
acf(na.omit(additive_decomp$random), main = "ACF of Residuals")
pacf(na.omit(additive_decomp$random), main = "PACF of Residuals")
```



On the ACF, the residuals do not showcase autocorrelation across different lags. On the PACF, the residual is only significant at lag 1.25, while others are insignificant, meaning the residuals are mainly random white noise and do not have a regular pattern.

(b)

```
# multiplicative decomposition
mult_decomp <- decompose(data_ts, type = "multiplicative")
par(mfrow=c(1,2))
acf(na.omit(mult_decomp$random), main = "ACF of Residuals")
pacf(na.omit(mult_decomp$random), main = "PACF of Residuals")
```



On the ACF, the residuals do not showcase autocorrelation across different lags. On the PACF, the residual is only significant at lag 0.5 and 1.25, while others are insignificant, meaning the residuals are mainly random white noise and do not have a regular pattern.

(c)

The ACF and PACF plots of the residuals of both additive and multiplicative decompositions are very similar. Both of the residuals of the additive and multiplicative methods are insignificant at most lags, indicating randomness and white noise. However, the PACF plot of the residuals of the additive decomposition has one less significant spike at a lag. So, additive is the better method here. Also, additive method is more straightforward to compute.

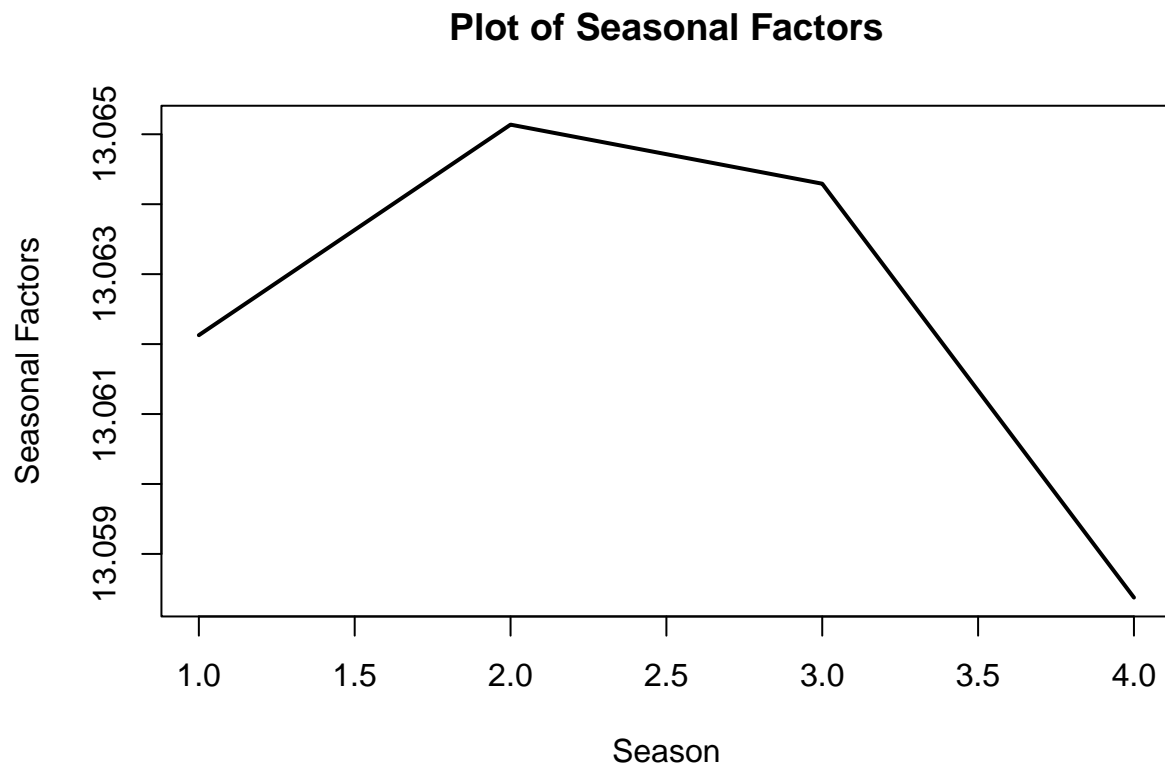
(d)

The models for the additive and multiplicative cycles would be similar, since their ACF and PACF plots are very similar. This means the residuals of both methods are mostly random (white noise) and do not have significant correlations at almost all the lags.

(e)

```
# fit seasonal factors
fit <- tslm(log(data_ts) ~ season + 0)
```

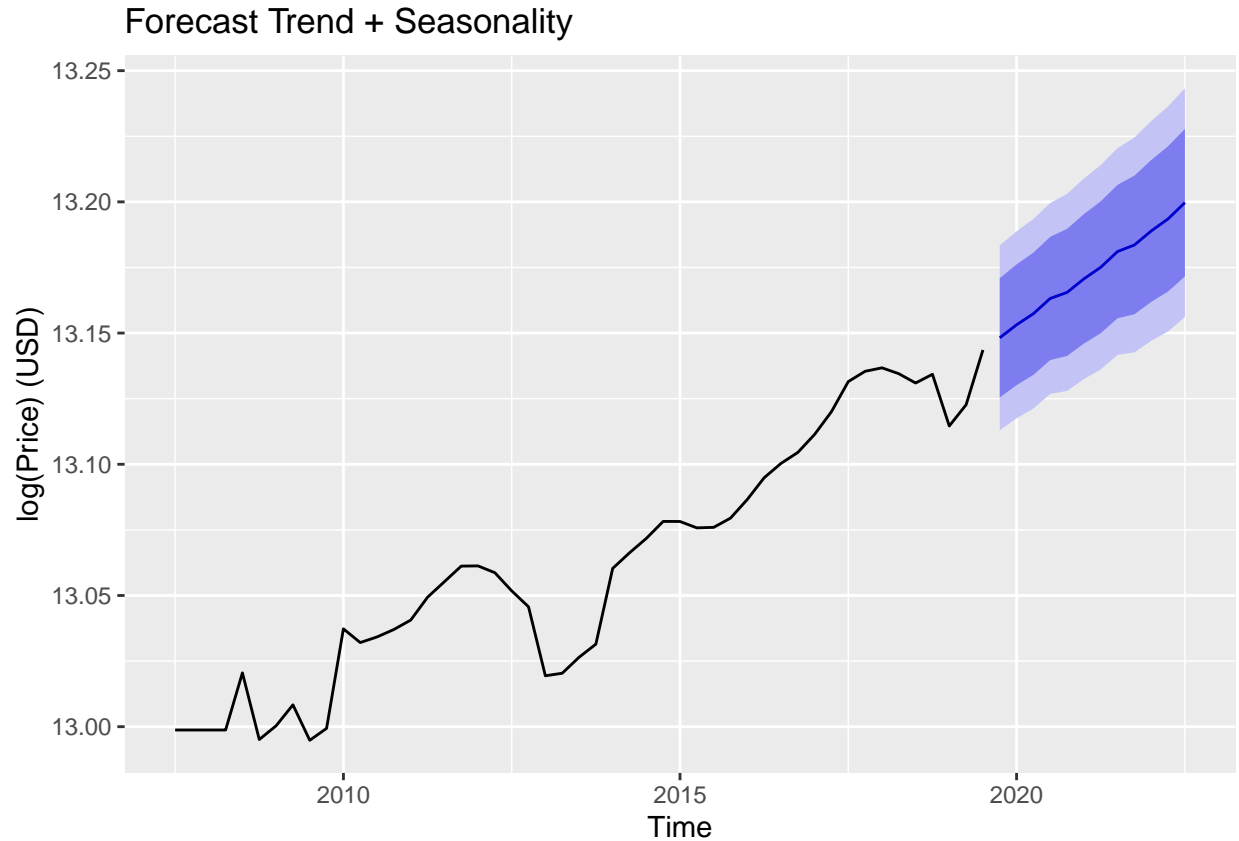
```
# plot of coefficient of seasonal factors
plot(fit$coef, type = 'l', ylab = 'Seasonal Factors', xlab = "Season", lwd = 2,
     main = "Plot of Seasonal Factors")
```



According to the plot of seasonal factors, there is an increase in seasonal effects from the season 1 to 2, and a peak at season 2, indicating high seasonal effects in the second quarter. After the second quarter, the seasonal effects decrease, with the fourth quarter having the lowest seasonal effects.

(f)

```
# fit and plot log quadratic model with seasonal component
log_quadratic_seasonal_m <- tslm(log(data_ts) ~ trend + I(trend^2) + season + 0)
autoplot(forecast(log_quadratic_seasonal_m, h = 12), ylab = "log(Price) (USD)",
         main = "Forecast Trend + Seasonality")
```



III. Conclusions and Future Work

Conclusions:

The report identifies the log-quadratic model as the most suitable for forecasting house price trends, as supported by its superior performance based on both AIC and BIC values. The model provides a reasonable forecast with confidence and prediction intervals, capturing the long-term trend and seasonality in the data.

1. Model Selection:

- The log-quadratic model with seasonal adjustments performs best due to its ability to capture the time series data's trend while incorporating seasonality.

2. Residual Analysis:

- Residual patterns indicate some periodic seasonality and possible outliers, but they mostly resemble white noise after incorporating seasonality, confirming the model's adequacy.

3. Forecasting Results:

- The model generates a reliable forecast for 24 quarters ahead with well-defined confidence and prediction intervals. These intervals provide a clear range of potential outcomes.

Future Work:

1. Handling Residual Skewness and Outliers:

- The residual histogram showed slight skewness and potential outliers on the left tail. Removing or accounting for these outliers could refine the model's accuracy.

2. **Exploration of Alternative Models:**

- Seasonal ARIMA or exponential smoothing methods could be tested to verify if they outperform the log-quadratic model in capturing both short-term fluctuations and long-term trends.

3. **Dynamic Seasonality:**

- Seasonality was modeled additively, but exploring dynamic seasonal components might better capture variations in seasonal intensity over time.

4. **External Factors:**

- Incorporating external variables, such as economic indicators (e.g., interest rates, inflation), could enhance the model's explanatory power.

5. **Regularization Techniques:**

- To avoid overfitting, especially in periodic models with higher complexity, ridge regression or lasso could be applied.

IV. References

Data Source:

https://www.kaggle.com/datasets/htagholdings/property-sales?resource=download&select=ma_lga_12345.csv