Advanced Statistical Modeling
Time Series

Eloi Campeny Roig
Maria Paraskeva
Omar López Rubio

UPC - FIB
January 2024

Contents

1 Introduction

In this assignment of the Time Series section from the subject Advanced Statistical Modeling we are going to explore the Box-Jenkins ARIMA methodology and its extensions as applied to a selected ”real-time series”.

The time series we will base our analysis on was chosen from the proposed time series given during the course. It shows the apparent cement consumption in Spain (Construction Indicator) in thousands of tonnes, and the plot of the series can be seen in 1. The time series plot shows an evident drop around 2008 which can be attributed to the 2008–2014 Spanish financial crisis, also known as the Great Recession in Spain. It makes sense that during a financial crisis, less construction sites are operating, resulting in a lower cement consumption across the country.

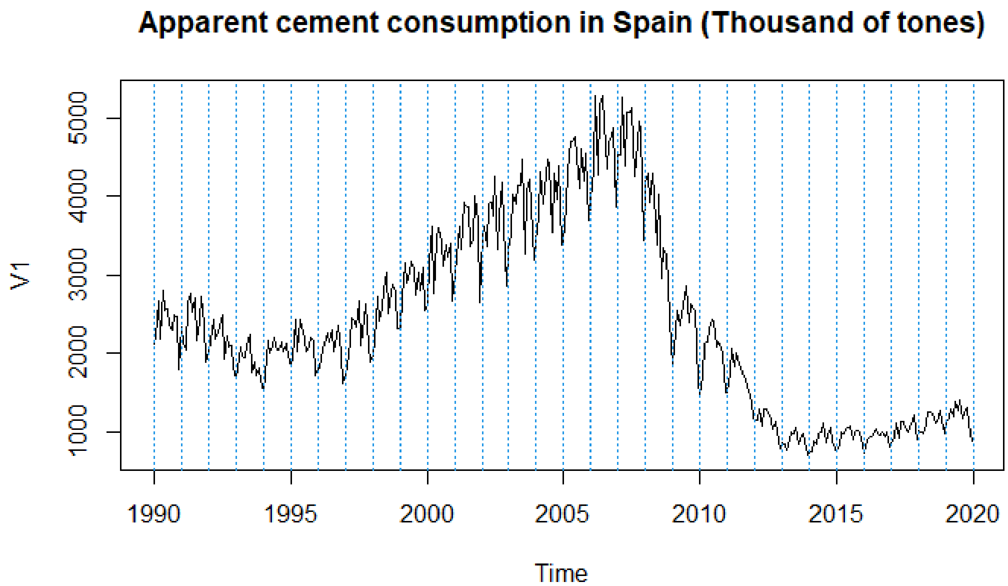

Figure 1: Original time series.

2 Identification

The first step in a time series analysis is to transform the time series to make it stationary and analyse the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify plausible models for the data. The first goal is to make the time series stationary. Stationarity is a key assumption in many time series models, as it simplifies the modeling process and ensures that statistical properties of the series do not change over time. By calculating the mean and variance per month and plotting the annual mean against the annual variance, we could see that the points mostly align with the regression line, suggesting a linear relationship as can be seen in the plot of Figure 2. Then we plot the cement consumption over the years so as to gain insights into potential trends or seasonal patterns in construction activity. In Figure 3 we can clearly see that the height of the boxes (IQR) is higher for higher values of the mean, so we proceed in a change of scale.

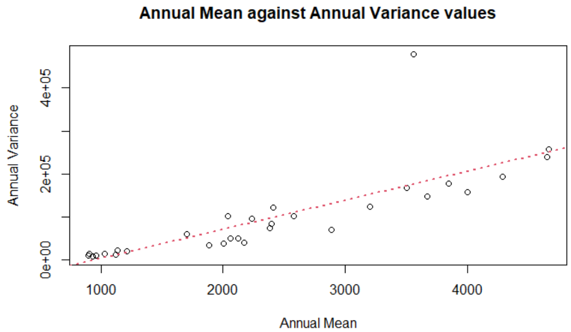

Figure 2: Linear relationship between annual mean and annual variance.

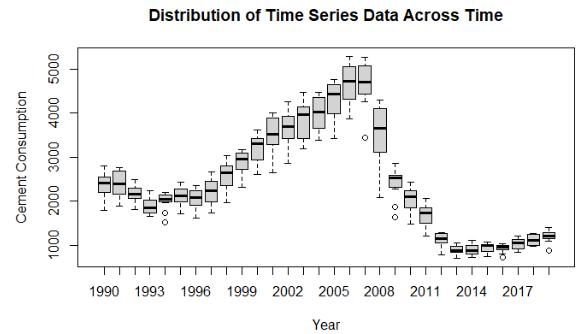

Figure 3: Boxplots showing the data distribution.

2.1 Transformations

The first transformation we make is a logarithmic one, as it is a common technique to stabilize variance and make the data more amenable to modeling. We first plot the logarithmic transformation of the data shown in Figure 4 and then we create the same plots as before, showing the annual mean and variance relationship in the transformed logarithmic values (Figure 5), as well as the distribution of the logarithmic data shown in the form of boxplots (Figure 6).

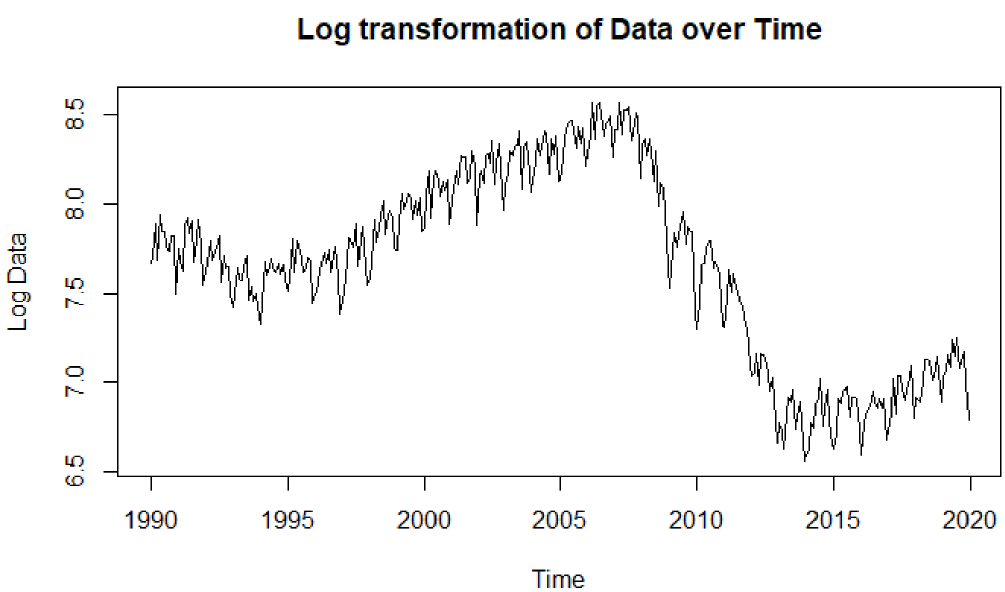

Figure 4: Logarithmic transformation.

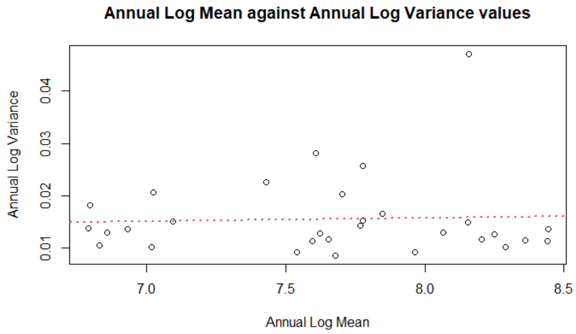

Figure 5: Relationship between annual log mean and annual log variance.

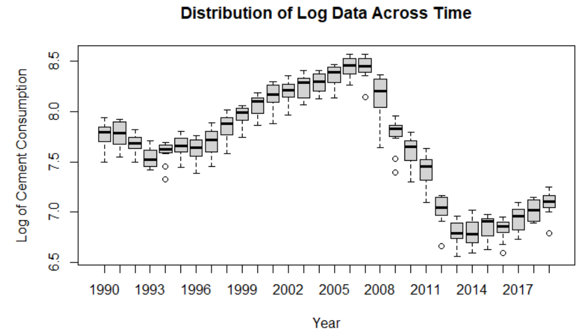

Figure 6: Boxplots showing the log data distribution.

In order to better understand and model the patterns within the data, we proceed to the decomposition of the time series by breaking down the series into its underlying components. Figure 7 shows the values being broken down into trend, seasonal, and residual (random) components. We can see the downfall of the financial crisis in the trend section, and there is also an evident seasonal pattern.

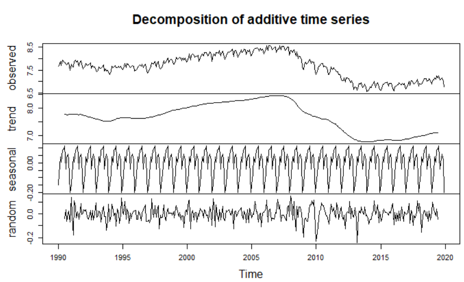

Figure 7: Time series decomposition.

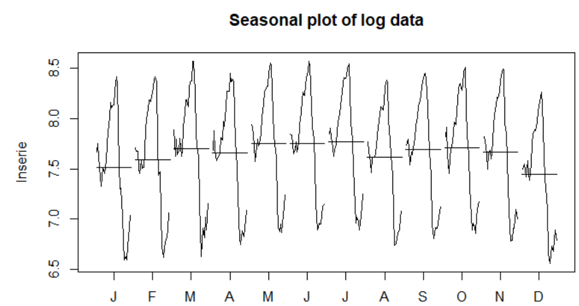

Figure 8: Seasonal plot of logarithmic transformation.

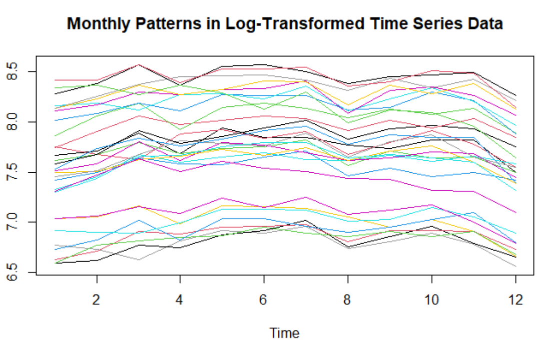

Figure 9: Monthly patterns.

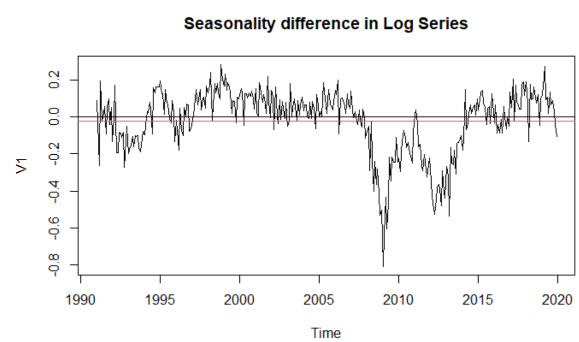

Figure 10: Seasonal difference in Log series.

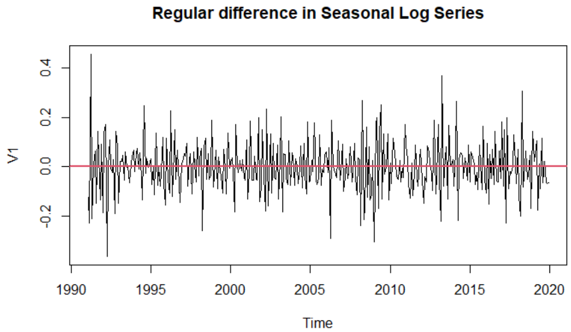

Figure 11: Regular difference.

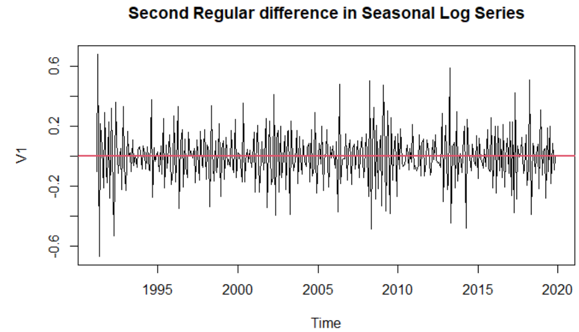

Figure 12: Second Regular difference.

In Figure 8 we can see that the variance is not completely stable, which shows us that we should do more transformations. Figure 9 shows the monthly patterns in the log transformed data and the seasonal patterns can be seen clearly. The next three plots show first the seasonality difference which has an unstable mean (10), then the regular difference (11) and the second regular difference (12). The last two plots show how the values at each time point change compared to the previous time point after taking one and then two differences. Fluctuations above and below the horizontal line at $y = 0$ indicate the direction and magnitude of changes in the second regular difference. We gather the variances of the three series (*d12lnserie*, *d1d12lnserie*, *d1d1d12lnserie*) in Table 1 in order to compare the variances and select the lower, which is that of the *d1d12lnserie*.

Series	Variance
d12lnserie	0.2866344
d1d12lnserie	0.01199032
d1d1d12lnserie	0.03492908

Table 1: Variance of the transformed series.

2.2 ACF and PACF Analysis

After achieving stationarity, the ACF and PACF are examined to identify potential models for the time series. The ACF function shows the correlation between a series and its lagged values. Peaks in the ACF indicate potential autocorrelation at those lags, while the PACF function displays the correlation between a series and its lagged values while removing the effects of intermediate lags. Peaks in the PACF suggest direct relationships without the influence of other lags. The autocorrelation and partial autocorrelation plots can be found side by side in Figure 13.

We identify the ACF as infinite decreasing lags, and PACF as the one with finite lags different from 0. For the regular part we see that the last lag different from zero could be lag 2 or lag 4 and for seasonal we think that the best is 4, (6 is too much and produces errors). So, the identified potential models are AR(2), AR(4), and SAR(4).

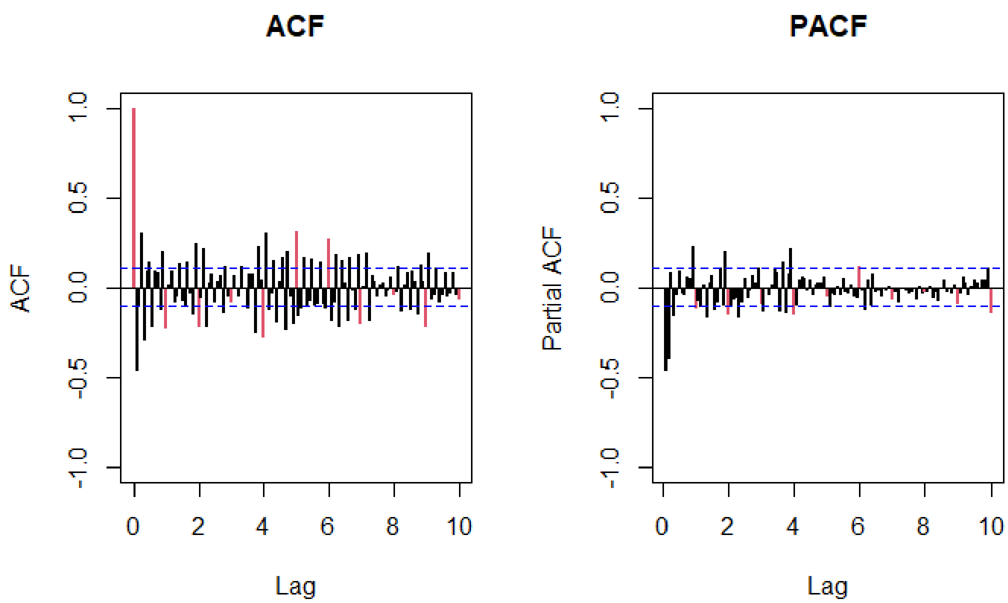

Figure 13: ACF and PACF plots of *d1d12lnserie*.

3 Estimation

We proceed with estimating the models using R. For the model AR(2) SAR(4) with zero regular and seasonal differences (d=0, D=0) in Figure 14, we test to see if the intercept is significant. Looking at the t-ratio we see that it is lower than 2 so the intercept is not significant. So we can use the model with the regular and seasonal differences (d=1, D=1) shown in Figure 15. The AIC is lower for the model without intercept and we see that all the weights are significant (all abs(t) values > 2).

```
Call:
arima(x = d1d12lnserie, order = c(2, 0, 0), seasonal = list(order = c(4, 0,
    0), period = 12))

Coefficients:
          ar1      ar2     sar1     sar2     sar3     sar4  intercept
      -0.4291  -0.2002  -0.5424  -0.6249  -0.4701  -0.5386      2e-04
s.e.   0.0590   0.0588   0.0473   0.0535   0.0509   0.0553      8e-04

sigma^2 estimated as 0.004824:  log likelihood = 421.88,  aic = -827.76
       ar1        ar2       sar1       sar2       sar3       sar4  intercept
 7.2721005  3.4055123 11.4712236 11.6786796  9.2403510  9.7323341  0.1980013
intercept
0.1980013
```

Figure 14: Model AR(2) SAR(4) with (d=0, D=0) metrics.

```
Call:
arima(x = lnserie, order = c(2, 1, 0), seasonal = list(order = c(4, 1, 0), period = 12))

Coefficients:
          ar1      ar2     sar1     sar2     sar3     sar4
       -0.429  -0.2000  -0.5423  -0.6247  -0.4700  -0.5387
s.e.    0.059   0.0587   0.0473   0.0535   0.0508   0.0553

sigma^2 estimated as 0.004825:  log likelihood = 421.86,  aic = -829.72
      ar1       ar2      sar1      sar2      sar3      sar4
 7.268824  3.407810 11.466640 11.674799  9.243763  9.732309
```

Figure 15: Model AR(2) SAR(4) with (d=1, D=1) metrics.

We follow the same practices for the model AR(4) SAR(4) with zero regular and seasonal differences (d=0, D=0) (Figure 16) to test if the intercept is significant. Looking at the t-ratio we see that it is lower than 2 so the intercept is not significant. So we can use the model with the regular and seasonal differences (d=1, D=1) (Figure 17). Here the AIC is lower (-832.89) for the model with regular and seasonal difference, but the coefficients AR3 and AR4 are not significant enough. Comparing both models, model 2 has lower AIC but the last two coefficients of AR do not pass the significance test. So we finally choose the first model AR(2) SAR(4) as it has almost the same AIC of model 2 but it is simpler and all coefficients are significant.

```
Call:
arima(x = d1d12lnserie, order = c(4, 0, 0), seasonal = list(order = c(4, 0,
    0), period = 12))

Coefficients:
          ar1      ar2     ar3      ar4     sar1     sar2     sar3     sar4  intercept
      -0.4047  -0.1692  0.0973  -0.0747  -0.5494  -0.6152  -0.4560  -0.5156      2e-04
s.e.   0.0568   0.0606  0.0588   0.0552   0.0484   0.0542   0.0529   0.0552      8e-04

sigma^2 estimated as 0.004749:  log likelihood = 425.47,  aic = -830.93
       ar1        ar2        ar3        ar4       sar1       sar2       sar3       sar4
 7.1223326  2.7895614  1.6550313  1.3524139 11.3468835 11.3596164  8.6162180  9.3439073
 intercept
 0.1933412
intercept
0.1933412
```

Figure 16: Model AR(4) SAR(4) with (d=0, D=0) metrics.

```
Call:
arima(x = lnserie, order = c(4, 1, 0), seasonal = list(order = c(4, 1, 0), period = 12))

Coefficients:
          ar1      ar2     ar3      ar4     sar1     sar2     sar3     sar4
      -0.4045  -0.1688  0.0976  -0.0745  -0.5493  -0.6150  -0.4560  -0.5157
s.e.   0.0569   0.0607  0.0592   0.0550   0.0484   0.0541   0.0527   0.0539

sigma^2 estimated as 0.004749:  log likelihood = 425.45,  aic = -832.89
      ar1       ar2       ar3       ar4      sar1      sar2      sar3      sar4
 7.106981  2.783658  1.648259  1.353665 11.358992 11.357675  8.647515  9.574784
```

Figure 17: Model AR(4) SAR(4) with (d=1, D=1) metrics.

4 Validation

We now proceed to the validation of our chosen model.

Residual analysis, Figure 18: The residuals appear to be random and centered around zero and there are some visible outliers. We could say that their variance is constant looking at the absolute residuals plot. There seems to be no clear pattern in the variability of the residuals over time, which suggests that the variance of the errors is constant (homoscedastic). The Q-Q plot shows some deviation in the tails, indicating that the residuals may not be perfectly normally distributed. The histogram shows that the residuals are roughly bell-shaped and centered around zero, even though there are more observations in the mean as well as the right side of the mean (heavy tail), indicating some slight skewness.

Looking at the statistical tests, we get a different interpretation as the normality tests (Shapiro-Wilk, Anderson-Darling, Jarque Bera) indicate that the residuals are not normally distributed (p-values are significant). Also, regarding homogeneity of variance, the Breusch Pagan test gives us ap-value of 0.531 suggests that there is no evidence of heteroscedasticity in the residuals.

Autocorrelation, Figure 19: The Ljung-Box test checks for autocorrelation in residuals at different lags and some lags (like 12 and 48) show significant p-values, suggesting possible autocorrelation issues at these lags. However, the Durbin-Watson test suggests no autocorrelation in residuals (DW close to 2).

The third plot that gives us another representation of the p-values for the Ljung-Box statistic over different lags indicates that some dots seem to be above the significance threshold (dashed lines), while most are below it. As the dots represent the p-value at each lag, we see that there is autocorrelation at the lags represented by the dots who are below the threshold. The Ljung-Box test fails after lag 9, so the samples are not independent.

Invertibility: There is no number in the results so we cannot make an interpretation. As our model is only comprised of an AR component and has no MA component, the *theta* variable is empty. Hence, the concept of invertibility does not apply to our model.

Causality, Figure 20: The model is stationary/causal as all the roots of the seasonal AR-characteristic polynomial lie outside of the unit circle with all modul >1. So, the model can be represented as a convergent MA(∞) expression with psi weights.

To sum up, outlier detection and its treatment should be the next step, and possibly, better validation results can be obtained. Also, calendar effects analysis should be performed.

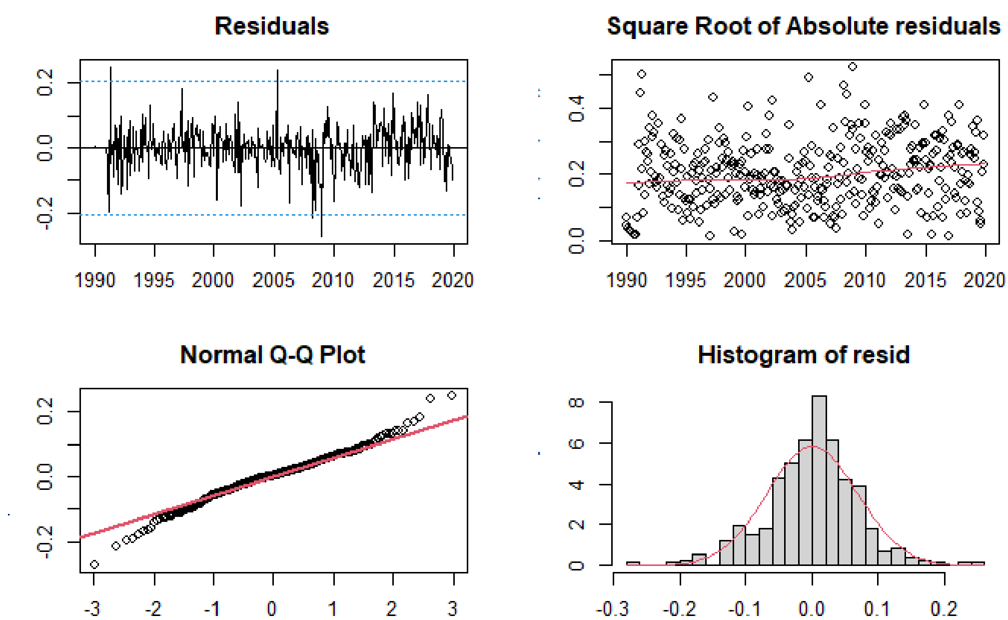

Figure 18: Residuals, baseline model.

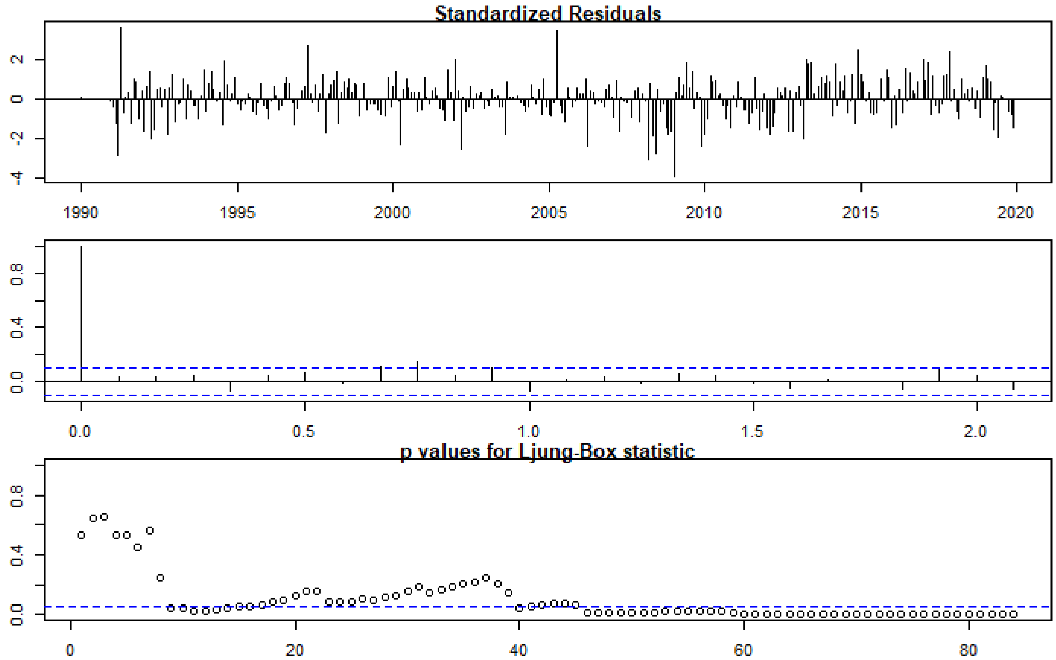

Figure 19: Residuals and Ljung-Box, baseline model.

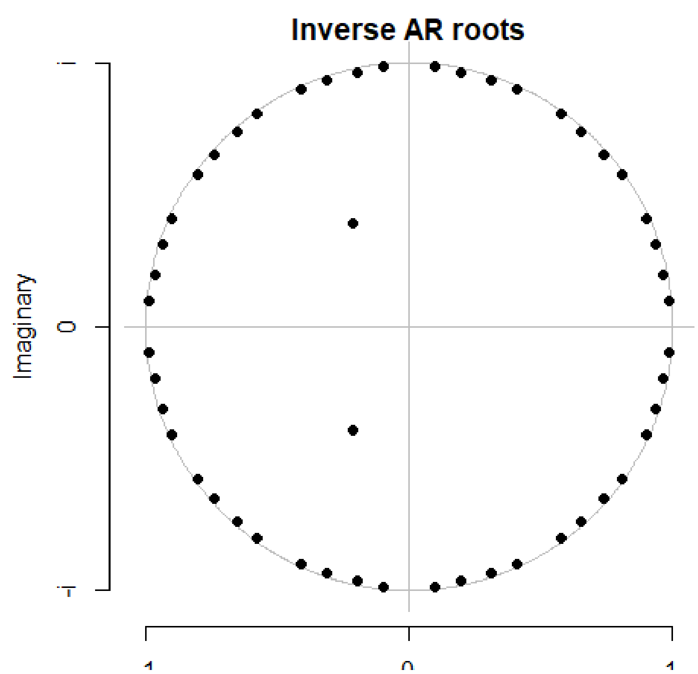

Figure 20: Inverse roots, baseline model.

5 Predictions

Using the model best model we proceed to make predictions.

Initially, we verify the stability of the model by comparing the coefficients between the model trained with the entire series and the model excluding the data from the last year (Figure 21).

```
Call:
arima(x = lnserie1, order = pdq, seasonal = list(order = PDQ, period = 12))

Coefficients:
         ar1      ar2     sar1     sar2     sar3     sar4
      -0.429  -0.2000  -0.5423  -0.6247  -0.4700  -0.5387
s.e.   0.059   0.0587   0.0473   0.0535   0.0508   0.0553

sigma^2 estimated as 0.004825:  log likelihood = 421.86,  aic = -829.72

Call:
arima(x = lnserie2, order = pdq, seasonal = list(order = PDQ, period = 12))

Coefficients:
          ar1      ar2     sar1     sar2     sar3     sar4
      -0.4457  -0.2208  -0.5248  -0.6374  -0.4547  -0.5314
s.e.   0.0602   0.0599   0.0486   0.0548   0.0519   0.0567

sigma^2 estimated as 0.004799:  log likelihood = 407.98,  aic = -801.97
```

Figure 21: Model with all the time series and without last year.

Second, we check the predictions of year 2019 (Figure 22).

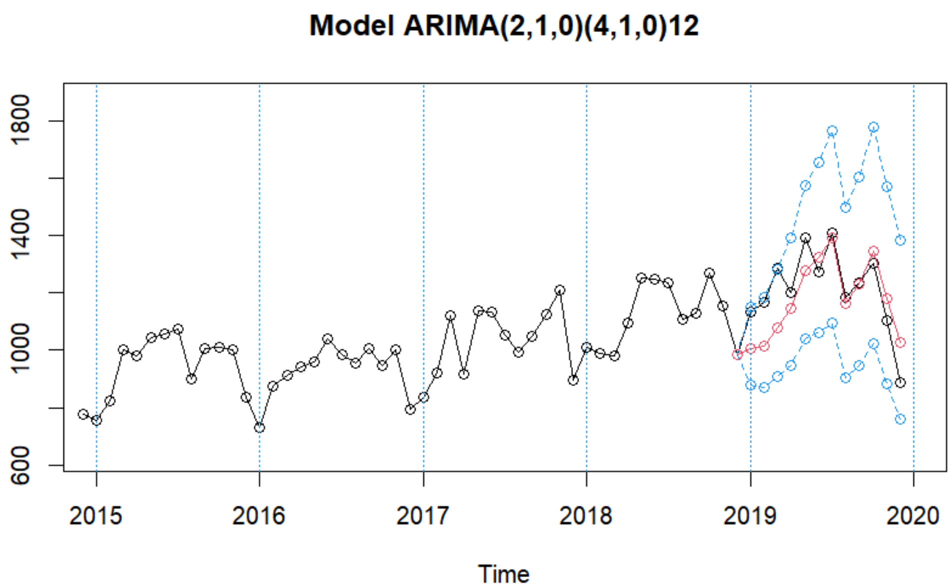

Figure 22: Predictions 2019.

In Figure 44, it is evident that the model tends to underestimate concrete consumption in the first month. This underestimation could be attributed to the calendar effects or outliers .

Then, we predict the year 2020 (Figure 23).

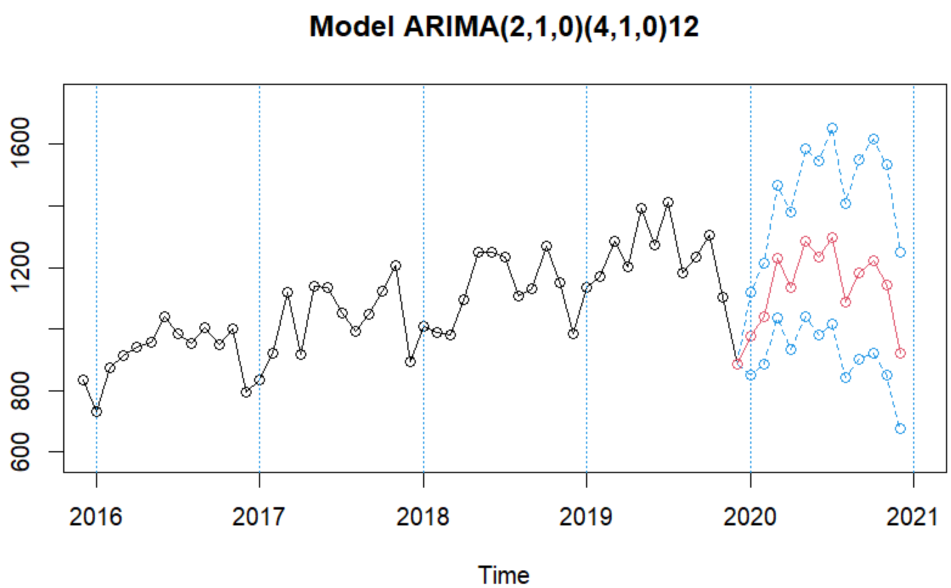

Figure 23: Model with all the time series and without last year.

6 Calendar Effects

In our investigation, we conducted training on three newly developed models to assess the significance of the Trading Day and Easter Week effects. Employing the optimal parameters pdq and PDQ obtained from our best-performing model to date, we conducted training using transformed time series data, incorporating additional information related to trading days and Easter week. The first model, as illustrated in Figure 24, utilized transformed series data along with information on trading days. Subsequently, the second model (Figure 25) incorporated transformed series data along with Easter week information. Lastly, the third model (Figure 26) integrated transformed series data with details on both trading days and Easter week.

```
Call:
arima(x = lnserie, order = pdq, seasonal = list(order = PDQ, period = 12), xreg = wTradDays)

Coefficients:
          ar1      ar2     sar1     sar2     sar3     sar4  wTradDays
      -0.3032  -0.1504  -0.6739  -0.5444  -0.3082  -0.3140     0.0119
s.e.   0.0559   0.0547   0.0544   0.0652   0.0631   0.0572     0.0009

sigma^2 estimated as 0.00397:  log likelihood = 461.24,  aic = -906.48
      ar1       ar2      sar1      sar2      sar3      sar4 wTradDays
 5.427796  2.751723 12.395774  8.352573  4.886567  5.491408 12.957543
```

Figure 24: Model for transformed time series and Trading days.

```
Call:
arima(x = lnserie, order = pdq, seasonal = list(order = PDQ, period = 12), xreg = wEast)

Coefficients:
          ar1      ar2     sar1     sar2     sar3     sar4    wEast
      -0.3177  -0.1233  -0.4294  -0.5156  -0.5516  -0.4308  -0.1139
s.e.   0.0581   0.0605   0.0496   0.0497   0.0474   0.0552   0.0095

sigma^2 estimated as 0.003723:  log likelihood = 467.66,  aic = -919.33
      ar1       ar2      sar1      sar2      sar3      sar4     wEast
 5.471862  2.038638  8.665087 10.372871 11.626368  7.807974 11.970752
```

Figure 25: Model for transformed time series and Easter week.

```
Call:
arima(x = lnserie, order = pdq, seasonal = list(order = PDQ, period = 12), xreg = data.frame(wTradDays,
    wEast))

Coefficients:
          ar1      ar2    sar1     sar2     sar3     sar4  wTradDays    wEast
      -0.1790  -0.0991  -0.525  -0.3973  -0.3589  -0.2023     0.0113  -0.1039
s.e.   0.0544   0.0579   0.064   0.0606   0.0779   0.0770     0.0008   0.0082

sigma^2 estimated as 0.002961:  log likelihood = 514.12,  aic = -1010.23
      ar1       ar2      sar1      sar2      sar3      sar4 wTradDays     wEast
 3.290223  1.712094  8.200912  6.551577  4.606271  2.626303 14.435149 12.687684
```

Figure 26: Model for transformed time series, Trading days and Easter week.

Models	Trading Days (T-Ratio)	Easter Week (T-Ratio)	AIC
Series	-	-	-829.717
Series + TD	0.0119 (13.0)	-	-906.48
Series + EW	-	-0.1139 (12.0)	-919.33
Series + TD + EW	0.0113 (14.4)	-0.1039 (12.7)	-1010.23

Table 2: Comparison Models with calendar effects.

From Table 2, it is evident that the calendar effect holds significance for all models, indicated by T-Ratios exceeding 2. Furthermore, the models exhibit a substantial decrease in AIC. Consequently, we acknowledge the importance of considering calendar effects in the formulation of the new models.

Utilizing the model that incorporates both Trading Days and Easter week, we will proceed to transform the time series. This transformation entails subtracting the calendar effects from the series, with the goal of identifying new models designed to explicitly accommodate calendar effects.

Utilizing the transformed series, denoted as $d1d12lnserieEC$, where $d1d12lnserie$ represents the series with the effects of the calendar subtracted, we generate autocorrelation function (ACF) and partial autocorrelation function (PACF) plots (Figure 27). These plots serve the purpose of identifying potential new models.

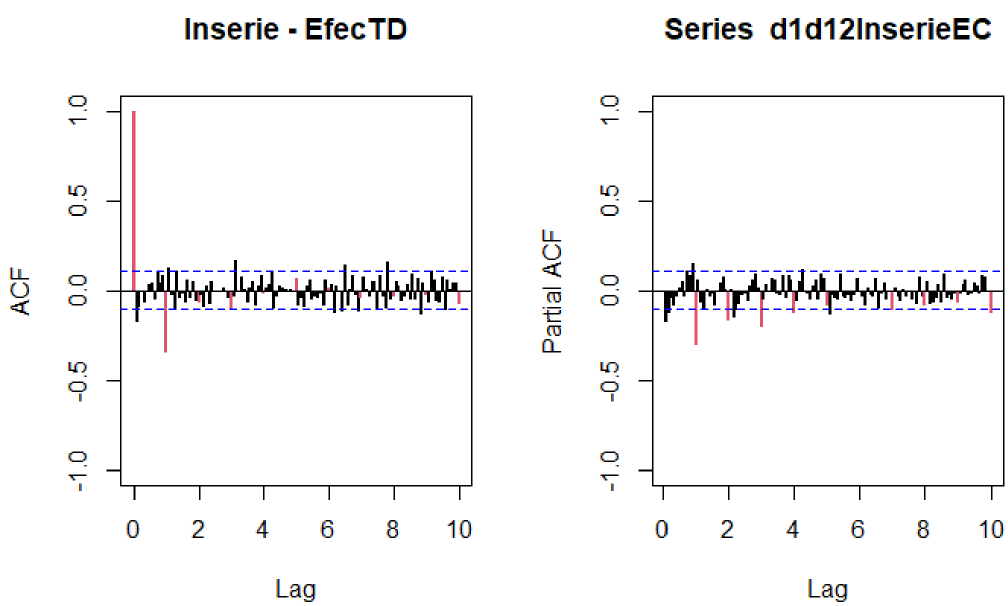

Figure 27: ACF and PACF for $d1d12lnserieEC$.

We identify AR(1), MA(2) for the regular part of the model and SAR(1) for the seasonal part.

```
Call:
arima(x = lnserie, order = c(1, 1, 0), seasonal = list(order = c(1, 1, 0), period = 12),
    xreg = data.frame(wTradDays, wEast))

Coefficients:
          ar1     sar1  wTradDays    wEast
      -0.1396  -0.3465     0.0117  -0.1055
s.e.   0.0539   0.0523     0.0006   0.0078

sigma^2 estimated as 0.003537:  log likelihood = 486.17,  aic = -962.34
      ar1       sar1 wTradDays      wEast
 2.590672  6.625657 19.925310 13.547765
```

Figure 28: Model for calendar effects AR(1) SAR(1).

```
Call:
arima(x = lnserie, order = c(0, 1, 2), seasonal = list(order = c(1, 1, 0), period = 12),
    xreg = data.frame(wTradDays, wEast))

Coefficients:
          ma1      ma2     sar1  wTradDays    wEast
      -0.1616  -0.1140  -0.3452     0.0116  -0.1032
s.e.   0.0546   0.0557   0.0526     0.0006   0.0077

sigma^2 estimated as 0.003476:  log likelihood = 489.19,  aic = -966.39
      ma1       ma2      sar1 wTradDays      wEast
 2.960200  2.048012  6.569362 19.478039 13.329819
```

Figure 29: Model for calendar effects MA(2) SAR(1).

We take as the best model MA(2), SAR(1) since have the lowest AIC (-962.34 vs -966.39).

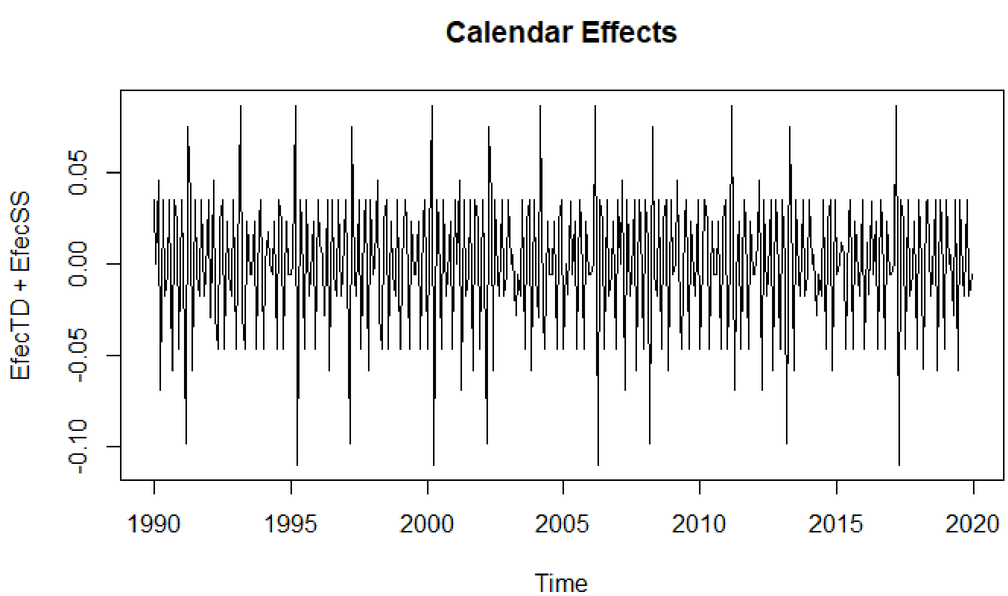

Figure 30: Calendar effect.

In Figure 30, the calendar effects estimated by our model are displayed. Figures 31 and 32 show the series with the calendar effects removed, representing the entire dataset and the last 5 years, respectively.

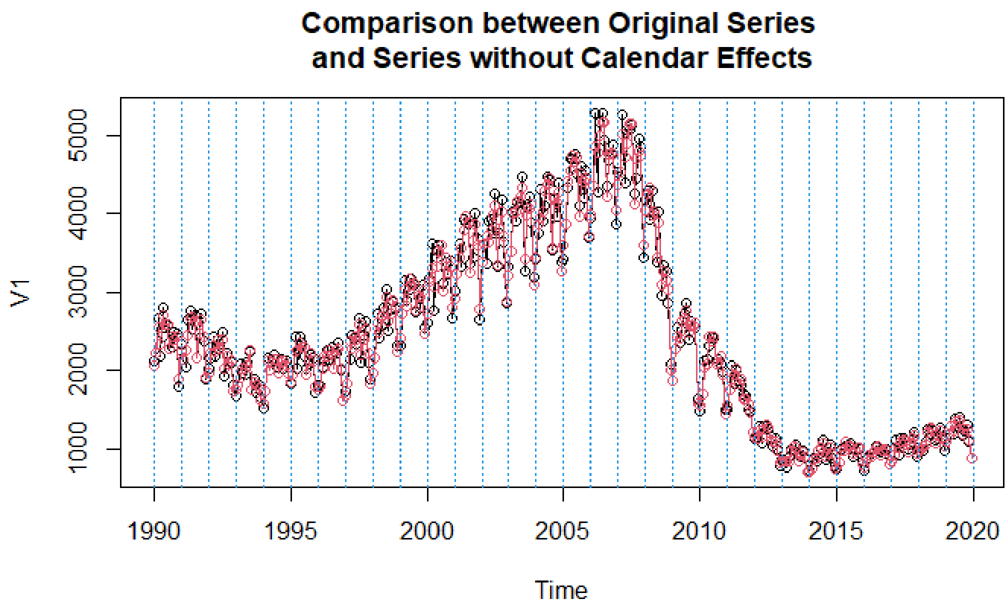

Figure 31: Comparison of Original series VS Without calendar effect.

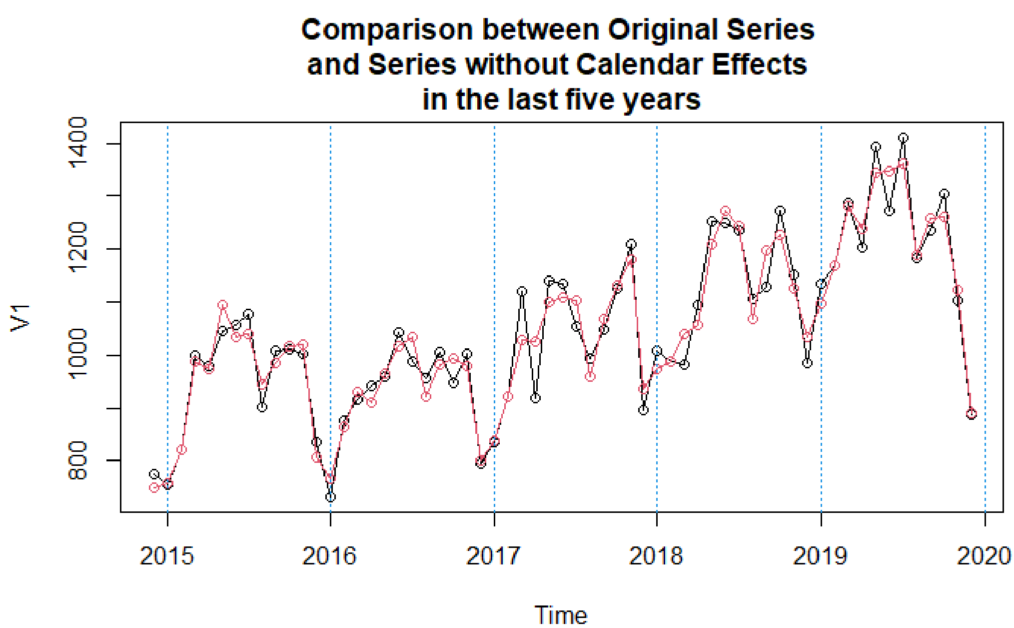

Figure 32: Comparison of Original series VS Without calendar effect in the last five years.

We attempted to validate the selected model.

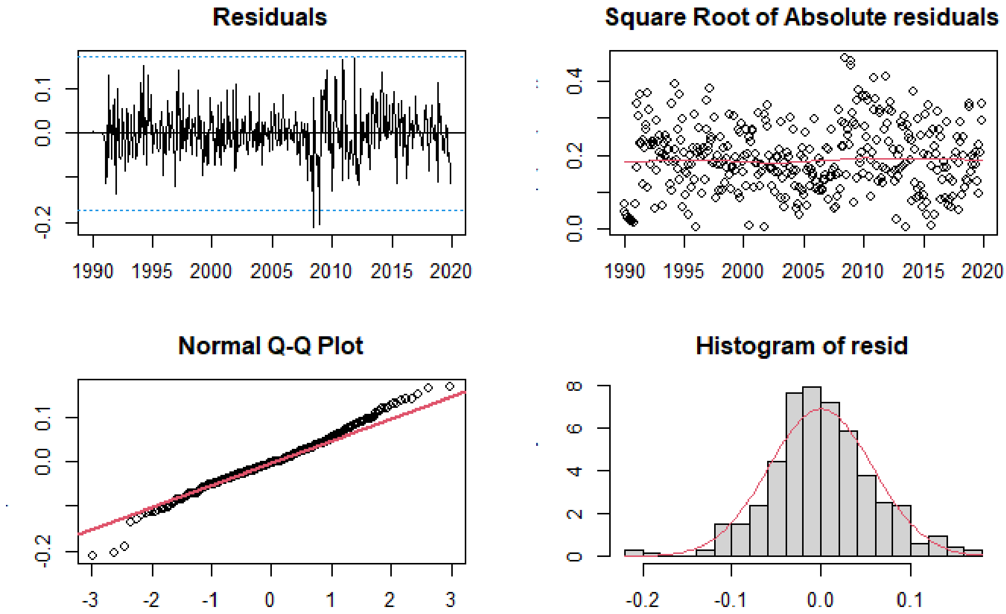

Figure 33: Residuals, model with calendar effects.

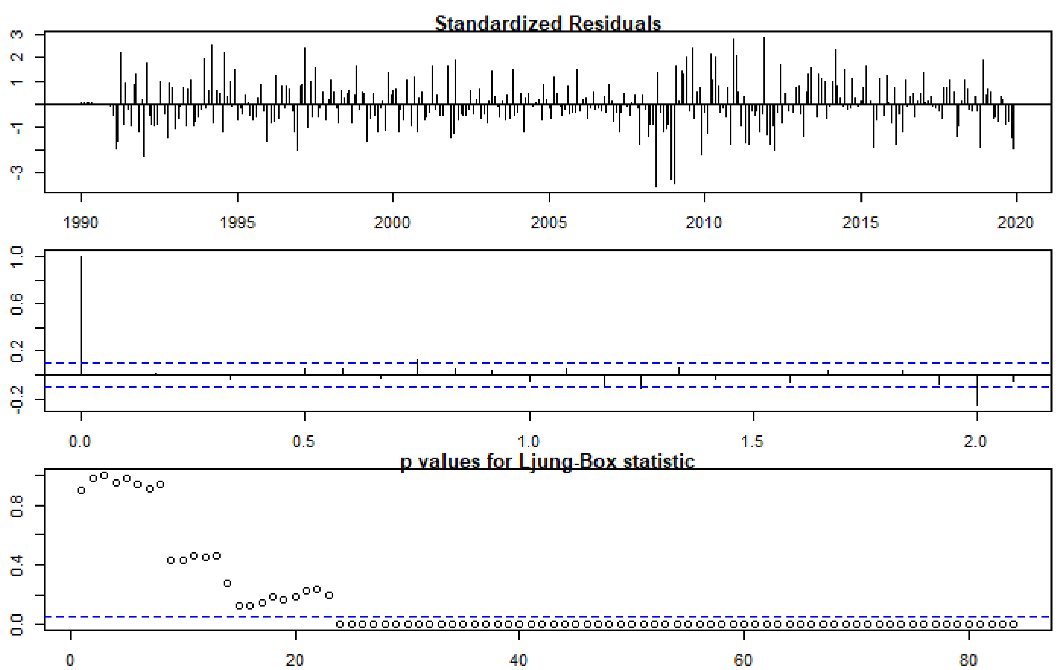

Figure 34: Residuals and Ljung-Box, model with calendar effects.

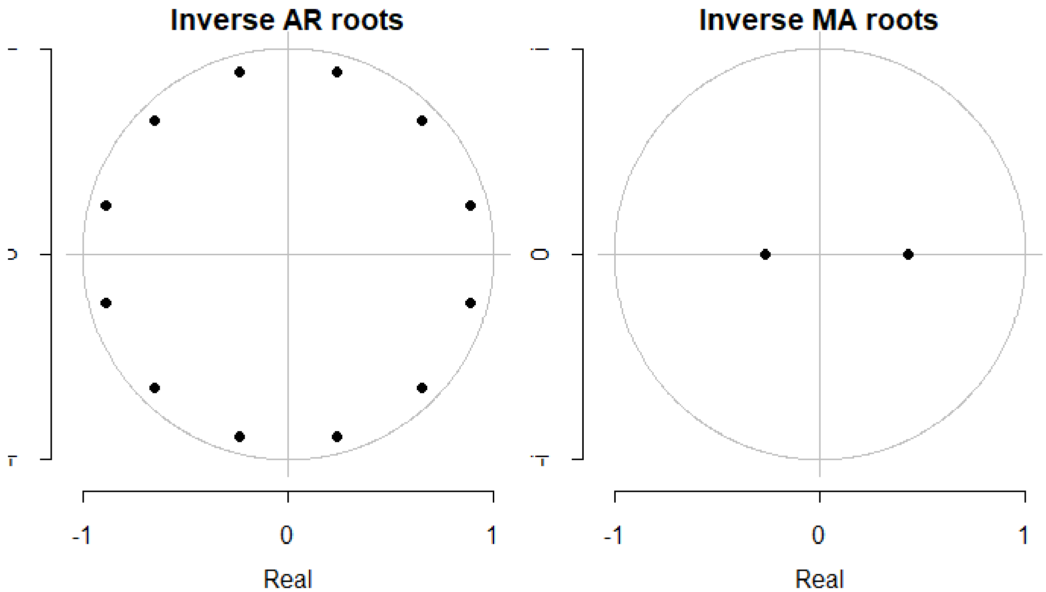

Figure 35: Inverse roots, model with calendar effects.

Figure 33 illustrates that the model residuals exhibit fewer outliers, with a more constant variance. Additionally, there are fewer outliers in the tails of the QQ-plot. In Figure 34, the Ljung-Box statistic indicates no correlation between samples until sample 24. Lastly, Figure 35 demonstrates that the model is both causal and invertible.

7 Outlier Treatment

We employed the *outdetec* function from the *atipics2.r* file to identify outliers in the series. The analysis utilized the previously mentioned model, which accounts for calendar effects, and a *crit* value of 3. The outcome of this analysis is presented in Table 3.

type detected	W coeff	Easter Week ABS L Ratio	Date	perc. Obs
AO	0.08952003	3.076751	Oct 1991	109.36492
AO	0.09057502	3.072872	Jul 1993	109.48036
TC	0.12457798	3.279579	Dic 1994	113.26703
AO	-0.09336765	3.126873	Mar 1996	91.08586
LS	-0.18798302	4.380461	Jun 2008	82.86288
LS	-0.15906805	3.867059	Dic 2008	85.29383
AO	-0.13213997	4.025069	Ene 2009	87.62183
TC	-0.20644367	4.846942	Dic 2009	81.34721
TC	-0.12593453	3.266848	Ene 2010	88.16726
TC	0.10932019	3.078051	Feb 2011	111.55195
TC	-0.13724247	3.508559	Mar 2013	87.17588
LS	0.13394251	3.302360	Dic 2018	114.33271

Table 3: Outliers detected.

Subsequently, with the detected outliers, we generated a new series excluding outliers (*lnserie.lin*) and another series without outliers and calendar effects (*lnserieEC.lin*).

We conducted an investigation to identify significant events occurring in months with detected outliers. Our findings revealed a noteworthy correlation between the majority of outliers and economic events, particularly the onset of the financial crisis in 2008.

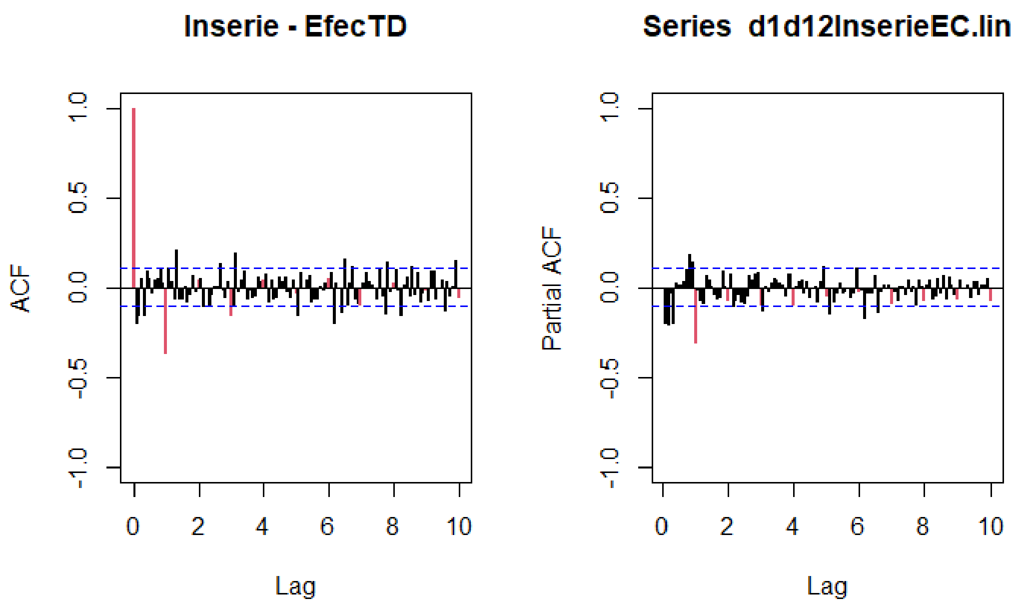

Figure 36: ACF and PACF for *lnserieEC.lin*.

We utilized the series without outliers and calendar effects (*lnserieEC.lin*) to explore and identify new models using ACF and PACF plots (Figure 36). The models identified from these plots include AR(2), AR(4), MA(2), MA(4) for the regular part, and SAR(1), SMA(1), SMA(3) for the seasonal part. Given the 12 potential models, we focus on the superior model, determined by the lower AIC. The optimal model, as illustrated in Figure 37, is identified as AR(2) SMA(3).

```
Call:
arima(x = lnserie.lin, order = c(2, 1, 0), seasonal = list(order = c(0, 1, 3),
    period = 12), xreg = data.frame(wTradDays, wEast))

Coefficients:
          ar1      ar2      sma1      sma2      sma3  wTradDays     wEast
      -0.2408  -0.1721   -0.6424   -0.1042   -0.2533     0.0114   -0.1015
s.e.   0.0531   0.0536    0.0810    0.0654    0.0557     0.0006    0.0064

sigma^2 estimated as 0.001754:  log likelihood = 592.38,  aic = -1168.75
      ar1       ar2      sma1      sma2      sma3 wTradDays     wEast
 4.532866  3.210997  7.929110  1.594150  4.549096 18.605050 15.887434
```

Figure 37: Best model with outliers and calendar effects.

We employed the identified model (AR(2) SMA(3)) to visualize the impact of outliers and to generate a series without calendar effects and outliers (Figures 38 and 39).

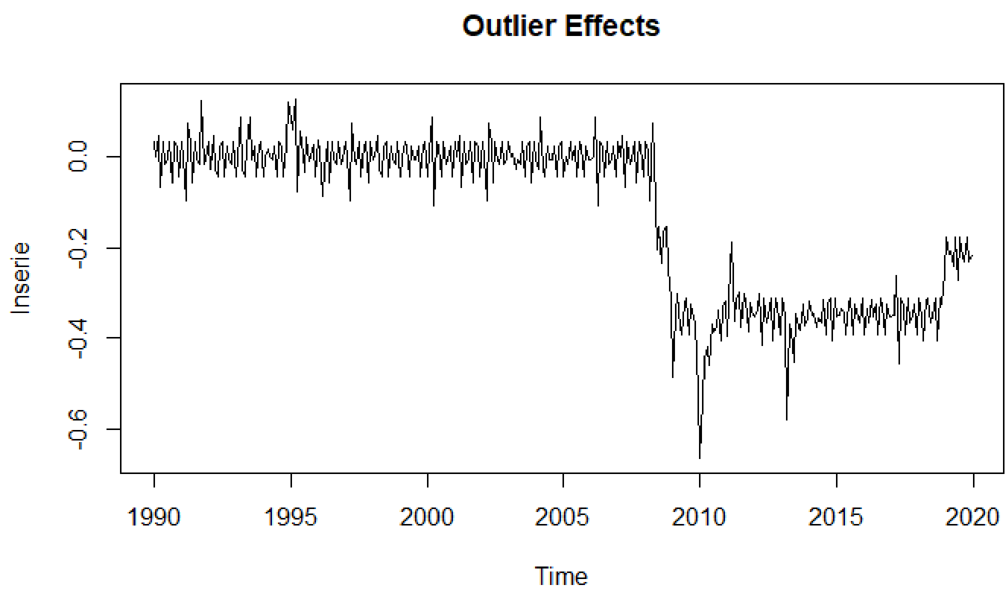

Figure 38: Outliers effect.

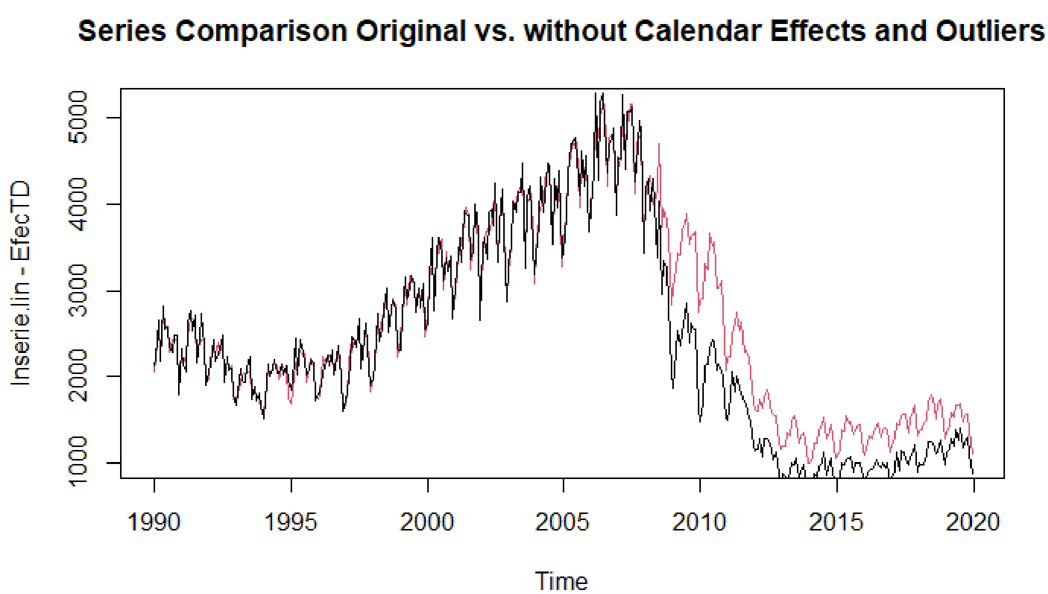

Figure 39: Comparison between original series and series without outliers nor calendar effects.

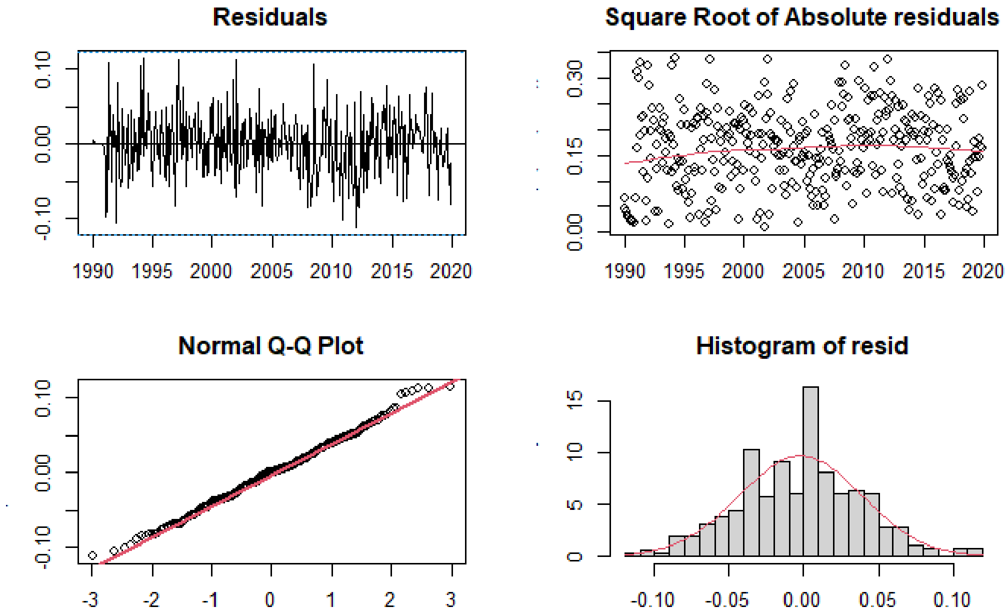

Figure 40: Residuals, model with calendar effects and without outliers.

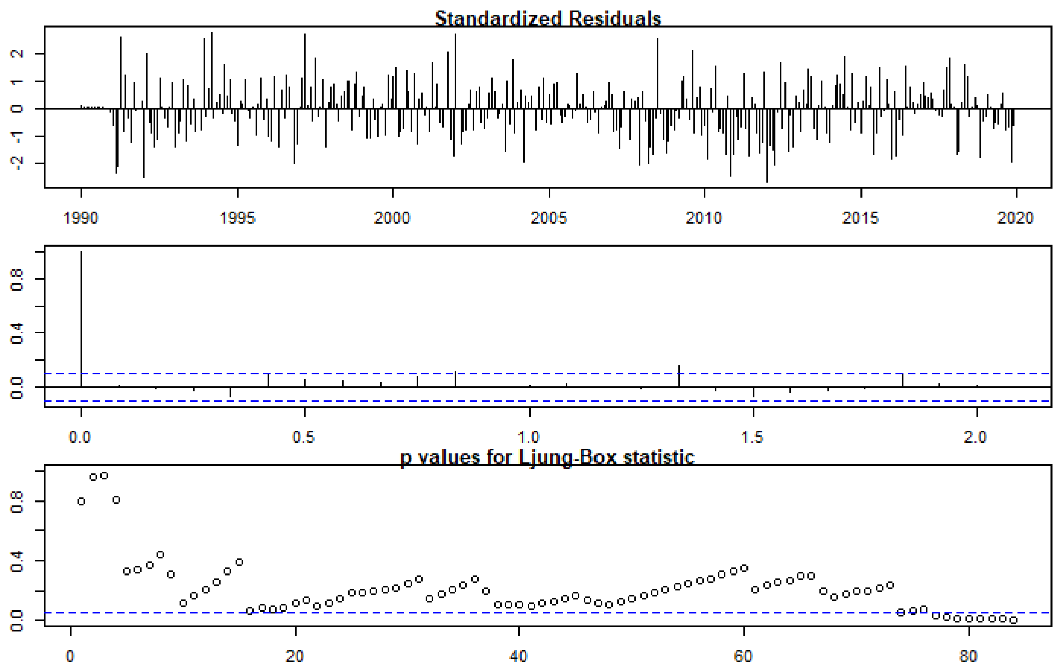

Figure 41: Residuals and Ljung-Box, model with calendar effects and without outliers.

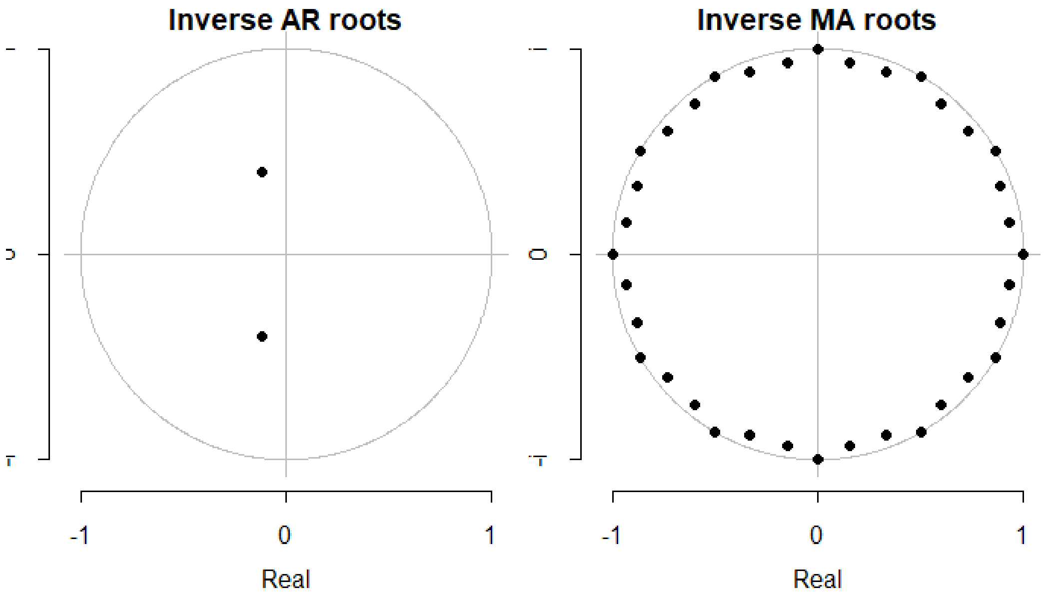

Figure 42: Inverse roots, model with calendar effects and without outliers.

From Figure 40 we can see that the model residuals have less outliers, the variance of the residuals looks that is not constant. And there are less outliers in the tales of QQ-plot. From Figure 41 we can see that Ljung-Box statistic that there is no correlation between samples until sample 77. And from Figure 42 we can see that the model is causal and invertible. We also check the numeric values of MA roots to check that they are greater than 1.

Figure 40 reveals that the model residuals exhibit fewer outliers, moreover the variance of the residuals appears to be constant. Additionally, there are fewer outliers in the tails of the QQ-plot. In Figure 41, the Ljung-Box statistic suggests no correlation between samples until sample 77. Moreover, Figure 42 demonstrates that the model is both causal and invertible. We also verified the numeric values of MA roots to ensure they are greater than 1.

8 Long Term Predictions

Using the model incorporating calendar effects and outlier treatment, we proceed to make predictions.

Initially, we verify the stability of the model by comparing the coefficients between the model trained with the entire series and the model excluding the data from the last year (Figure 43).

```
Call:
arima(x = lnserie.lin1, order = pdq, seasonal = list(order = PDQ, period = 12),
    xreg = data.frame(wTradDays, wEast))

Coefficients:
          ar1      ar2     sma1     sma2     sma3  wTradDays     wEast
      -0.2408  -0.1721  -0.6424  -0.1042  -0.2533     0.0114   -0.1015
s.e.   0.0531   0.0536   0.0810   0.0654   0.0557     0.0006    0.0064

sigma^2 estimated as 0.001754:  log likelihood = 592.38,  aic = -1168.75

Call:
arima(x = lnserie.lin2, order = pdq, seasonal = list(order = PDQ, period = 12),
    xreg = data.frame(wTradDays2, wEast2))

Coefficients:
          ar1      ar2     sma1     sma2     sma3  wTradDays2    wEast2
      -0.2516  -0.1811  -0.6447  -0.0858  -0.2695      0.0114   -0.1009
s.e.   0.0540   0.0545   0.0807   0.0669   0.0565      0.0006    0.0065
```

Figure 43: Model with all the time series and without last year.

Second, we check the predictions of year 2019 (Figure 44).

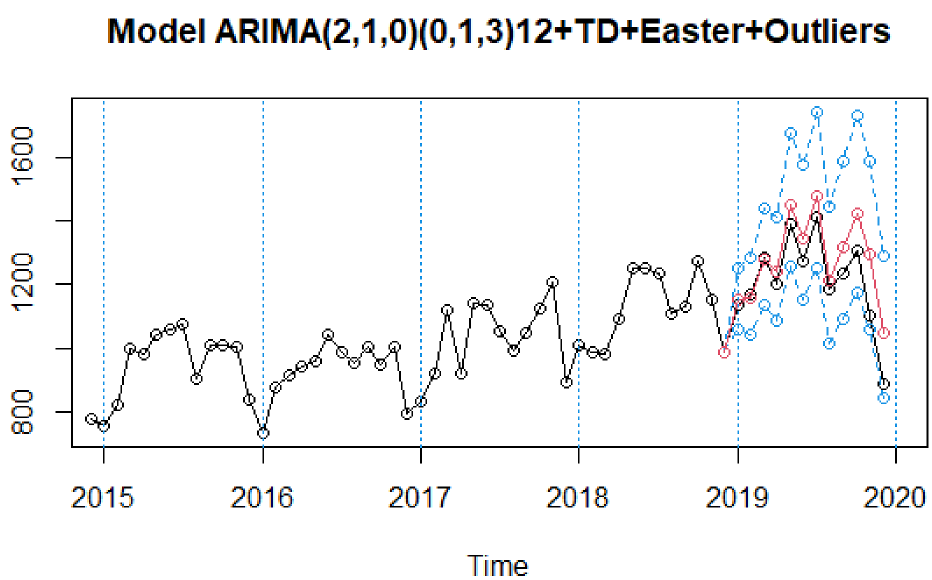

Figure 44: Predictions 2019.

In Figure 44, it is evident that the model tends to overestimate concrete consumption in the last moth. This overestimation could potentially be attributed to the uncertainties surrounding the initial cases of COVID-19.

Then, we predict the year 2020 (Figure 45).

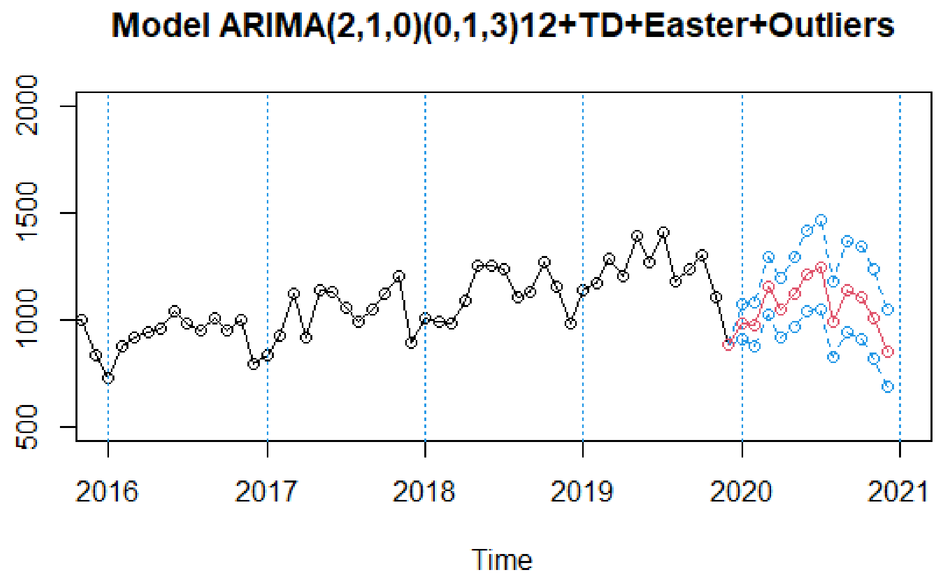

Figure 45: Model with all the time series and without last year.

The model predicts a decline in concrete consumption, potentially influenced by the decrease observed in December 2018. Upon examining the data for the year 2008, we observe that December marked the lowest month of that year, preceding a sustained decline in concrete consumption. This historical pattern leads us to interpret the model's projection as indicative of an anticipated downturn.

Models	RMSE	MAE	RMSPE	MAPE	CI
Original	103.958	84.136	0.09019	0.07209	502.150
With calendar effects	101.829	73.809	0.10131	0.06742	518.881
With calendar effects and Outliers	90.809	71.147	0.08380	0.06204	374.089

Table 4: Error comparison of three models predicting year 2019.

It is evident that the best model is the one incorporating calendar effects and outlier treatment, as it exhibits the lowest error (across all metrics) and the highest precision, reflected in the narrowest confidence intervals.

9 Conclusions

In this project we have meticulously analysed a time series showing the cement consumption in Spain from 1990 to 2019. We applied the techniques and methods seen during this course and made conclusions from the discovered insights.

We have observed that, evidently, certain seasonal patterns, such as a preference for vacations in August or reduced construction activity during winter, contribute to a lower consumption of concrete.

Furthermore, the real estate bubble burst is apparently a result of the 2008 crisis, which led to a substantial decline in concrete consumption and can be seen in the trend of the analysed time series.

Given the significant impact of the economy on the construction sector, and subsequently on concrete consumption, it is challenging for a model to achieve high accuracy without accounting for these influential factors.