TIME SERIES ANALYSIS FOR HUNGARY CHICKEN POX CASES

PROF. Zinovy Radovilsky

Ruchika Balani (ac9968)
Aishwarya Khinvasara (xv6580)
Parthiv Patel (hk9113)
Sphurti Pendke (fq2684)
Mengchi Xie (kd8118)

INTRODUCTION

Our dataset contains weekly reports of chickenpox cases in Hungary from 2005 to 2014, which we collected from www.kaggle.com. The original data includes county-level adjacency matrix and time series of county-level cases. However, to simplify our time series analysis, we aggregated the county level cases to country level and monthly data. Aggregating the data allows for more accurate results and easier identification of trends and patterns. We converted the data into a time series and divided it into training, validation, and future partitions. Our goal is to identify the best forecasting method to predict monthly chickenpox cases in Hungary for the 12 months of 2015. To achieve this, we first analyzed the data components and will explore various forecasting methods such as Two-level forecasting, Regression model with trend and seasonality, and Auto ARIMA model, along with baseline models like naïve and seasonal naïve. We will compare the accuracy measures of these models, including MAPE and RMSE, to identify the best model. Once we have the best model, we will apply it to predict the cases for 2015.

EIGHT STEPS OF FORECASTING

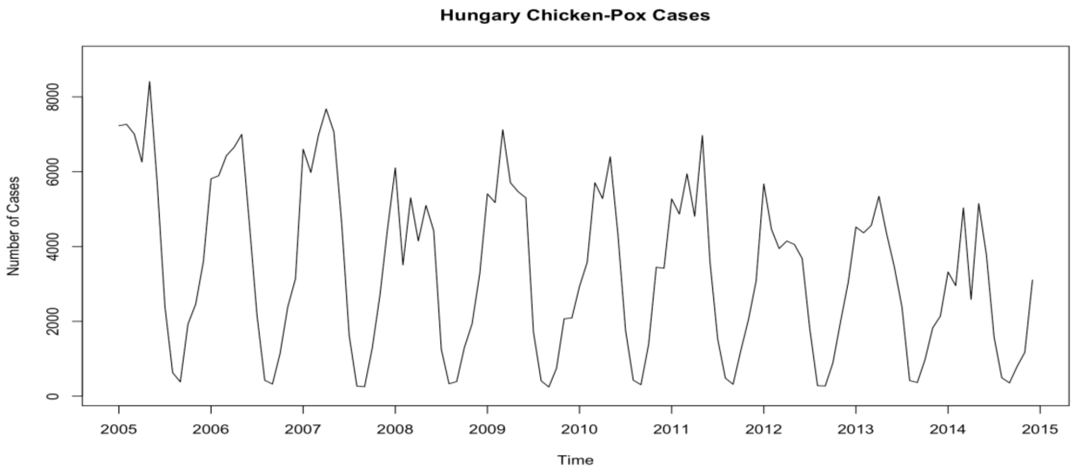
Step 1: Define Goal:

The objective of this project is to generate numerical forecasts for chickenpox cases in 2015. To achieve this, we aim to create a predictive model that considers the trend and seasonal components of the historical data and can effectively forecast for the year 2015. Our focus is on identifying the most accurate model among the available options. To develop these forecasting models, we utilized the R language.

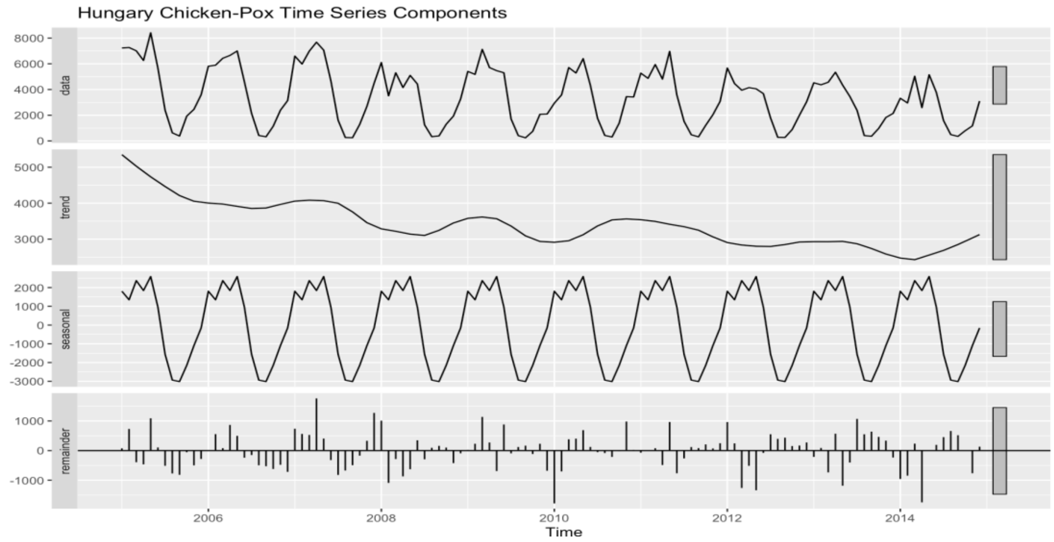
Step 2: Get Data:

We obtained the dataset for chickenpox cases from www.kaggle.com, covering the period from January 2005 to December 2014, reported monthly. We partitioned the data into three parts: training, validation, and future. The training period includes data from January 2005 to December 2012, the validation period includes data from January 2013 to December 2014, and the future period includes the year 2015.

Step 3: Explore and Visualize Series:

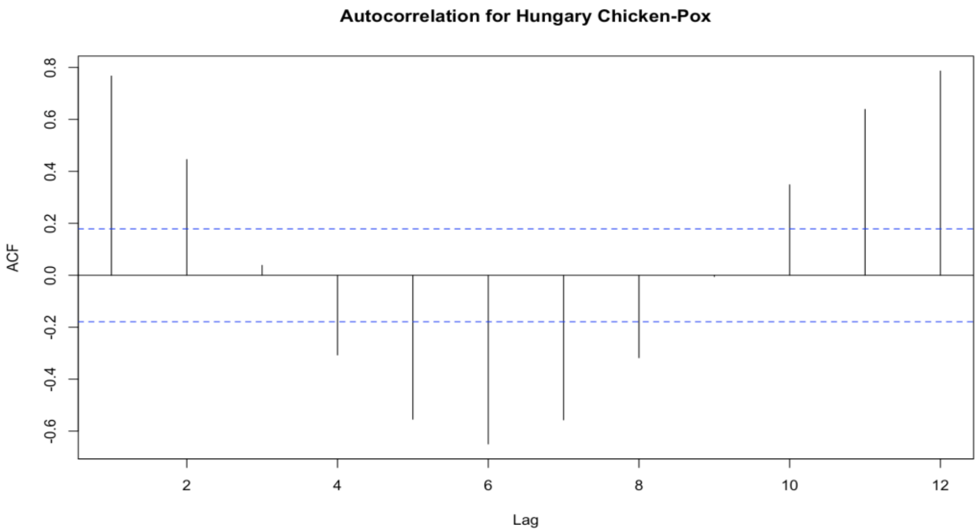


The plot above displays the number of chickenpox cases from 2005 to 2014, showing a decreasing trend over the years. Additionally, there is a clear seasonality pattern, with the highest number of cases occurring from around March to May each year, and the lowest number of cases observed from approximately June to August.



1. The initial chart labeled "data" displays historical data spanning from 2005 to 2014.
2. The second chart reveals a decreasing trend from the start of 2005, which sharply declines in 2014. This trend is considered multiplicative since the rate of reduction in the number of chickenpox cases per year remains constant.
3. The third chart, labeled as the "seasonal" chart, illustrates a cyclic pattern of high and low values that repeats annually. At the beginning of the year, the number of cases tends to be low, but it gradually increases and reaches its peak in May or June before decreasing towards the end of the year. This cyclic pattern repeats every year, indicating seasonality. The seasonality observed in this chart is additive in nature, implying that the variations have a stable magnitude.

4. Finally, the "Remainder" chart displays two components: a systematic component of level and a non-systematic component of noise. The level represents the long-term average of the data, while the noise comprises all the random fluctuations observed in the historical data.



The highest positive auto correlation coefficient of 0.786 is at lag 12 which suggests monthly seasonality in the time series data. The second largest positive auto correlation coefficient occurs at a lag of 1 indicating the trend component in the time series data. In this scenario, we observe decreasing auto correlation coefficients, which indicate the presence of a level component in the time series data. The magnitude of the ACF value of -0.649 at lag 6 indicates that this negative correlation is moderately strong between the time series and its value six-time steps in the past. There is no significant autocorrelation at lag 6 and lag 9.

PREDICTIBILITY OF DATASET:

```
Series: hungary_chickenpox.ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1       mean
      0.7802  3482.5787
s.e.  0.0572   565.5372

sigma^2 = 1990058:  log likelihood = -1039.95
AIC=2085.91   AICc=2086.11   BIC=2094.27

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE     MASE      ACF1
Training set -36.85291 1398.889 1118.882 -62.66861 86.42157 1.560343 0.260635
```

The equation of the AR (1) model for entire data set is:

$$\mathbf{y_t = 3482.5787 + 0.7802y_{t-1}}$$

This is an autoregressive model, where the mean of 3482.5787 represents the intercept (or alpha), and ar1 of 0.7802 represents the beta one value (i.e., the coefficient of the predictor).

```
> z.stat
[1] -3.842657
```

```
> p.value
[1] 6.085467e-05
```

```
> if (p.value<alpha) {
+   "Reject null hypothesis"
+ } else {
+   "Accept null hypothesis"
+ }
[1] "Reject null hypothesis"
```

Our goal is to test whether our dataset is predictable or random using hypothesis testing. We set the slope coefficient (β) to 1 and create a null hypothesis (H0: β = 1) that assumes the data is a random walk. The alternative hypothesis (H1: β ≠ 1) states that the data is not a random walk. If we reject H0, it suggests that the data is not a random walk. We use a significance level (α) of 0.05, which corresponds to a 95%

confidence level. To calculate the associated p-value, we use the Z statistic, and the resulting p-value is 6.085467e-05. Because the p-value is less than the alpha value of 0.05, we reject the null hypothesis and conclude that the dataset is predictable and not a random walk.

Step 4: Data Preprocessing:

Initially, the data consisted of weekly records of the number of chickenpox cases in each county of Hungary. These records were then merged to create a weekly total for the entire country. Next, the weekly data was further combined to obtain monthly totals for Hungary. The final dataset used for analysis contained the number of chickenpox cases per month in Hungary from the years 2005 to 2014. In total, there were 120 observations, representing 12 periods per year over a 10-year period. Sometimes it is useful to aggregate the data as it gives more accurate results and helps in easier identification of trends and patterns.

Step 5: Partition Series:

The training and validation data were partitioned using an 80-20 ratio, where 80% of the data was allocated for training (equivalent to 96 periods) and 20% for validation (equivalent to 24 periods). The training data comprised observations from January 2005 to December 2012, while the validation data consisted of data from January 2013 to December 2014.

Step 6: Apply Forecasting

REGRESSION MODELS

Time series regression is a statistical method for predicting a future response based on the response history and the transfer of dynamics from relevant predictors. One of the advantages of the regression models is that they can be used to capture important relationships between the forecast variable of interest and the predictor variables. Three regression-based models are implemented:

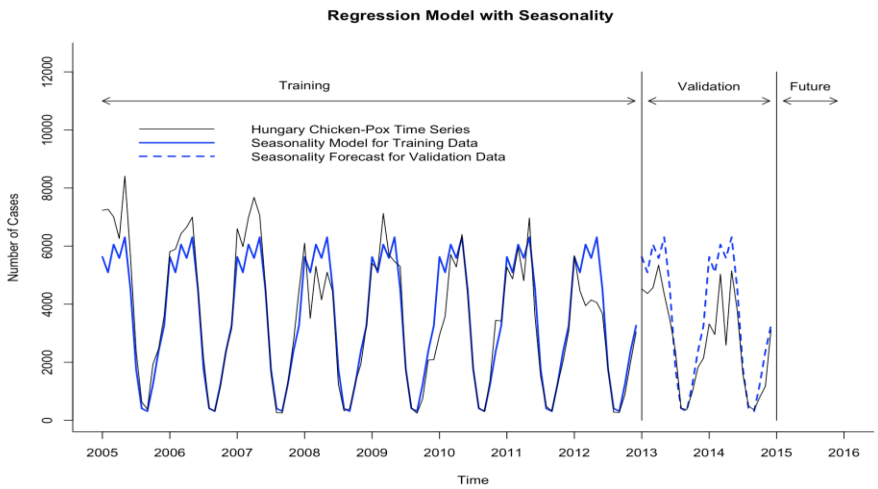
1. Regression based model with Seasonality.

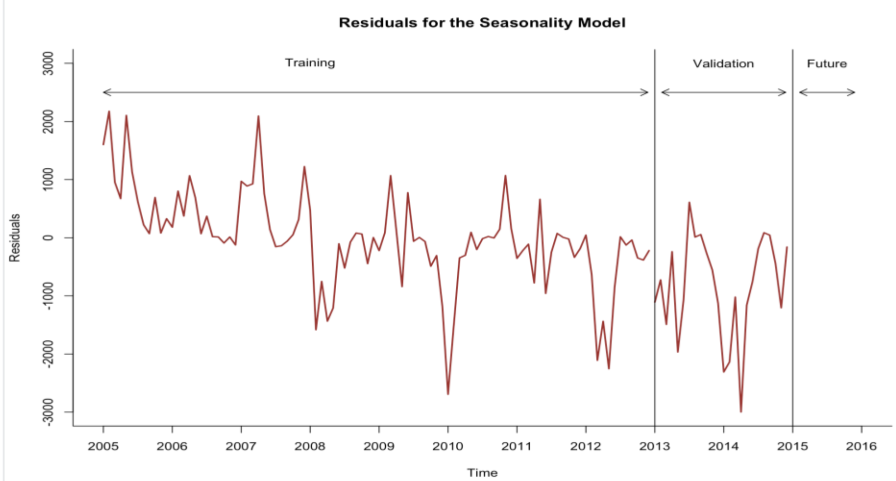
```
Call:
tslm(formula = train.ts ~ season)

Residuals:
    Min      1Q  Median      3Q     Max 
-2693.4  -303.4     6.0   337.3  2175.2 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5629.38     311.40  18.078  < 2e-16 ***
season2      -537.63     440.38  -1.221   0.2256    
season3       426.00     440.38   0.967   0.3362    
season4       -44.38     440.38  -0.101   0.9200    
season5       677.87     440.38   1.539   0.1275    
season6     -1099.38     440.38  -2.496   0.0145 *  
season7     -3859.75     440.38  -8.765 1.77e-13 ***
season8     -5223.12     440.38 -11.860  < 2e-16 ***
season9     -5318.62     440.38 -12.077  < 2e-16 ***
season10    -4393.13     440.38  -9.976 6.50e-16 ***
season11    -3251.63     440.38  -7.384 1.02e-10 ***
season12    -2365.00     440.38  -5.370 6.88e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 880.8 on 84 degrees of freedom
Multiple R-squared:  0.8731,	Adjusted R-squared:  0.8565 
F-statistic: 52.53 on 11 and 84 DF,  p-value: < 2.2e-16
```

The equation of the regression model includes seasonality and is given by,

$$\mathbf{y_t = 5629.38 - 537.63D_2 + 426.00D_3 \ldots\ldots - 2365D_{12}.}$$

The model has an R-squared value of 0.8731, indicating that the seasonal component explains about 87% of the variance in the time series and shows a strong relationship between the dependent and independent variable. The intercept's regression coefficient has a statistically significant p-value, while the p-values for dummy variables D_2, D_3, D_4, and D_5 are not significant. However, dummy variables D_6 to D_{12} are significant. Overall, the model seems to be a good fit in the training partition of the historical data.





Upon examining the regression model with seasonality depicted in the above chart, it becomes clear that while the line fits reasonably well in the training set, it fails to capture the peaks accurately. Moreover, the line does not fit well in the validation set, suggesting overfitting of the model in the validation partition. Furthermore, the residual chart reveals a negative trend in errors, which is a sign of overfitting of the model.

2. <u>Regression Model with Linear Trend and Seasonality</u>

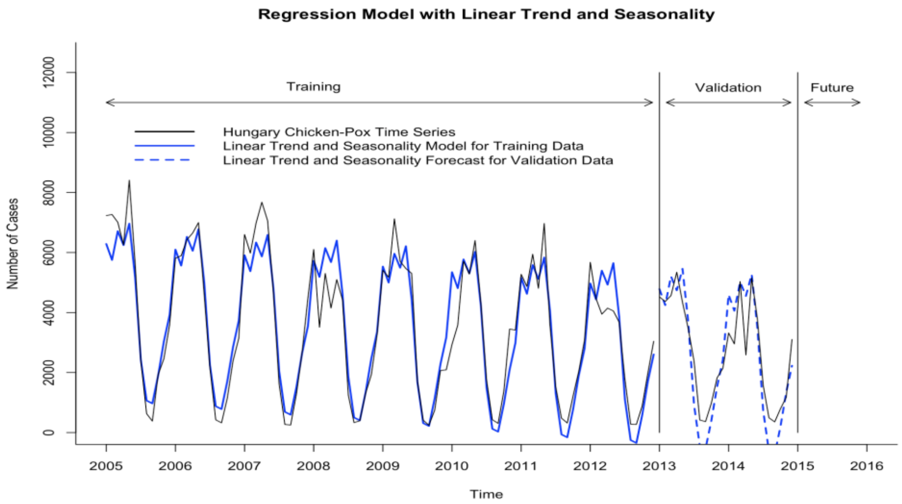
```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
     Min       1Q   Median       3Q      Max 
-2411.14  -406.17    29.91   430.95  1810.76 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6303.603    292.899  21.521  < 2e-16 ***
trend         -15.680      2.803  -5.594 2.77e-07 ***
season2      -521.945    377.542  -1.382  0.17053    
season3       457.359    377.574   1.211  0.22921    
season4         2.664    377.626   0.007  0.99439    
season5       740.594    377.698   1.961  0.05326 .  
season6     -1020.976    377.792  -2.702  0.00834 ** 
season7     -3765.672    377.906  -9.965 7.74e-16 ***
season8     -5113.367    378.041 -13.526  < 2e-16 ***
season9     -5193.187    378.197 -13.731  < 2e-16 ***
season10    -4252.007    378.374 -11.238  < 2e-16 ***
season11    -3094.828    378.571  -8.175 2.91e-12 ***
season12    -2192.523    378.789  -5.788 1.23e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 755.1 on 83 degrees of freedom
Multiple R-squared:  0.9078,	Adjusted R-squared:  0.8945 
F-statistic: 68.13 on 12 and 83 DF,  p-value: < 2.2e-16
```

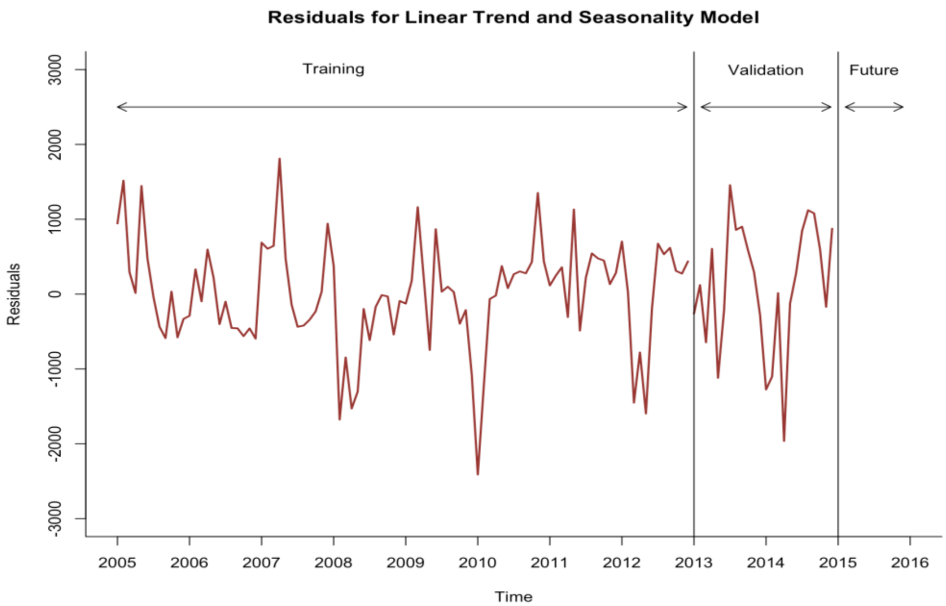
The equation of the regression model includes both linear trend and seasonality and is given by

$$\mathbf{y_t = 6303.603 - 15.680\ t - 521.945D_2 + 457.359D_3 \ldots\ldots -2192.523D_{12}.}$$

The fitted model has an R-squared value of 0.9078, indicating that the trend and seasonal components explain around 91% of the variance in the time series and demonstrate a strong relationship between the dependent and independent variables.

Examining the p-values, all the regression coefficients, including the intercept, trend, and seasons, are statistically significant, except for seasons D_2 to D_6. Overall, the equation suggests that the model appears to be a notably good fit in the training partition of the historical data.





After examining the regression model with linear trend and seasonality displayed in the above chart, it becomes apparent that while the line fits reasonably well in the training set. However, it appears to be a reasonably good fit in the validation partition as well. We cannot say it is an absolute good fit in both the training and validation sets, but rather a reasonably good fit.

Furthermore, upon reviewing the residual chart, it is evident that the trend is straight in errors, which indicates a reasonably good fit of the model. The model can explain most of the variation in the data, and any remaining variation is due to random chance.

3. <u>Regression Model with Quadratic Trend and Seasonality</u>

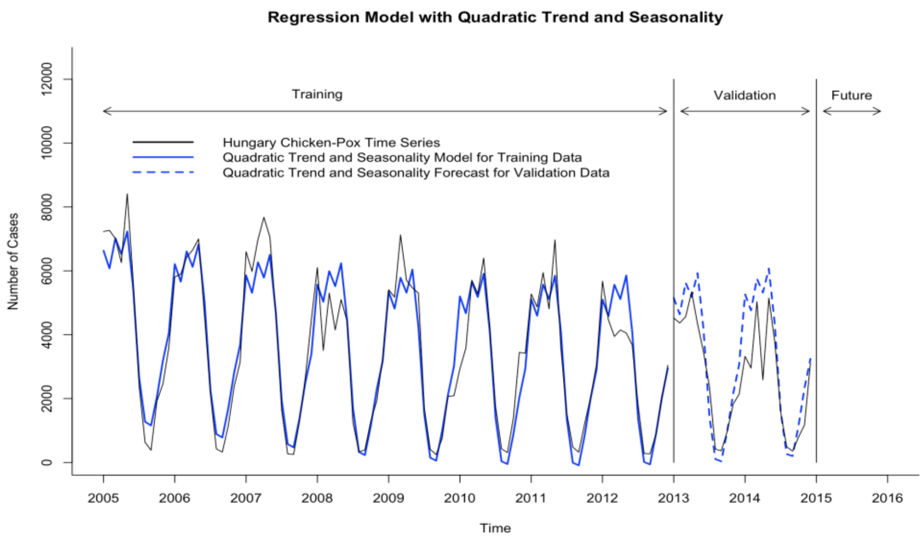
```
Call:
tslm(formula = train.ts ~ trend + I(trend^2) + season)

Residuals:
     Min       1Q   Median       3Q      Max
-2263.32  -405.52    79.02   378.64  1891.01

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6671.0450   334.4886  19.944  < 2e-16 ***
trend         -38.4395    11.0079  -3.492 0.000775 ***
I(trend^2)      0.2346     0.1099   2.135 0.035744 *
season2      -519.5989   369.7024  -1.405 0.163665
season3       461.5829   369.7366   1.248 0.215434
season4         8.2955   369.7917   0.022 0.982157
season5       747.1638   369.8664   2.020 0.046639 *
season6     -1013.9372   369.9599  -2.741 0.007523 **
season7     -3758.6325   370.0719 -10.156 3.67e-16 ***
season8     -5106.7971   370.2022 -13.795  < 2e-16 ***
season9     -5187.5559   370.3514 -14.007  < 2e-16 ***
season10    -4247.7840   370.5202 -11.464  < 2e-16 ***
season11    -3092.4813   370.7096  -8.342 1.46e-12 ***
season12    -2192.5230   370.9212  -5.911 7.49e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 739.4 on 82 degrees of freedom
Multiple R-squared:  0.9127,    Adjusted R-squared:  0.8989
F-statistic: 65.94 on 13 and 82 DF,  p-value: < 2.2e-16
```

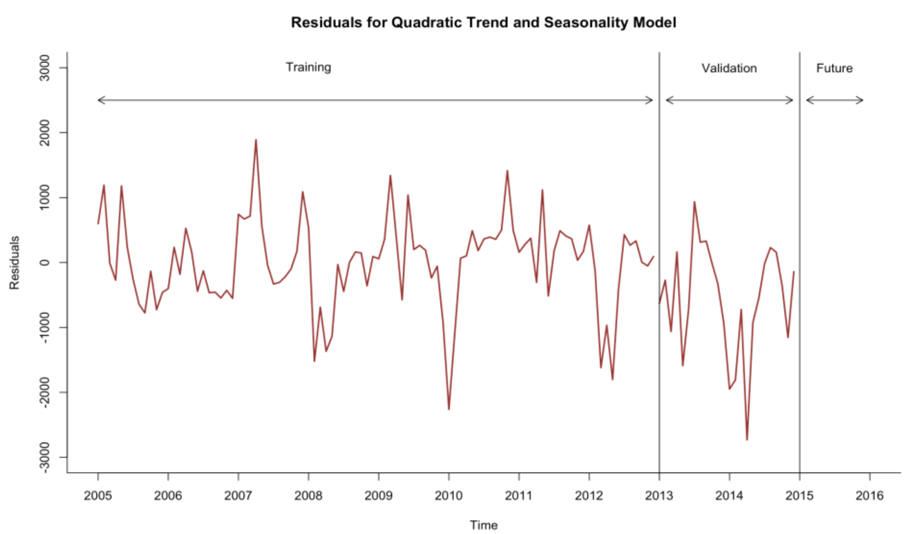
The regression model with quadratic trend and seasonality is represented by the equation:

$$\mathbf{y_t = 6671.0450 - 38.4395\,t + 0.2346\,t^{\wedge}2 - 519.5989D_2 + 461.5829D_3 \ldots\ldots - 2192.5230D_{12}.}$$

The model has an R-squared value of 0.9127, indicating that the trend (both linear and quadratic) and seasonal components explain approximately 91% of the variance in the time series, demonstrating a very strong relationship between the independent and dependent variables.

Examining the p-values, all the regression coefficients, including the intercept, linear and quadratic trend, and seasons, except for seasons D_2, D_3, and D_4, are statistically significant. Overall, the equation suggests that the model appears to be a significantly good fit in the training partition of the historical data.





Upon reviewing the regression model with quadratic trend and seasonality shown in the chart, it is noticeable that the line fits reasonably well in the training set. However, there is a visible discrepancy in the validation set, indicating that the model is overfitting in the validation partition. Furthermore, upon analyzing the residual chart, it becomes evident that there is a slight decreasing trend in errors, which implies overfitting of the model.

AUTO - ARIMA MODEL

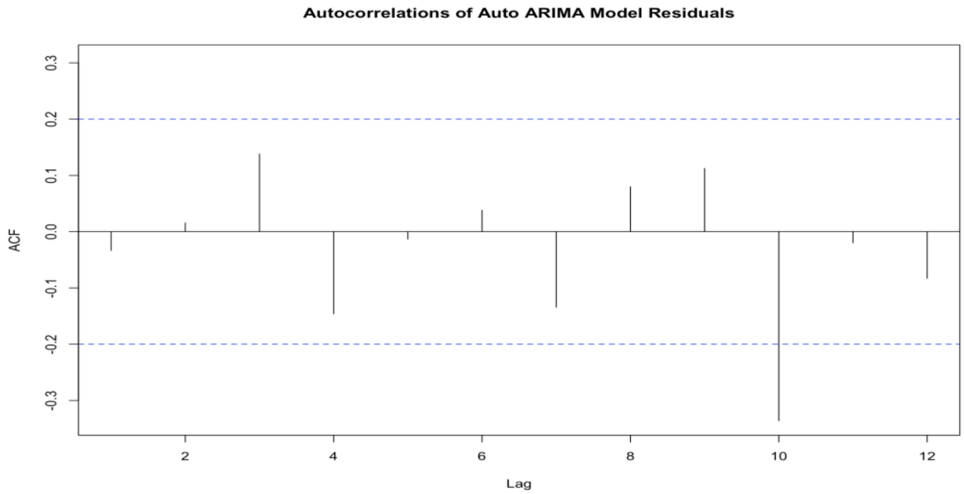
ARIMA (Autoregressive Integrated Moving Average) models are complex with several parameters involved and complex relationships between the model parts. ARIMA (p, d, q) (P, D, Q) [m] model is used to forecast data with level, trend, and seasonality components.

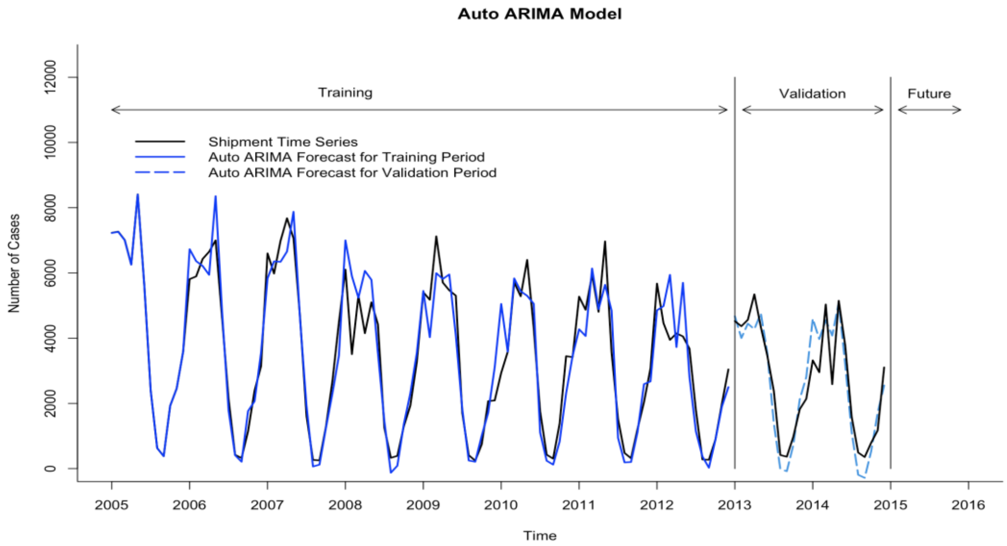
```
Series: train.ts
ARIMA(1,0,0)(2,1,0)[12] with drift

Coefficients:
         ar1      sar1      sar2      drift
      0.5272   -0.4283   -0.3434   -17.3477
s.e.  0.1110    0.1260    0.1193     9.1028

sigma^2 = 621832:  log likelihood = -679.75
AIC=1369.5   AICc=1370.27   BIC=1381.66

Training set error measures:
                      ME     RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set -0.02671112 719.8564 497.0492 5.708914 21.24356 0.6707063 -0.03352976
```

The Equation of the AUTO-ARIMA model for training partition is:

$$\mathbf{y_t = -17.3477 + 0.5272(y_{t-1}) - 0.4283(y_{t-1} - y_{t-2}) - 0.3434(y_{t-2} - y_{t-3})}$$

The Autoregressive model's best seasonal and non-seasonal component parameters have been determined by the computer. The model can be characterized as an autoregressive model of the first order for non-seasonal factors, without differencing being utilized to eliminate a first-order linear trend, and a moving average applied to error lags of the zeroth order. Seasonal components are represented by an autoregressive model of the second order, with differencing employed to remove the first-order linear trend, and a moving average applied to error lags of the zeroth order. The seasonal components pertain to a seasonal period of 12 (i.e., monthly). The autoregressive coefficient for non-seasonal components is 0.5272. For the seasonal components, the autoregressive coefficients are -0.4283 and -0.3434.





The ACF chart for the ARIMA (1,0,0) (2,1,0) model of residuals reveals that all the lags except lag 10 are statistically significant i.e., are random, suggesting that most of the relationships (both linear and seasonal) have been accounted for in the forecasting process. Additionally, the Auto ARIMA model appears to be a good fit for both the training and validation sets.

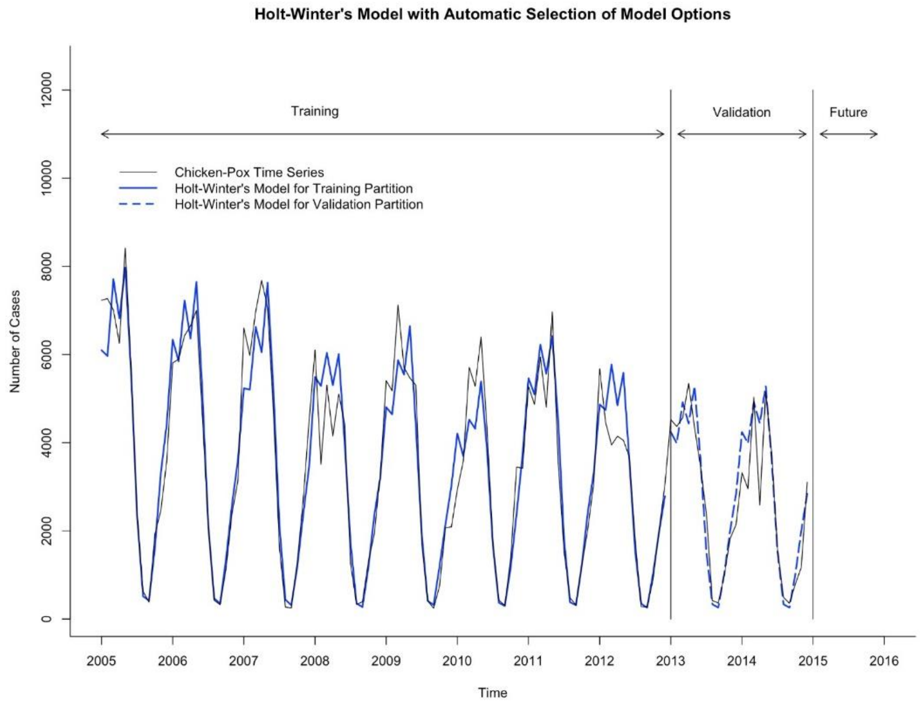
HOLT'S - WINTER MODEL

A Holt-Winters model is defined by its three order parameters, alpha, beta, gamma. Alpha specifies the coefficient for the level smoothing. Beta specifies the coefficient for the trend smoothing. Gamma specifies the coefficient for the seasonal smoothing.

HW ZZZ model with automated selection of error, trend and seasonality and optimal alpha, beta and gamma.

```
ETS(M,N,M)

Call:
 ets(y = train.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.2197
    gamma = 1e-04

  Initial states:
    l = 4129.2821
    s = 0.9895 0.6978 0.3552 0.0901 0.1162 0.5293
           1.2618 1.8386 1.5449 1.7118 1.3881 1.4765

  sigma:  0.1909

     AIC     AICc      BIC
1638.731 1644.731 1677.196
```

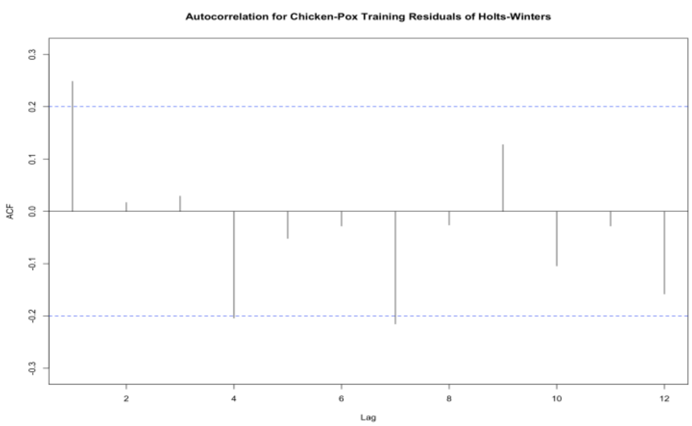
The model has a (M, N, M) structure, indicating Multiplicative error, no trend, and Multiplicative seasonality. The smoothing parameter alpha, set to 0.2197, determines the weighting of recent versus older observations when estimating the level or trend component. A higher alpha value indicates a slower rate of forgetting past observations. The smoothing parameter gamma, set to 1e-04, indicates a high degree of smoothing applied to the seasonal component of the model. The initial state for the level component is 4129.2821, while "s" represents the seasonal component's estimates at the beginning of the time series.The values of AIC (Akaike Information Criterion), AICc (Corrected AIC) and BIC (Bayesian Information Criterion) values provide a means to evaluate the goodness of fit of the model.

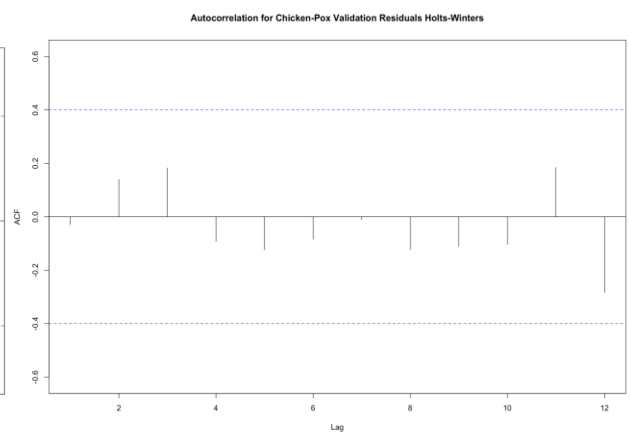


The above chart shows a good fit in both training and validation partition. It covers all the peaks and valleys points in the training and validation part.

TWO – LEVEL MODEL (HW + AR)

The approach involves a dual-level model in which the seasonal and trend components of the time series are modeled using the Holts-Winter method, while the residual variability in the data is captured using the AR(1) model.





We attempted to perform two-level forecasting using Holt-Winters exponential smoothing and an AR (1) model for residuals. However, based on the chart, it seems that the autocorrelation coefficients of the residuals in both the training and validation sets are not statistically significant. There are only slight indications of significance for the autocorrelation coefficients in the training set at lag 1 and lag 7, which we can disregard. This suggests that the residuals are random and there are no discernible relationships between them.

Step 7: Compare And Evaluate Performance:

Accuracy Measures of various models for Validation Partition	MAPE	RMSE
Regression model with seasonality	34.74	1201.25
Regression model with linear trend and seasonality	63.89	851.65
Regression model with quadratic trend and seasonality	36.81	1014.79
Auto ARIMA	41.26	667.46
Holt’s Winter with automated parameters	21.36	632.46

The table indicates that the Regression model with quadratic trend and seasonality, Auto ARIMA, and Holt's winter models with automated parameters exhibit better accuracy measures (MAPE, RMSE) than other models that only include seasonality or linear trend and seasonality. Although the MAPE value for the regression model with seasonality is lower than that of the model with quadratic trend and seasonality, the RMSE values differ significantly. Furthermore, the regression model with quadratic trend and seasonality addresses the trend component, providing an advantage over the other models.

Therefore, to forecast the Hungary dataset for the year 2015 based on the entire dataset, we have applied the following models based on the accuracy measures mentioned above:

1. Regression model with quadratic trend and seasonality
2. Auto ARIMA model
3. Holt’s Winter model with automated parameters

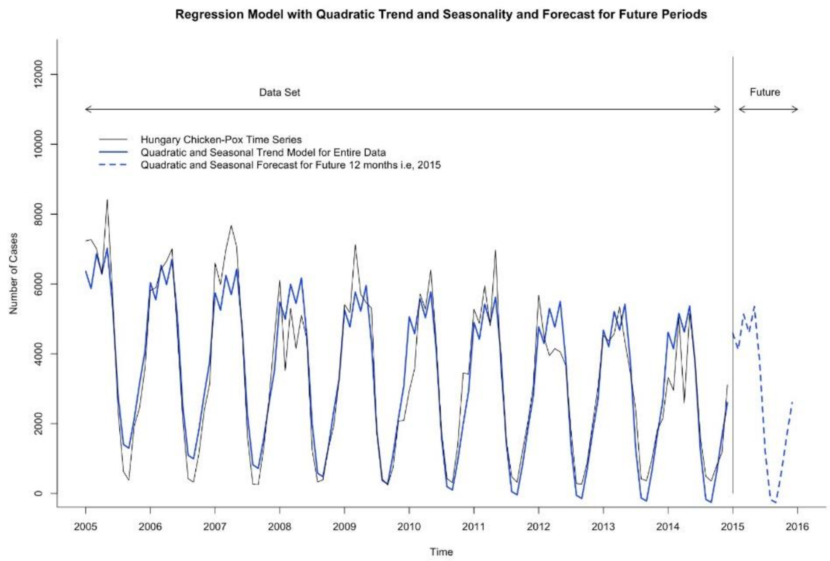
REGRESSION MODEL WITH QUADRATIC TREND AND SEASONALITY ON ENTIRE DATASET

```
> summary(quad.season)

Call:
tslm(formula = H_Chickenpox.ts ~ trend + I(trend^2) + season)

Residuals:
     Min       1Q   Median       3Q      Max 
-2120.73  -463.47    91.61   429.17  1984.69 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.393e+03  3.065e+02  20.856  < 2e-16 ***
trend       -2.898e+01  8.036e+00  -3.606 0.000475 ***
I(trend^2)   1.160e-01  6.432e-02   1.804 0.074065 .  
season2     -4.662e+02  3.381e+02  -1.379 0.170858    
season3      5.484e+02  3.381e+02   1.622 0.107830    
season4      2.101e+01  3.382e+02   0.062 0.950584    
season5      7.700e+02  3.382e+02   2.277 0.024812 *  
season6     -8.622e+02  3.383e+02  -2.549 0.012237 *  
season7     -3.383e+03  3.383e+02  -9.999  < 2e-16 ***
season8     -4.764e+03  3.384e+02 -14.078  < 2e-16 ***
season9     -4.845e+03  3.385e+02 -14.313  < 2e-16 ***
season10    -3.985e+03  3.386e+02 -11.770  < 2e-16 ***
season11    -2.935e+03  3.387e+02  -8.666 5.56e-14 ***
season12    -1.987e+03  3.388e+02  -5.866 5.15e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 756 on 106 degrees of freedom
Multiple R-squared:  0.8977,	Adjusted R-squared:  0.8852 
F-statistic: 71.57 on 13 and 106 DF,  p-value: < 2.2e-16

> |
```

The regression model with quadratic trend and seasonality is represented by the equation:

$\mathbf{y_t = 6.393e{+}03 - 2.898e{+}01\ t + 1.160e{-}01\ t\hat{}2 - 4.662e{+}02D_2 + 5.484e{+}02D_3 \ldots\ 1.987e{+}03D_{12}}$.

The R-squared value of the model is 0.8977, which means that the linear and quadratic trend as well as seasonal components explain approximately 89% of the variations in the time series. This outcome indicates a strong association between the dependent and independent variables. Upon analyzing the p-values, it can be concluded that all regression coefficients, including the intercept, linear trend (with the exception of the quadratic trend), and most seasonal components are statistically significant. However, the seasonal components D2, D3, and D4 are not statistically significant. In summary, the equation indicates that the model is a suitable match for the entire dataset and is notably effective.



The chart unmistakably shows that the blue lines, representing the quadratic and seasonal trends for the complete dataset, offer a considerable degree of precision in modeling the historical data, as they capably capture the peaks and valleys.

AUTO - ARIMA MODEL

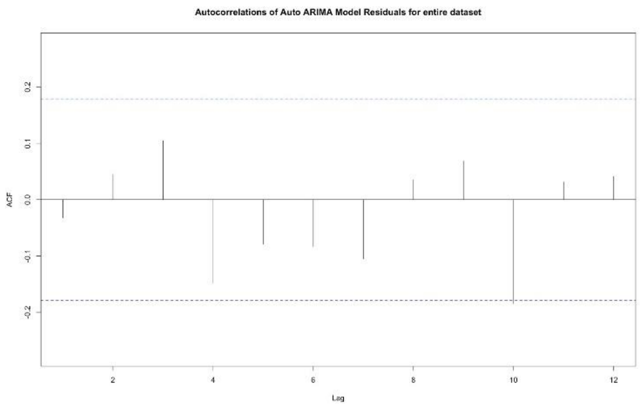
Summary of the Auto ARIMA model used on entire dataset:

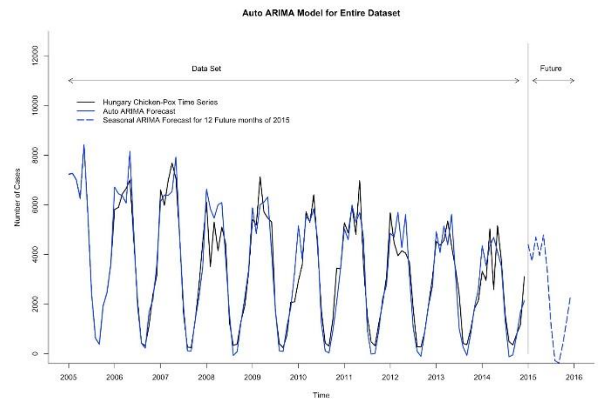
```
Series: hungary_chickenpox.ts
ARIMA(1,0,0)(0,1,1)[12] with drift

Coefficients:
         ar1     sma1      drift
      0.3993  -0.7718  -15.1894
s.e.  0.0937   0.1168     3.6447

sigma^2 = 555664:  log likelihood = -871.52
AIC=1751.04   AICc=1751.42   BIC=1761.76

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE        ACF1
Training set -20.12142 697.2846 491.0725 9.563898 25.51778 0.6848282 -0.03256568
```

The Equation of the AUTO-ARIMA model for training partition is:

$$\mathbf{y_t = -15.1894 + 0.3993(y_{t-1}) - 0.7718(\rho_{t-1})}$$

The computer has determined the optimal seasonal and non-seasonal parameters for the Autoregressive model using the entire dataset. The model can be described as a first-order autoregressive model for non-seasonal factors, without using differencing to remove a first-order linear trend, and a moving average applied to error lags of the zeroth order. For seasonal components, the model employs differencing of order one to remove the linear trend and a moving average applied to error lags of the first order. The seasonal components are based on a seasonal period of 12 (monthly data). The autoregressive coefficient for non-seasonal factors is 0.3993.





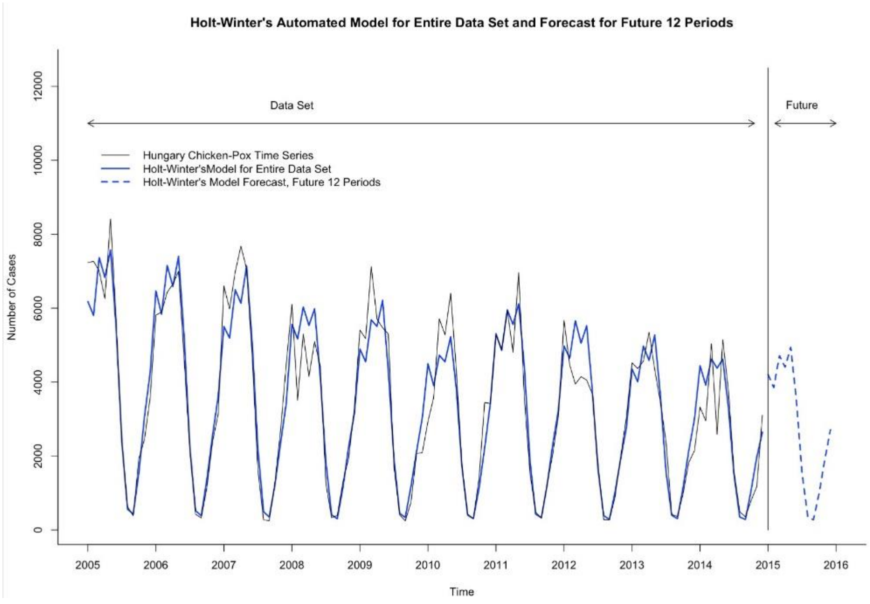
The autocorrelation function (ACF) plot for the residuals of the ARIMA (1,0,0) (0,1,1) model indicates that all the lags, except for lag 10 (which is insignificantly different), are statistically significant and random. This implies that most of the linear and seasonal relationships have been effectively captured in the forecasting process. Furthermore, the Auto ARIMA model seems to be an appropriate fit for the entire dataset.

HOLT'S - WINTER MODEL WITH AUTOMATED PARAMETERS

Automated selection of HW model

```
ETS(M,N,M)

Call:
 ets(y = hungary_chickenpox.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.1461
    gamma = 1e-04

  Initial states:
    l = 4087.0123
    s = 0.9751 0.6824 0.3564 0.0979 0.1285 0.5471
           1.2565 1.7782 1.5843 1.6963 1.3851 1.5121

  sigma:  0.2045

     AIC     AICc      BIC
2081.150 2085.765 2122.962
```

The model appears to be (M, N, M) which means Multiplicative error, no trend and Multiplicative seasonality. The smoothing parameter alpha which is **0.1461** represents the smoothing of the level component and means that the model puts more weight on the most recent observations when estimating the level or trend component, compared to the older observations. The value of alpha can be interpreted as the rate at which the model forgets the past observations, with higher values leading to a slower rate of forgetting and smoothing parameter gamma which is **1e-04** indicates the smoothing of the seasonal component i.e., high degree of smoothing has been applied to the seasonal component of the model. The initial states l of 4087.0123 represents the initial level component of time series and s represents the estimates of seasonal components at the start of the time series. The values of AIC (Akaike Information Criterion), AICc (Corrected AIC) and BIC (Bayesian Information Criterion) helps in evaluating how good the model fits.



The chart displayed above exhibits a good fit to the entire dataset, as it captures all the high and low points in the time series.

Accuracy Measures of various models for Entire Dataset	**MAPE**	**RMSE**
Regression model with quadratic trend and seasonality	35.1	710.571
Auto ARIMA Model	25.52	697.29
Holt’s Winter Model with automated parameters	17.24	658.78
Naïve Model	69.64	1481.26
Seasonal Naïve Model	25.52	1016.40

According to the comparison table, among all the models, Holt's Winter model offers the most precise predictions for the Hungary dataset when considering the MAPE and RMSE. As a result, to make a forecast for the Hungary dataset in the year 2015, it is advisable to use Holt's Winter model.

CONCLUSION

The Holt's Winter model has been identified as the optimal choice for generating accurate predictions. This provides evidence that incorporating recent data in forecasting leads to better results than relying on historical data alone. As previously noted, it is advisable to periodically reassess the chosen model every six months to ensure precise forecasts for the upcoming fiscal period.