línea horizontal

Homework 3

Time Series

**Statistical Data Analysis – 10th of January, 2020**

**Group 5**

Antoñanzas Martínez, Guillermo

Pueblas Núñez, Rodrigo

Martínez Capella, Ignacio

Burrell, David

## 

## 

## 

## 

## Brief data description, preparation and manipulation

We were assigned the dataset 5 (“data\_g5.xlsx”). In this dataset, we can see the percentage of housing approvals by month that were subsidised. Observations were taken at the end of each period and correspond to data from January 1990 to December 2007. The source is Banco de España ([www.bde.es](http://www.bde.es)). No cleaning was needed, and only two columns were present (Date and Percentage).

## 

## Research questions, plots and findings

### **Question 1**

### **Plot the series and briefly comment on the characteristics you observe (stationary, trend, seasonality…).**

### C:\Users\rodri\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\A4948650.tmpC:\Users\rodri\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\97EAD0DE.tmp

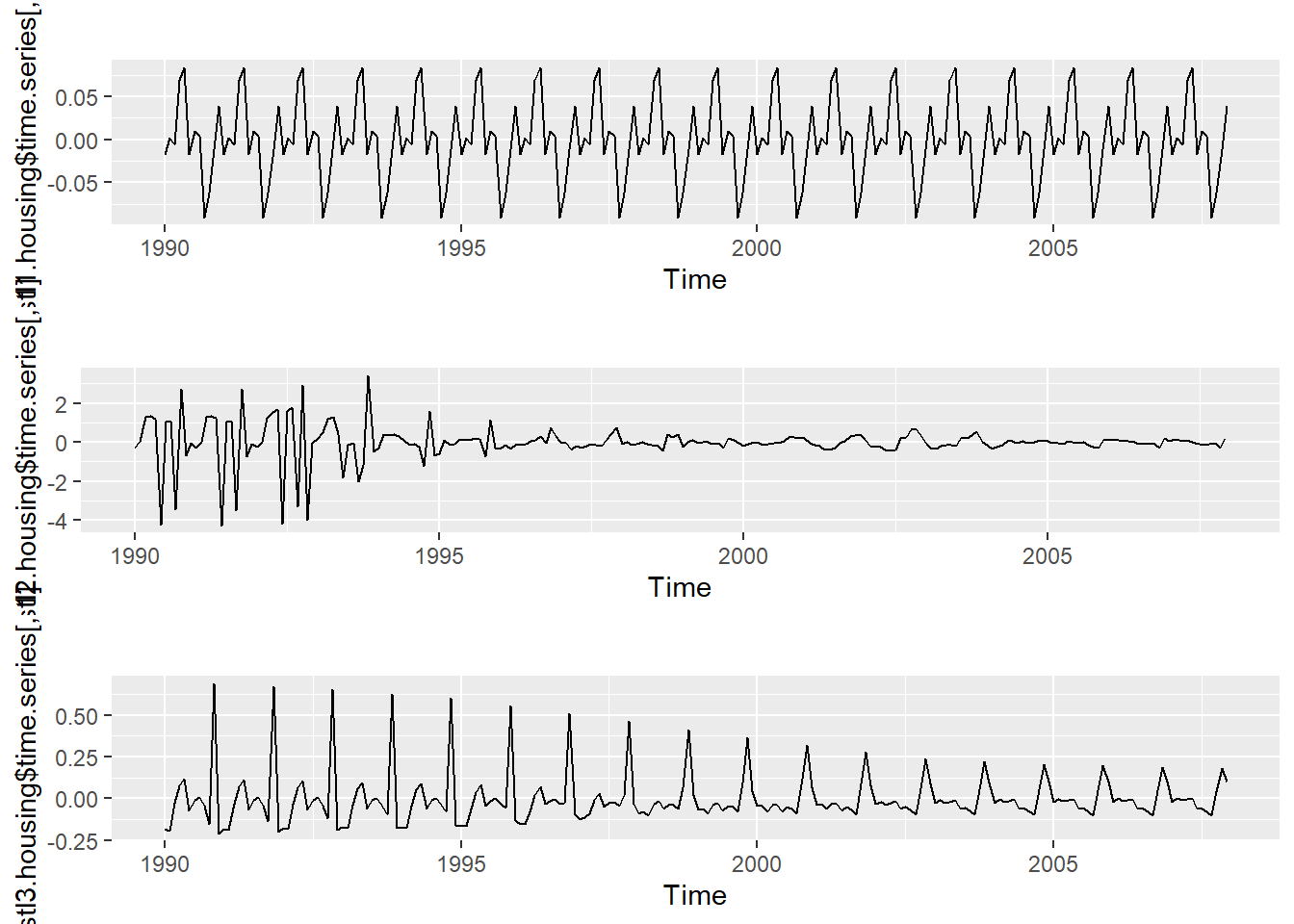
### We can address each of the properties separately:

* Trend: There is a trend over time, and we can see different patterns. There is a changing trend that has a positive slope between 1991 and 1993, a negative slope between 1993 and 2003, and a positive slope from 2003 and onwards.
* Seasonality: It seems that there is no seasonality, as there is not a regularly repeating pattern related.
* Cyclical components: There is no evidence of any cyclic behaviour in the data.

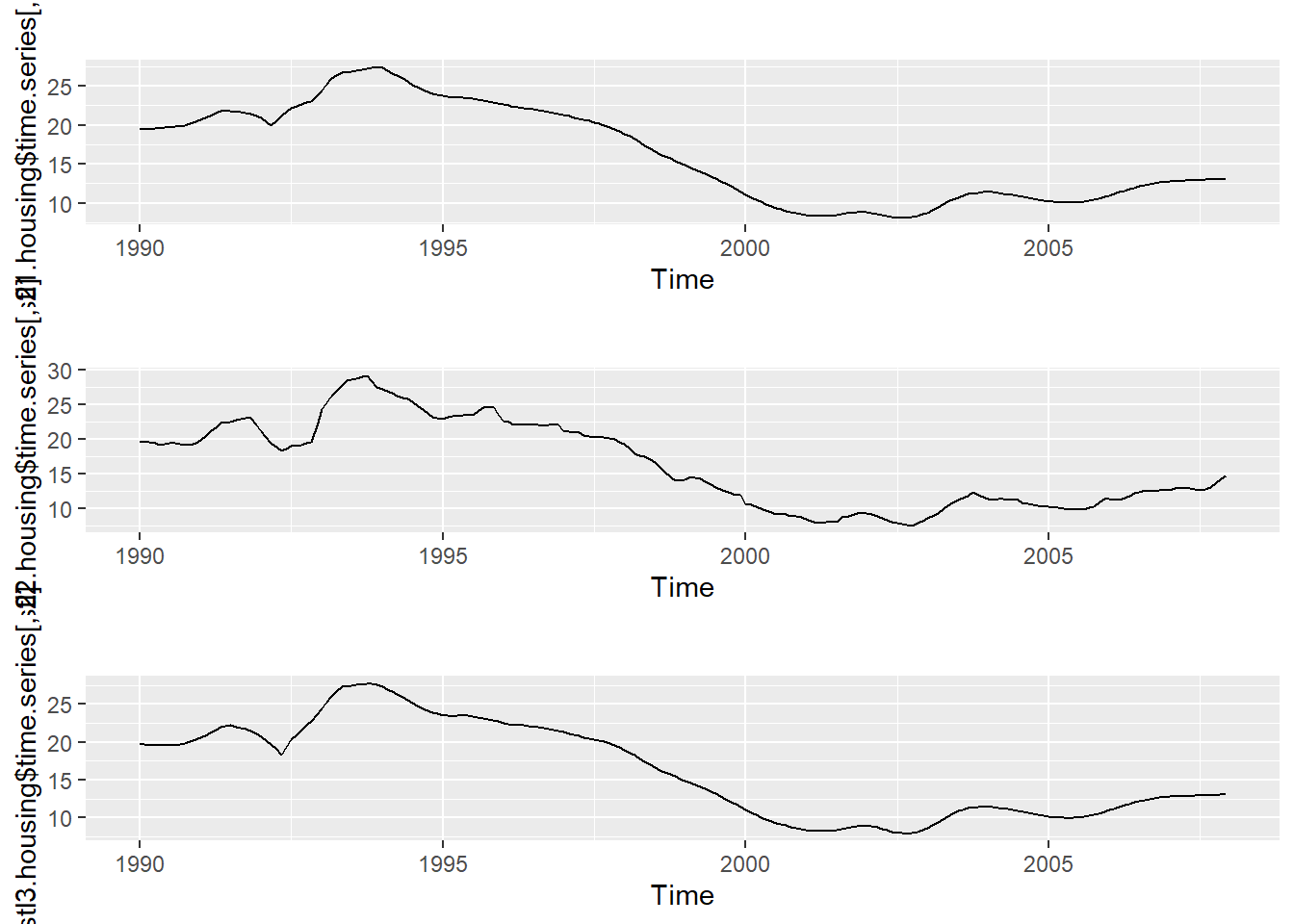
Taking these observations into account, we can state that we are working with a non-stationary time series. Even if the time series shows no seasonal or cyclical components, the presence of trends makes it non-stationary.

### **Question 2**

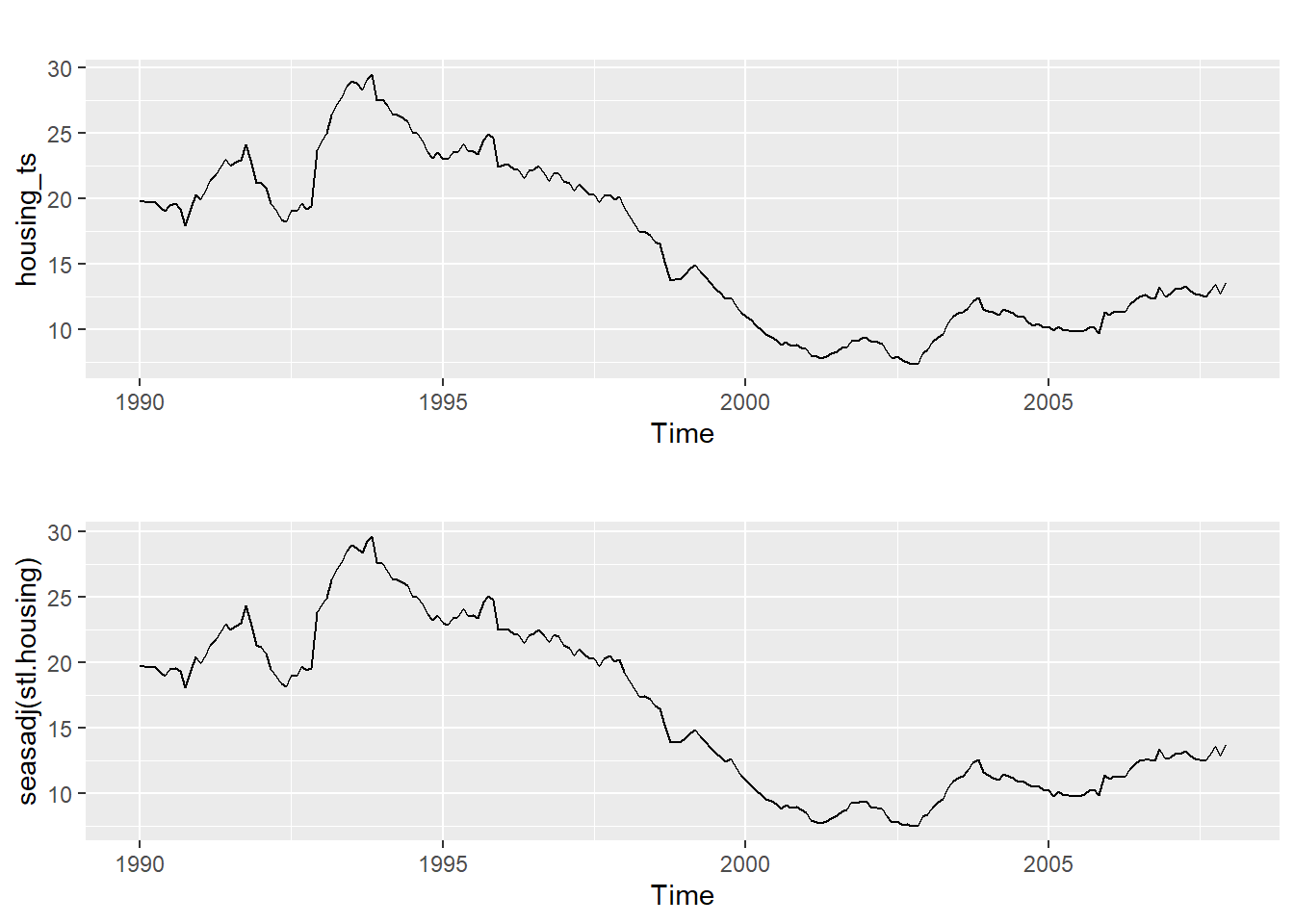
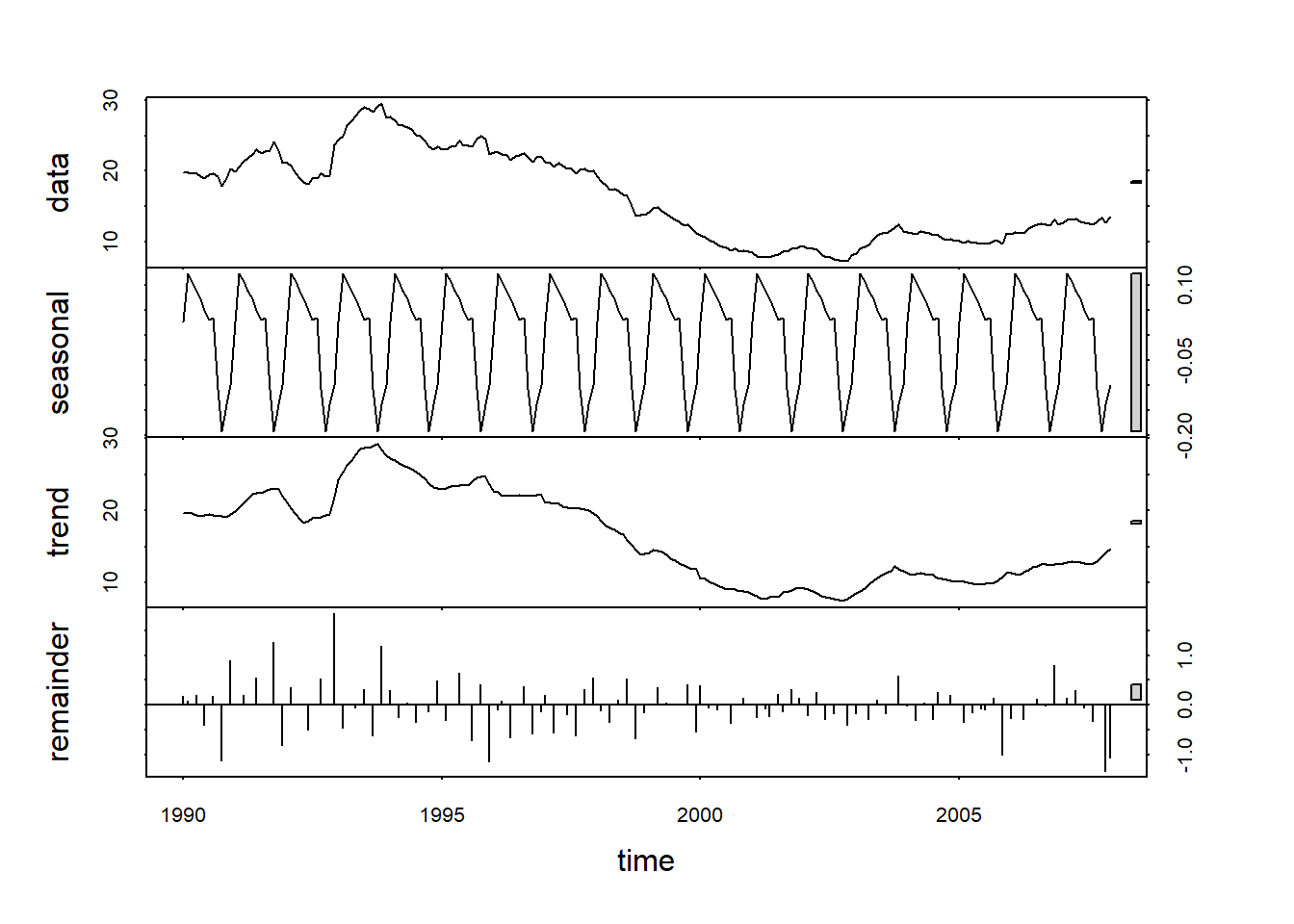
**2a) Obtain a plot of the decomposition of the series, using stl(). Use an additive decomposition or a multiplicative one, depending on your data.**



We have used an additive decomposition, since there are no changes in seasonal components. We can see that using the first seasonal window (s.window = periodic), the seasonal components are small, centered around zero and more or less constant. Since the seasonality plots obtained before showed no influence of seasonal components, we will choose this seasonal window.

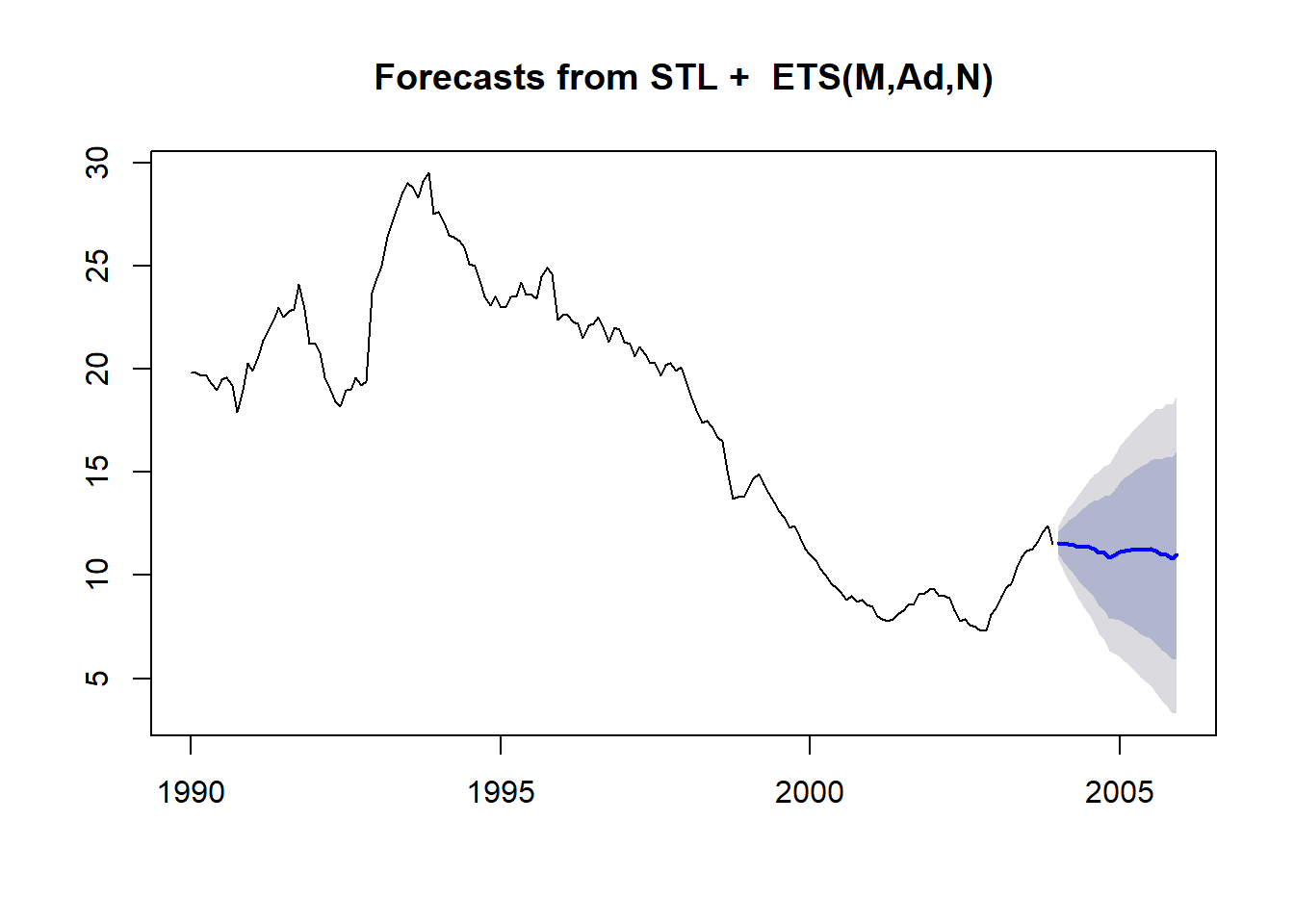
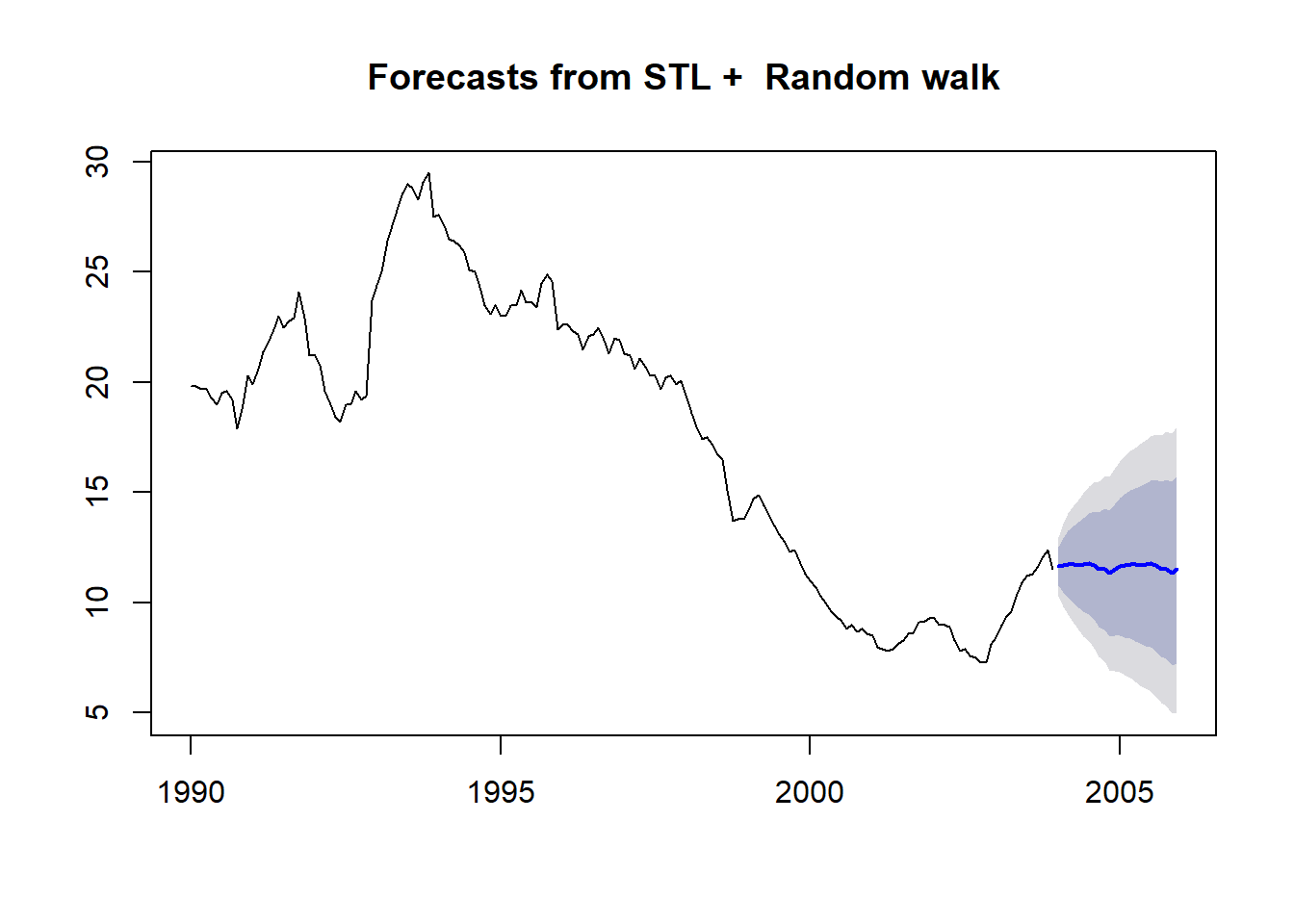


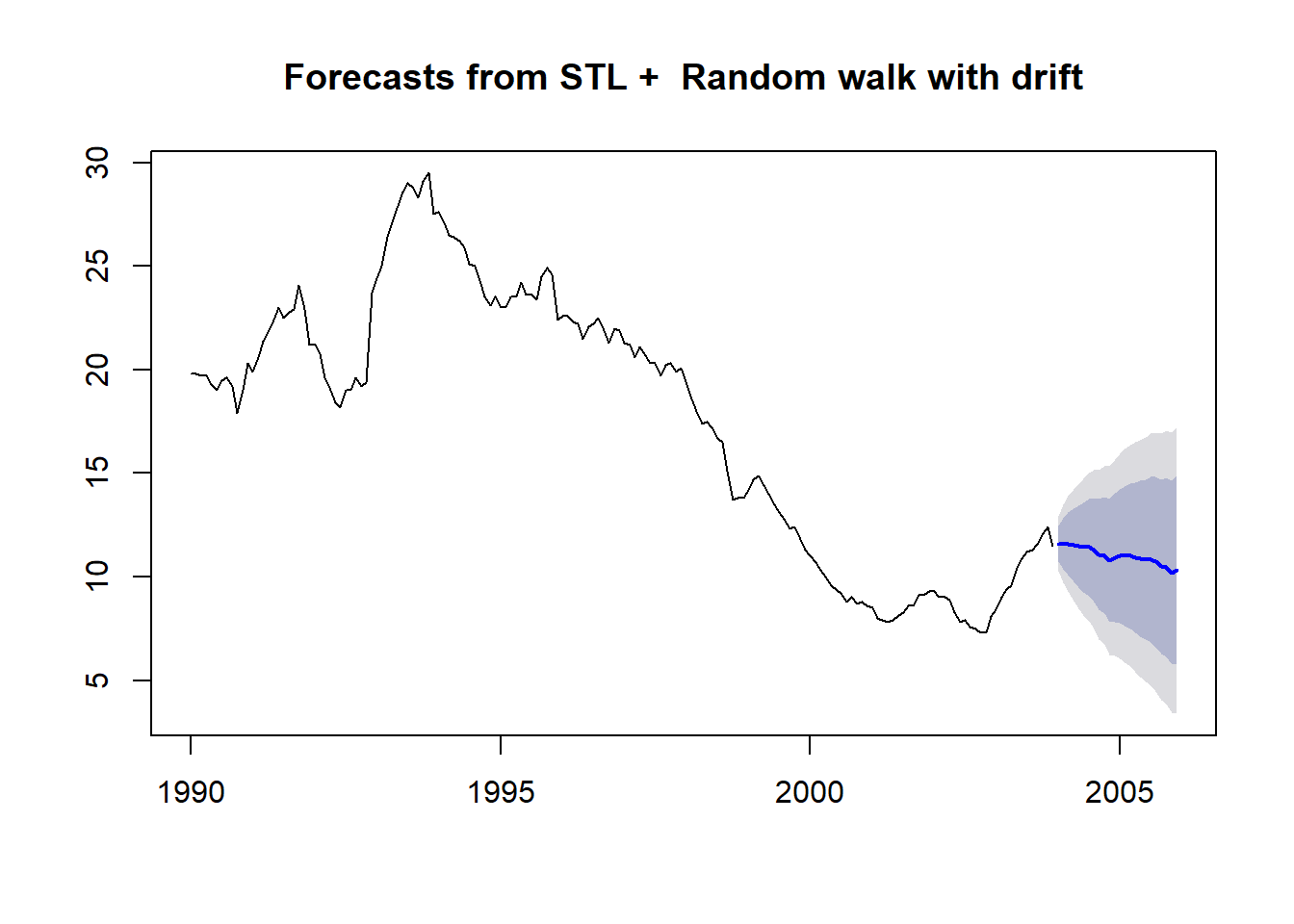
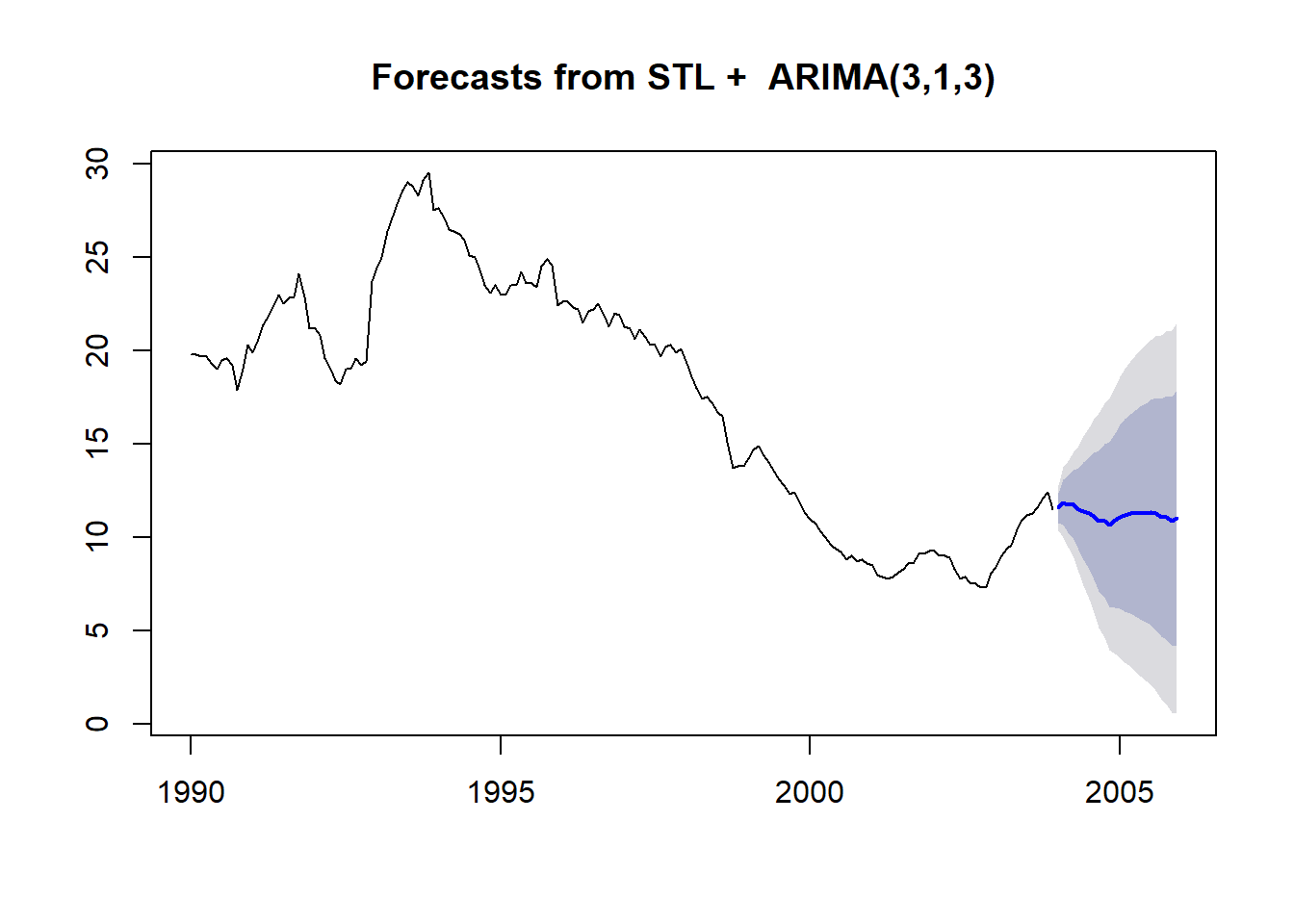
Keeping a periodic window for the seasonal components (assuming they are equal across the years), we can see that using the second trend window (t.window = 5) produces the trend that best caputres the structure of the data. Thus, we will choose it.



Watching the decomposition plot resulting from the selected windows, we can see that the seasonal components have low constant values that are centred around zero ([-0.192, 0.124]). The structure of the data is captured well by the trend component and the seasonal adjusted series is just like the original one. The remainder values are somewhat big ([-1.34, 1.84]), but have zeron mean (-0.0153).

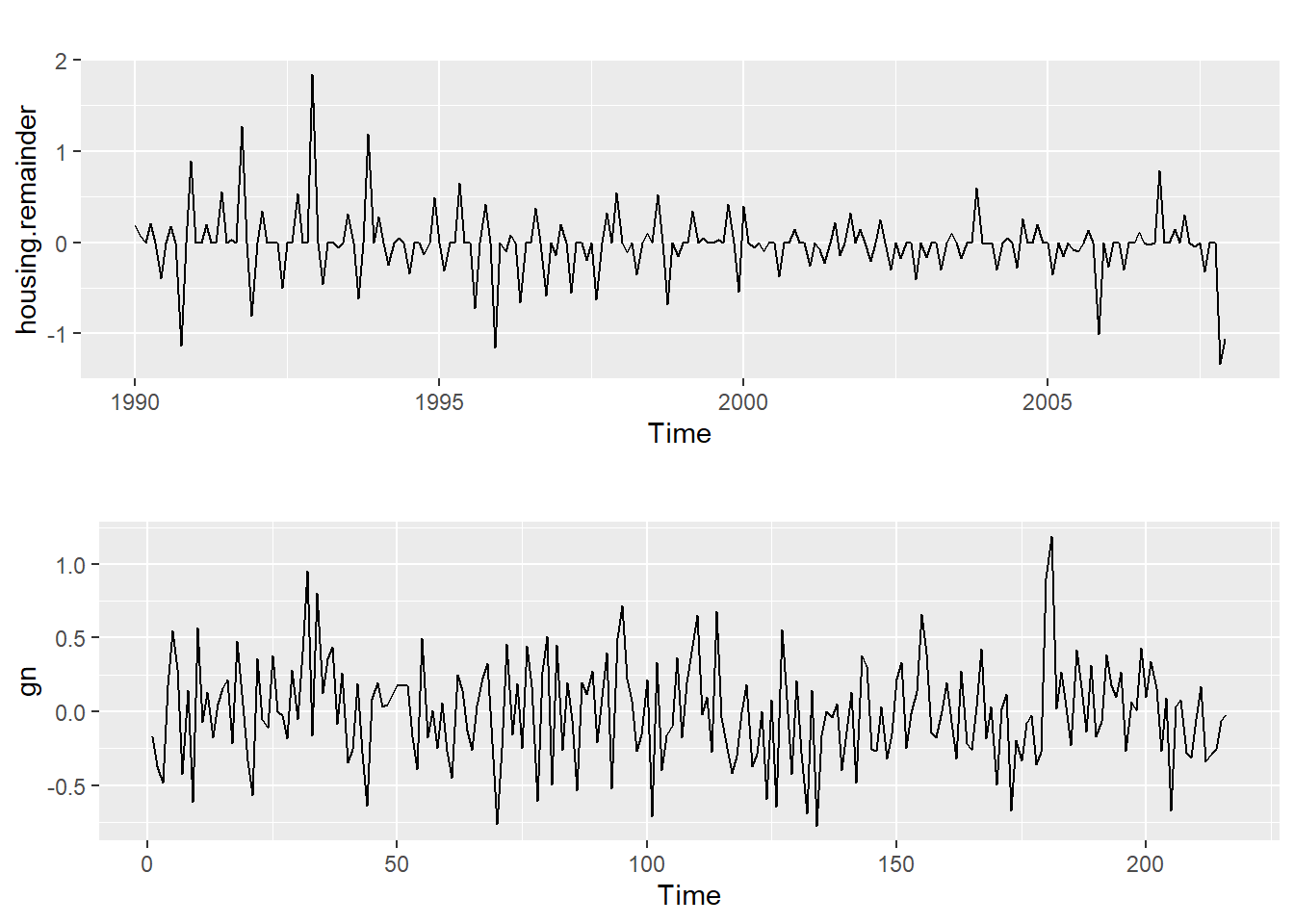
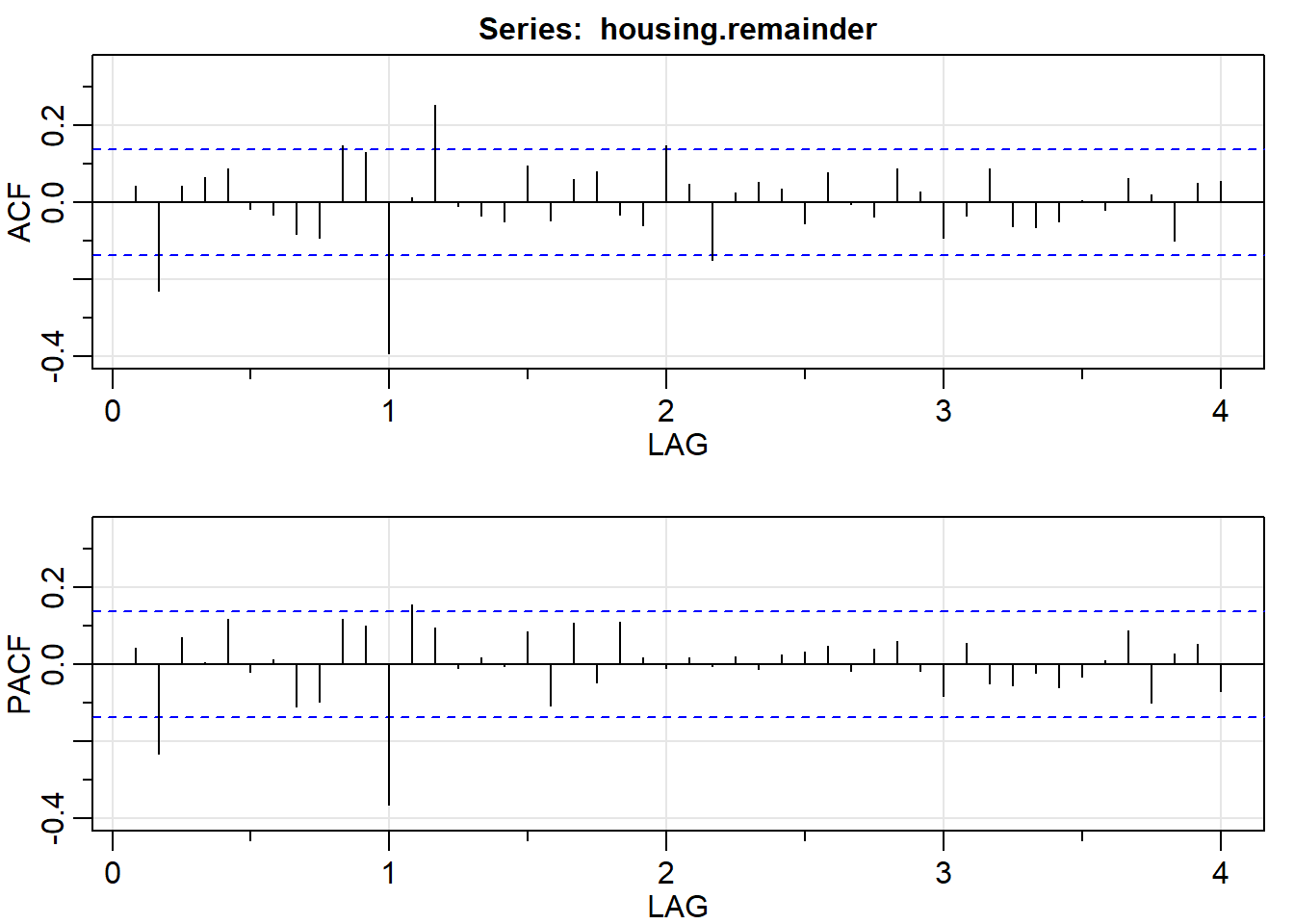
**Use the function forecast() to forecast future values.**





Using stl() to forecast, we can see that using the random walk with drift method to predict future values results in the lowest RMSE.

**Does the remainder look like a white noise to you?**

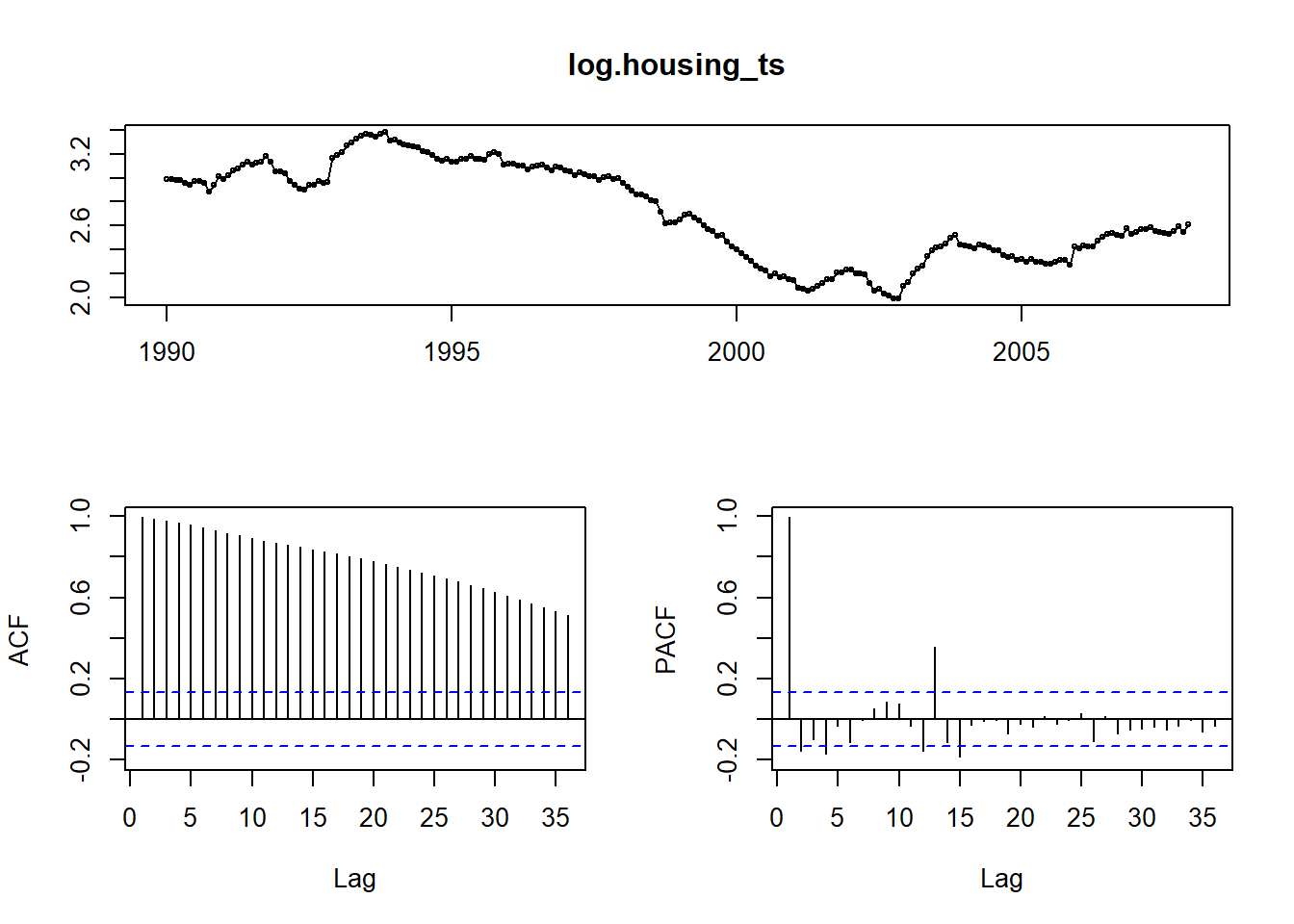
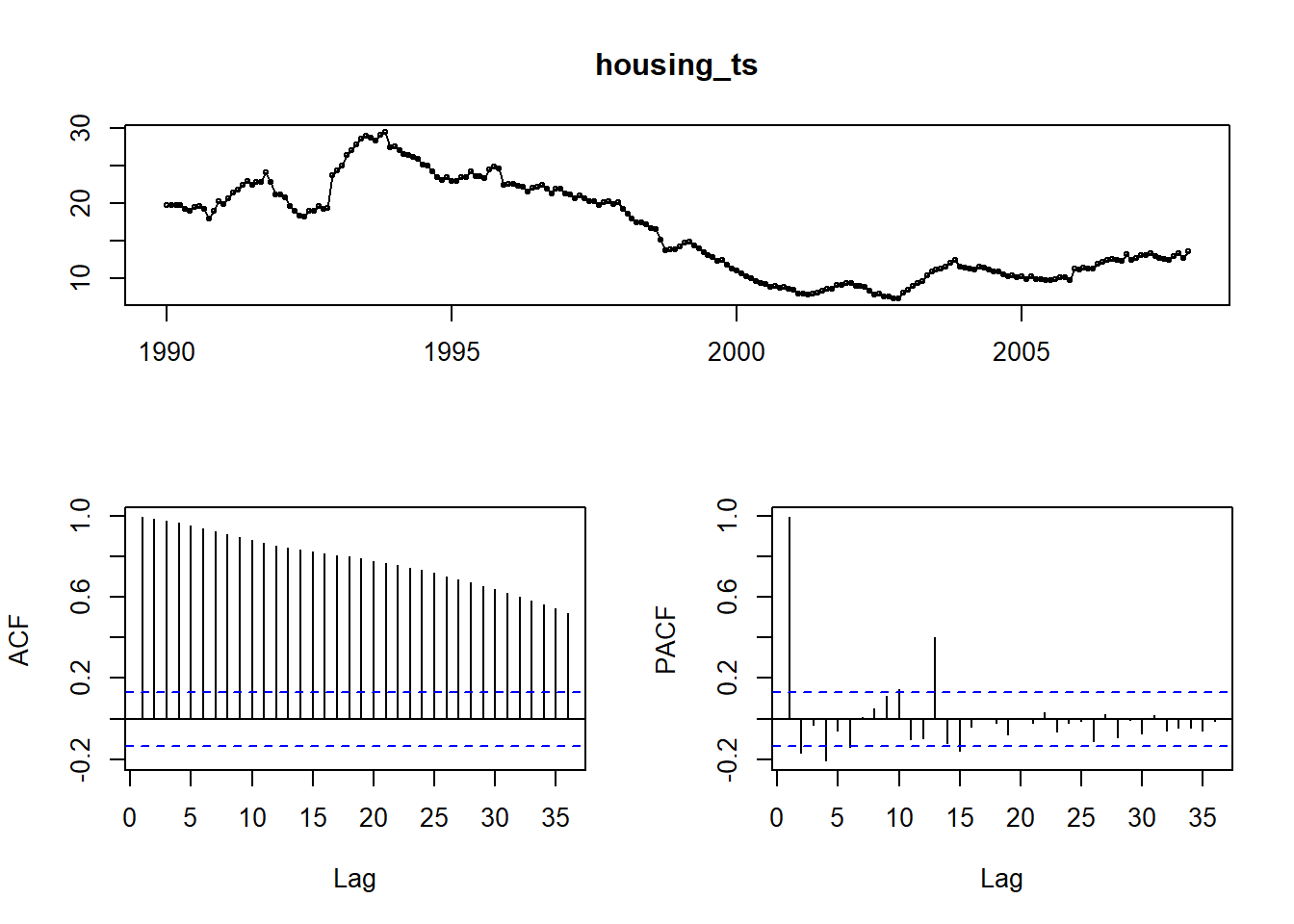


The remainder does not look like gaussian noise, since it does not have a constant variance. In the preceding years, the data has much more variance than in the most recent ones.

### **Question 3**

**Fit an ARIMA model to your time series. Some steps to follow:**

### **3a) Decide on whether to work with your original variable or with the log transform one.**

There is barely any difference, so we will use the original data.

### **3b) Are you going to consider a seasonal component? If the answer is yes, identify s.**

As previously stated, we can see that there is no seasonality, and this can be seen pretty easily by using the function ggseasonplot with the chosen data. We also plotted the Seasonal Decomposition to see that the values of the season are really low, between -0.2 and 0.1.

### **3c) Decide on the values of d and D to make your series stationary.**

A formal ADF test does not reject the null hypothesis, but the differenced time series does. We get a p-value=0.1, so we can reject the hypothesis of non-stationarity. This indicates that it may be a good idea to consider differencing the data, given that the chosen default data structure may be somewhat stationary. From now on, we will be using the differenced time series.

### **3d) Identify values for p and q for the regular part and P and Q for the seasonal part. Check also the correlation between the coefficients of the model.**

TODO

### **3e) Make diagnostic of the residuals for the final model chosen (autocorrelations, zero mean, normality). Use plots and tests.**

TODO

### **3f) Once you have found a suitable model, repeating the fitting model process several times if necessary, use it to make forecasts. Plot them.**

TODO

### **3g) Use the function getrmse to compute the test set RMSE of some of the models you have already fiited. Which is the one minimizing it? Use the last year of observations (12 observations for monthly data, 4 observations for quarterly data) as the test set.**

TODO

### **3h) You can also use the auto.arima() function with some of its parameters fixed, to see if it suggests a better model that the one you have found. Don’t trust blindly its output. Automatic found models aren’t based on an analysis of residuals but in comparing some other measures like AIC. Depending on how complex the data set is, they may find models with high values for p, q, P or Q (greater than 2).**

TODO