



# Gasoline's price prediction in Italy

Time series for Business and Financial Data

by Pietro Ferrazzi, Sergio Postigo, Md Jamiur Rahman Rifat

# Project specifications

- **Monthly** predictions, 3 months in advance prediction
- Price of gasoline for **final consumer (EUR)**
- Use **any data** that you find **relevant**



# Methodology

Data Collection



Propose additional predictors and look for their data sources

Exploratory Data Analysis



Analyze the relationship between predictors and the response

Data Preprocessing



Cleaning and merging the data  
Infer missing values

Data Modeling



Propose and fit different families of models

# Data Collection

Complementary Datasets



Monthly oil's price in Italy (1996-today)



Yearly vehicle gas emissions (1990-today)



Monthly employment rate in Italy (2004-today)

Main Dataset



Ministero dell'Ambiente  
e della Sicurezza Energetica



Daily EUR-USD exchange rate (2004-today)

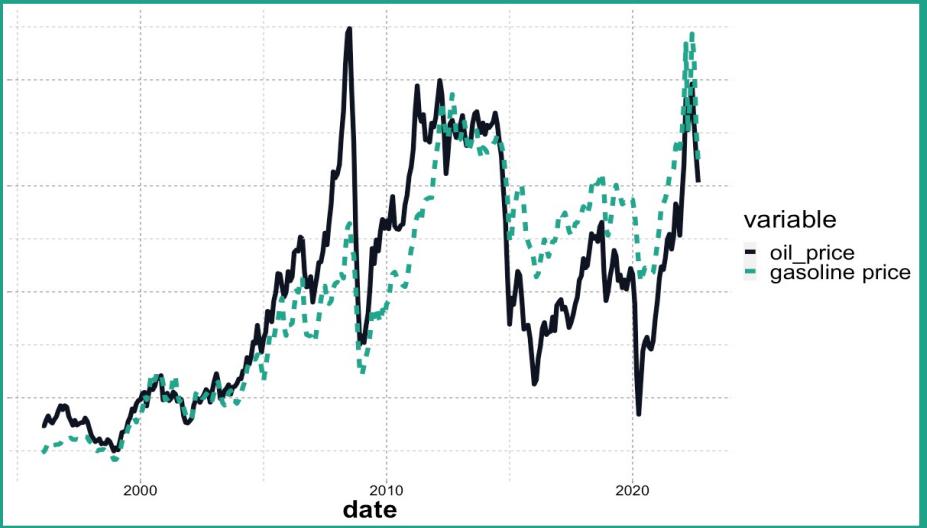


Daily ENI stock price (1996-today)

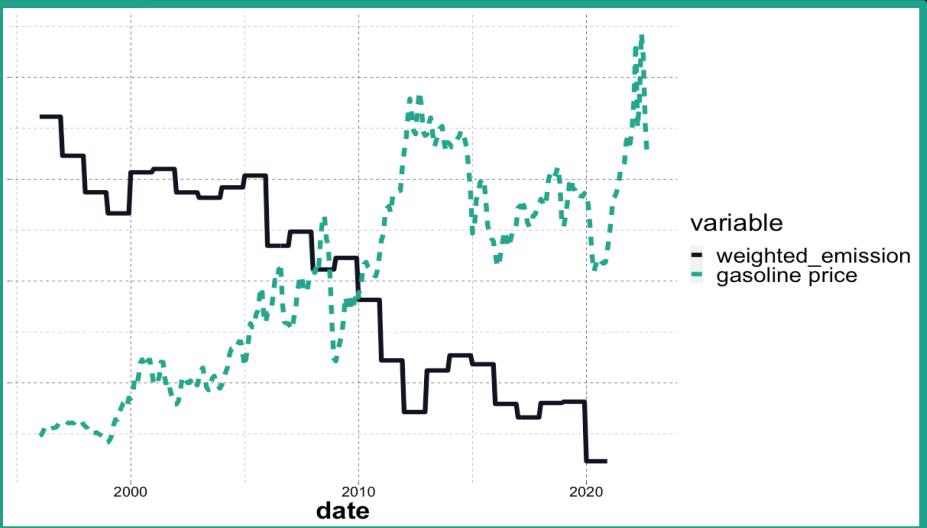
Monthly gasoline's price in Italy (1996 – today)

# Exploratory Data Analysis

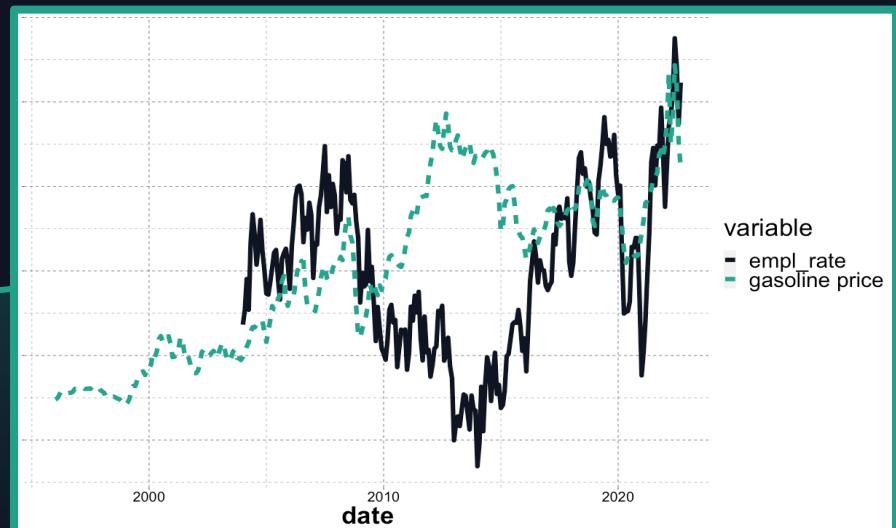
*Gasoline's price vs Oil's price*



*Gasoline's price vs Vehicle Emissions*

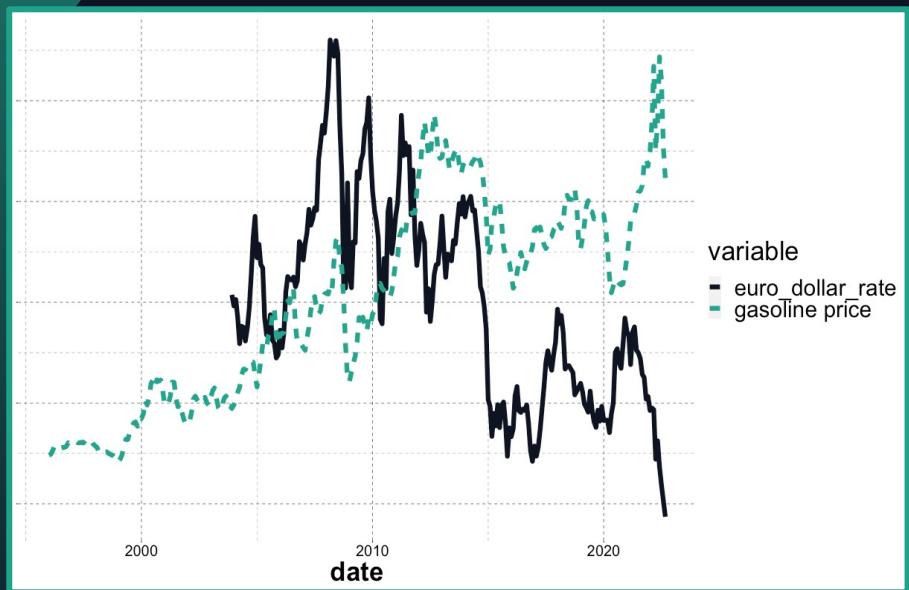


*Gasoline's price vs Employment rate in Italy*

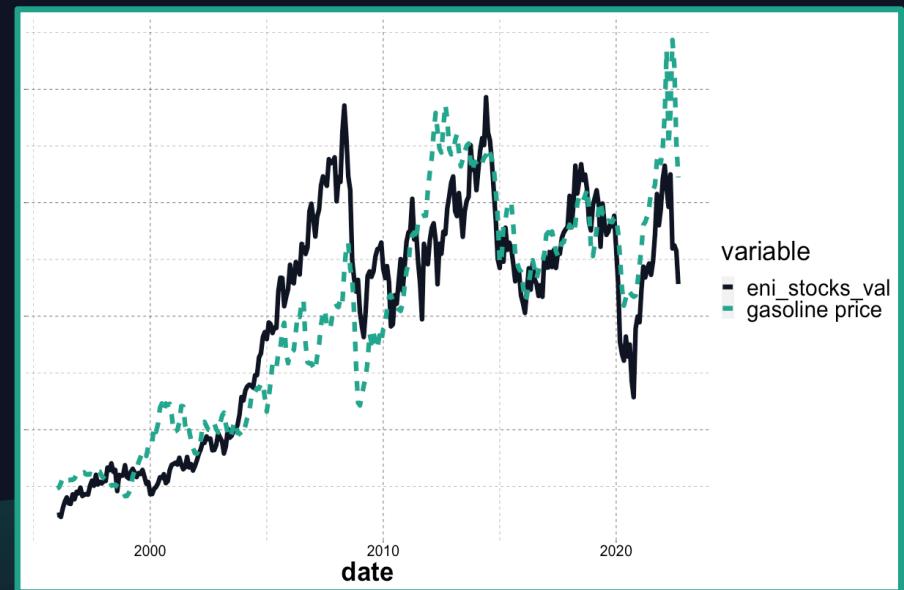


# Exploratory Data Analysis

*Gasoline's price vs Euro-Dollar rate*



*Gasoline's price vs Eni stock's*



# Data Preprocessing



# Data Preprocessing



# Data Preprocessing

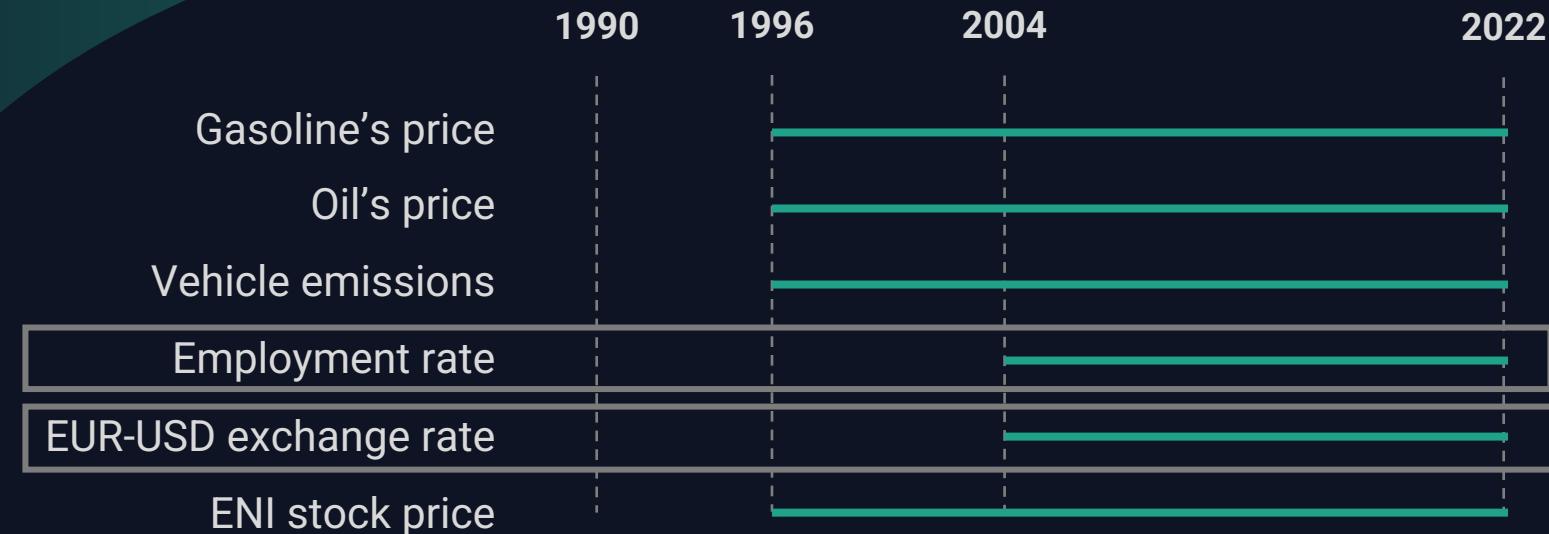


# Data Preprocessing



# Data Preprocessing

POLYNOMIAL  
REGRESSION



# Data Preprocessing

POLYNOMIAL  
REGRESSION

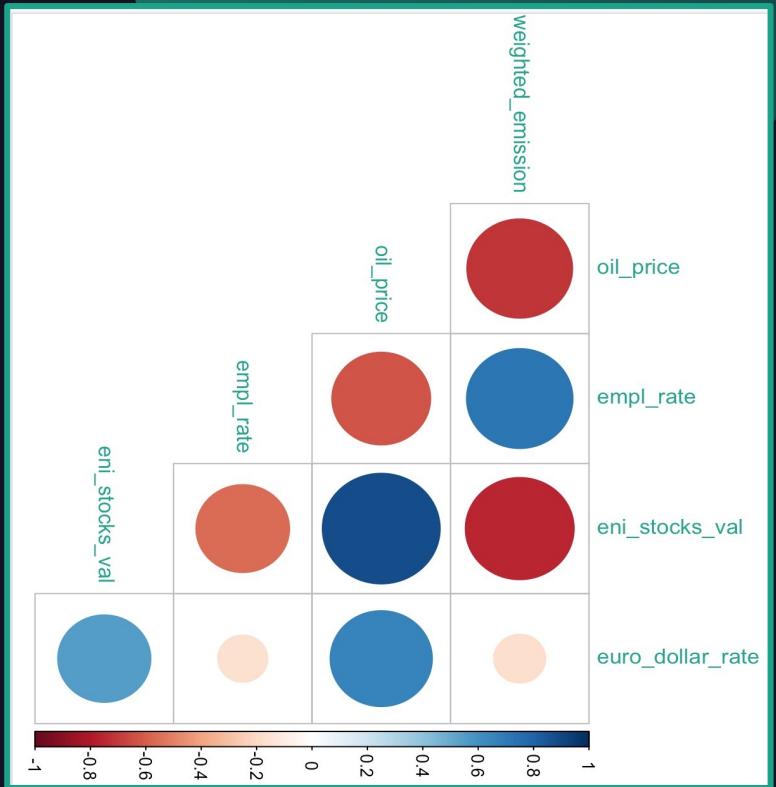


# Data Preprocessing

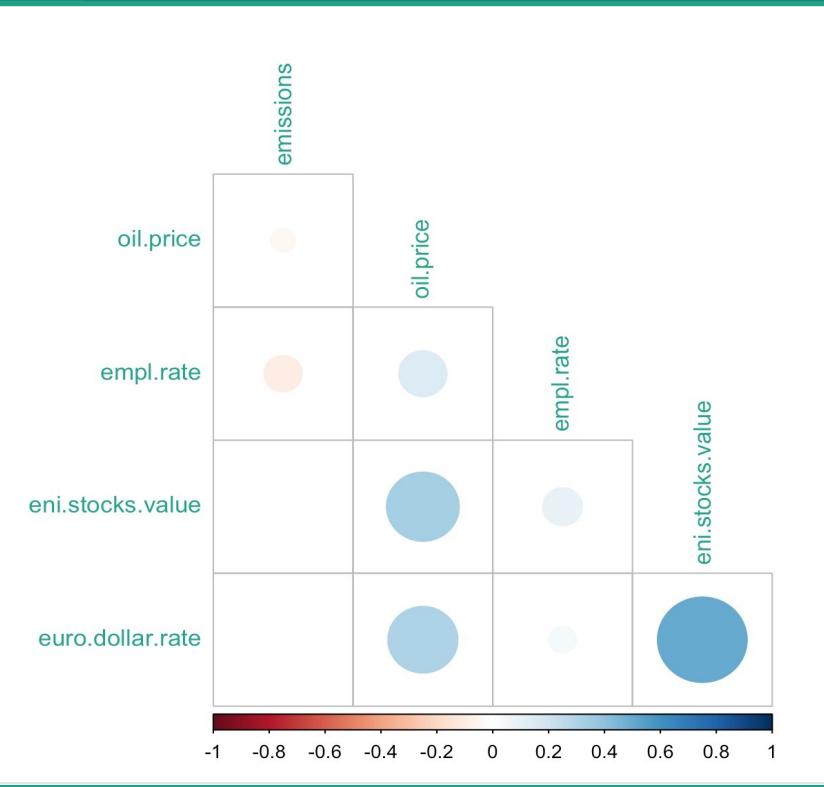


# Exploratory Data Analysis

Before  
removing time  
dependency



Correlation matrix without the time-dependent component



After removing  
time dependency

# Data Modeling

## Linear Regression

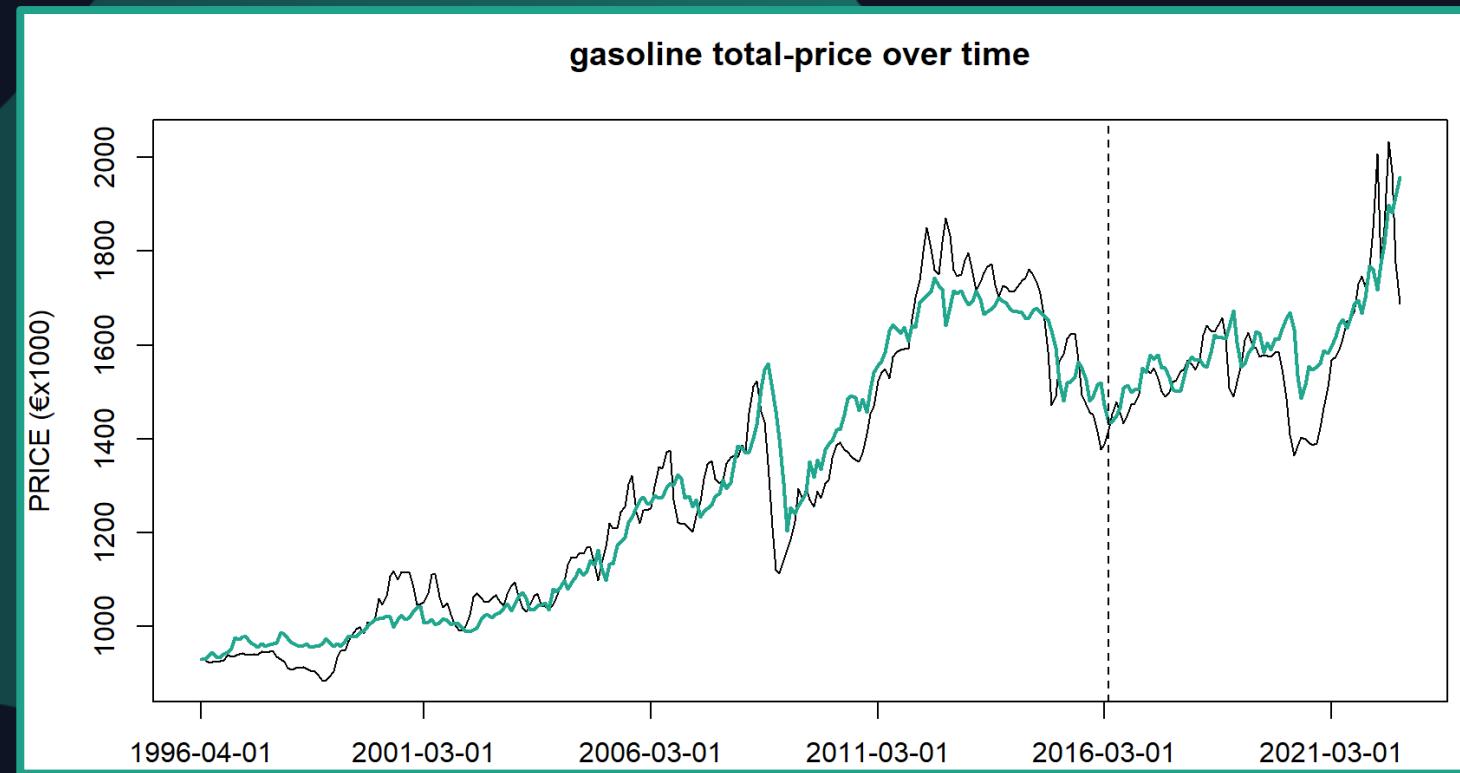
### Motivation

- Simple and interpretable
- Assume linear relationship of predictors vs response

Test RMSE = 0.101 €

### Results

- Month
- Oil's price
- Vehicle emissions
- Employment rate
- EUR-USD exchange rate



# Data Modeling

## Linear Regression

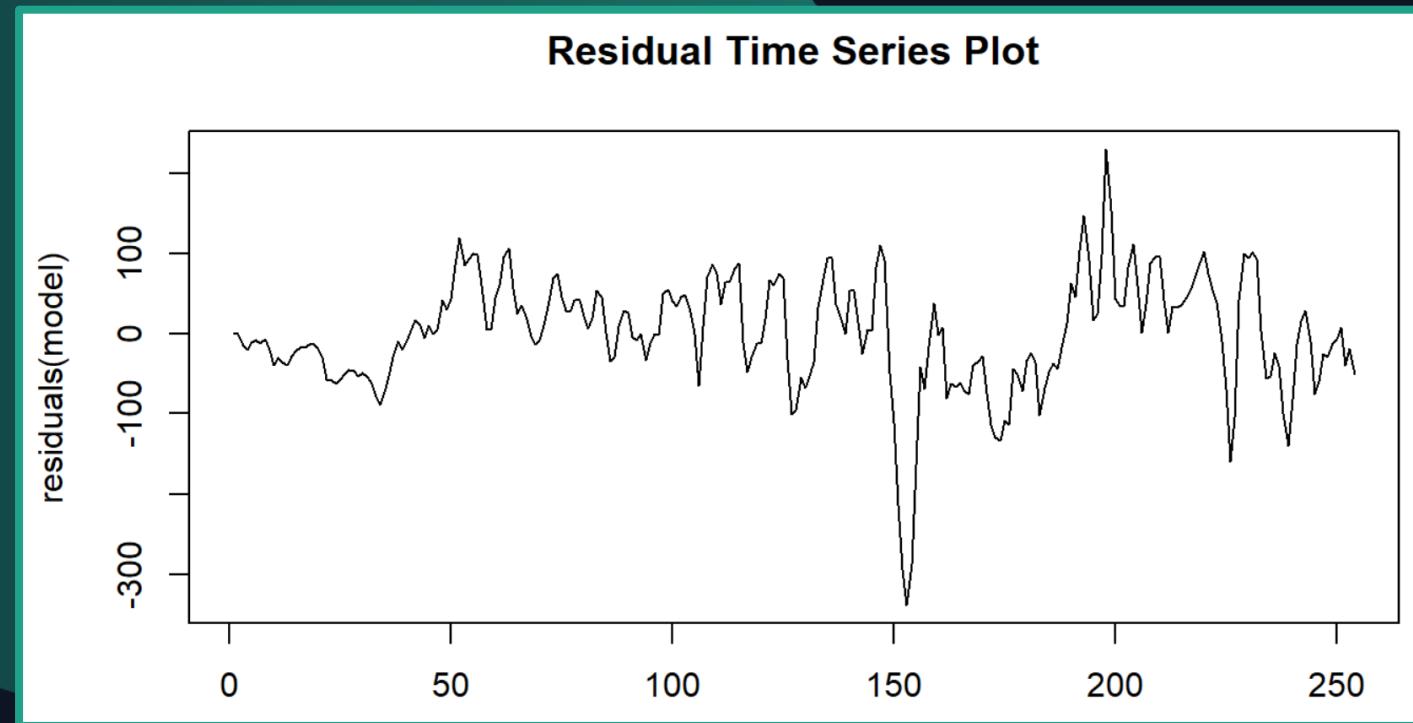
### Results

- Month
- Oil's price
- Vehicle emissions
- Employment rate
- EUR-USD exchange rate

### Motivation

- Simple and interpretable
- Assume linear relationship of predictors vs response

Test RMSE = 0.101 €



# Data Modeling

## Linear Regression

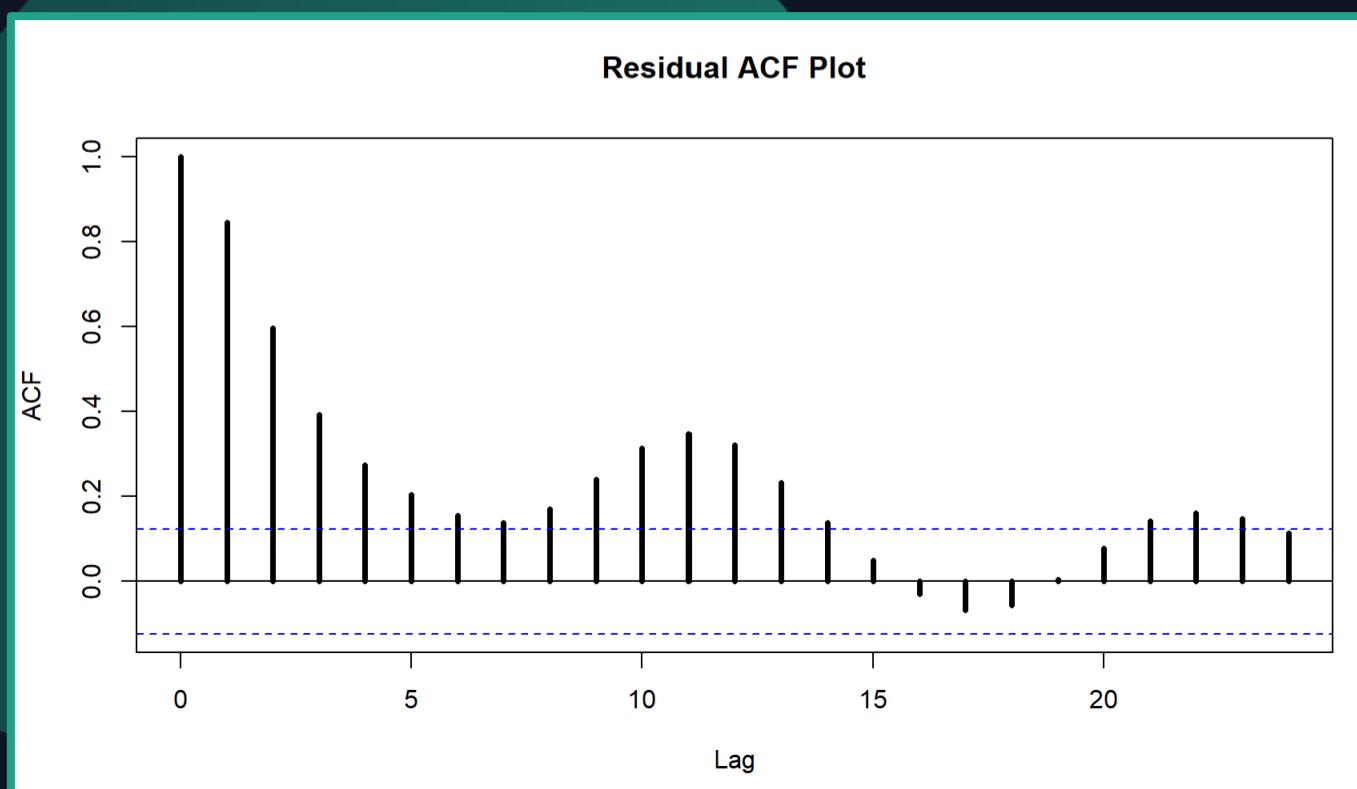
### Motivation

- Simple and interpretable
- Assume linear relationship of predictors vs response

Test RMSE = 0.101 €

### Results

- Month
- Oil's price
- Vehicle emissions
- Employment rate
- EUR-USD exchange rate



# Data Modeling

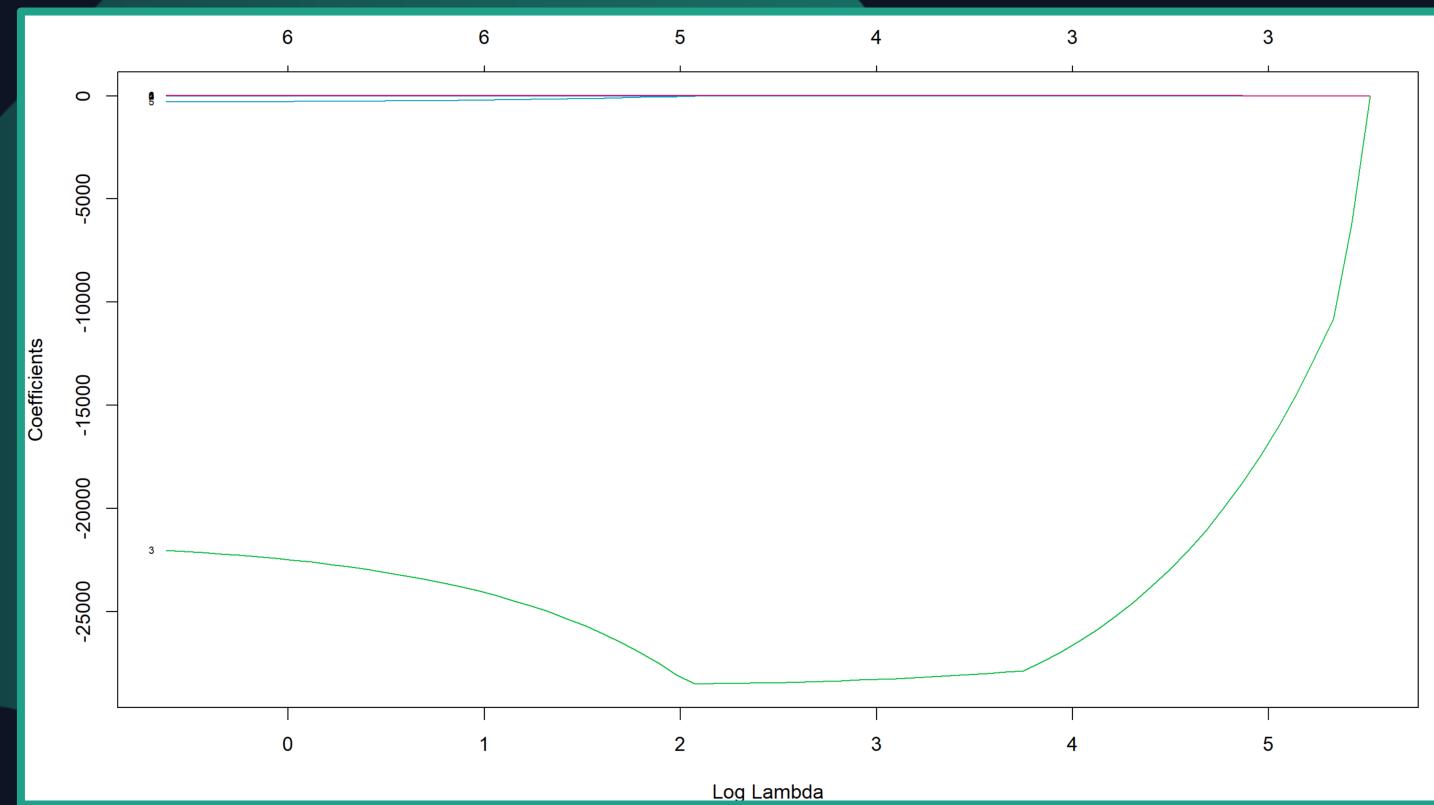
## Lasso Regression

### Motivation

- Some predictors show correlation
- Improve the linear regression results

### Results

- Time-trend
- Oil's price
- Vehicle emissions
- Employment rate



# Data

# Modeling

## Lasso Regression

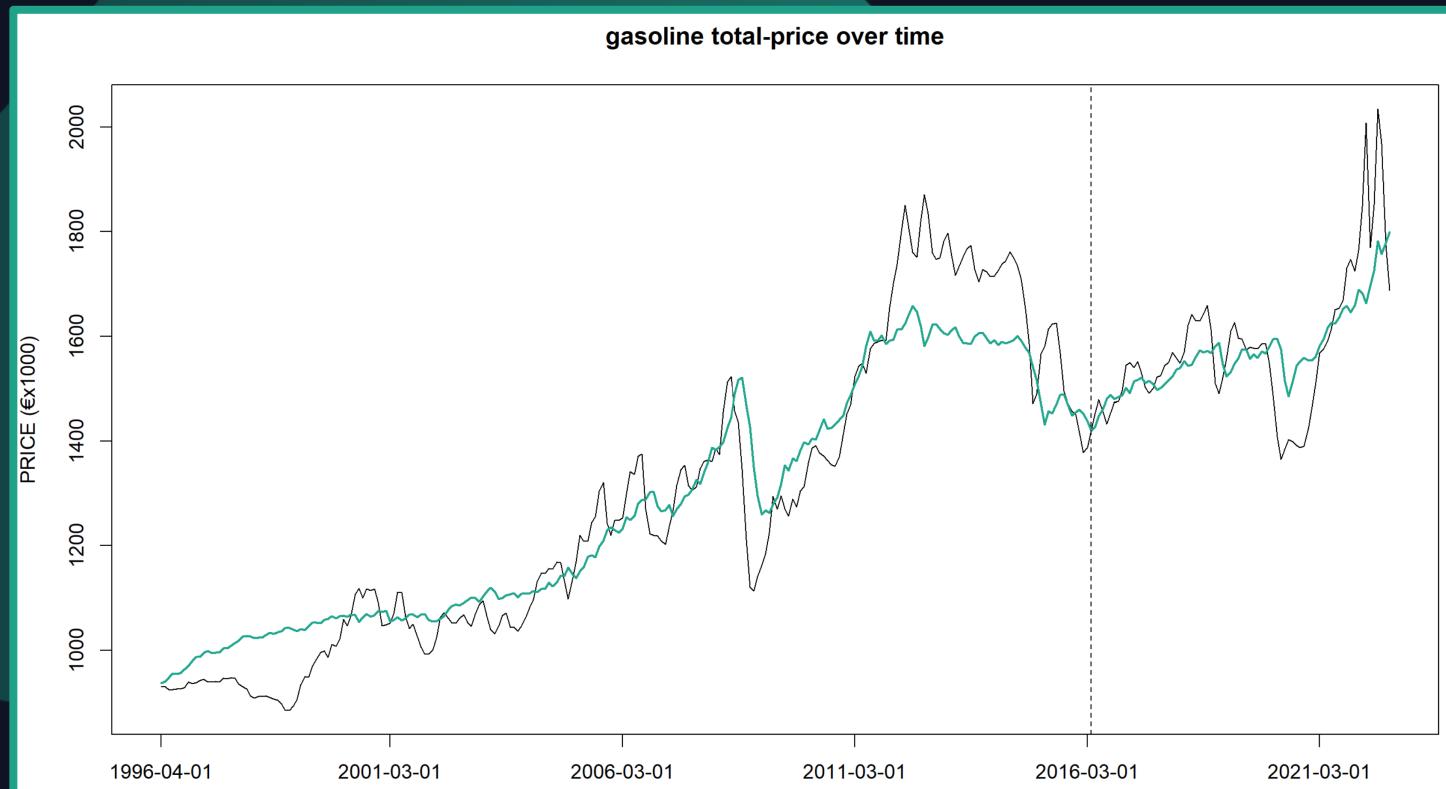
## Results

- Time-trend
- Oil's price
- Vehicle emissions
- Employment rate

## Motivation

- Some predictors show correlation
- Improve the linear regression results

Test RMSE = 0.098 €



# Data Modeling

## Lasso Regression

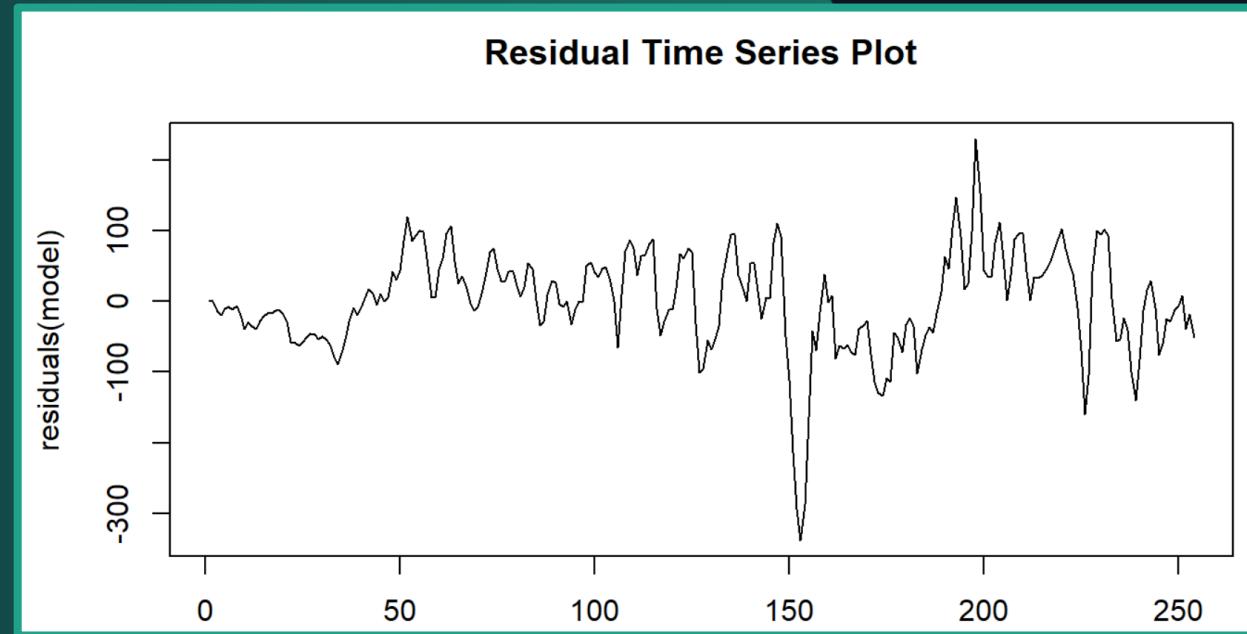
### Motivation

- Some predictors show correlation
- Improve the linear regression results

Test RMSE = 0.098 €

### Results

- Time-trend
- Oil's price
- Vehicle emissions
- Employment rate



# Data Modeling

## Lasso Regression

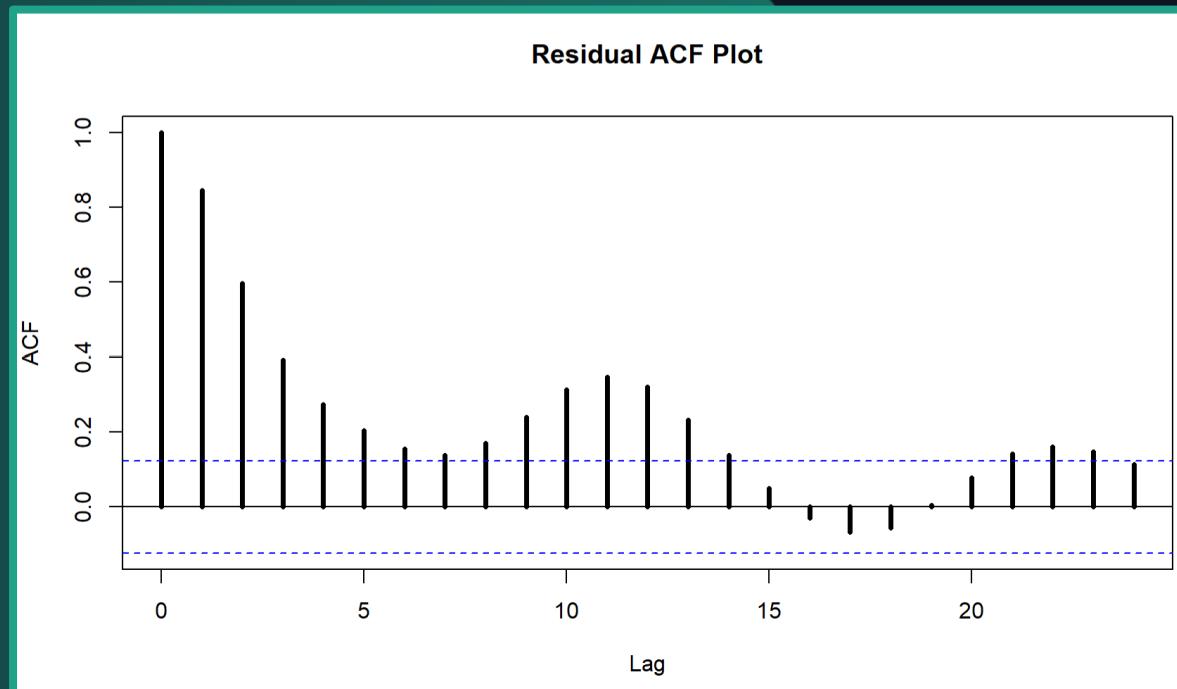
### Motivation

- Some predictors show correlation
- Improve the linear regression results

Test RMSE = 0.098 €

### Results

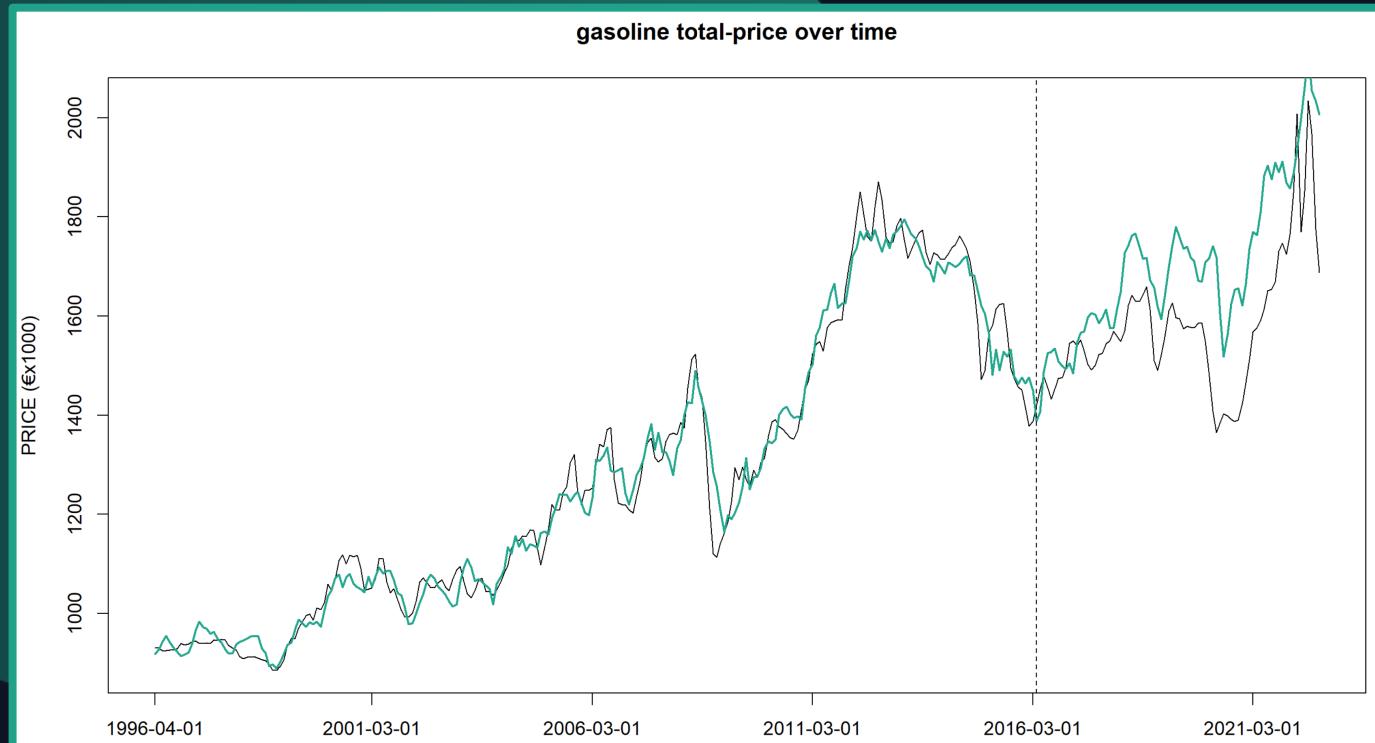
- Time-trend
- Oil's price
- Vehicle emissions
- Employment rate



# Data Modeling

## Generalized Additive Model (smooth | spline)

Test RMSE = 0.169 €



## Motivation

- The relationship may not be linear

## Results

- s(Month)
- s(Oil's price)
- s(Vehicle emissions)
- Employment rate
- s(EUR-USD exchange rate)

# Data Modeling

## Generalized Additive Model (smooth | spline)

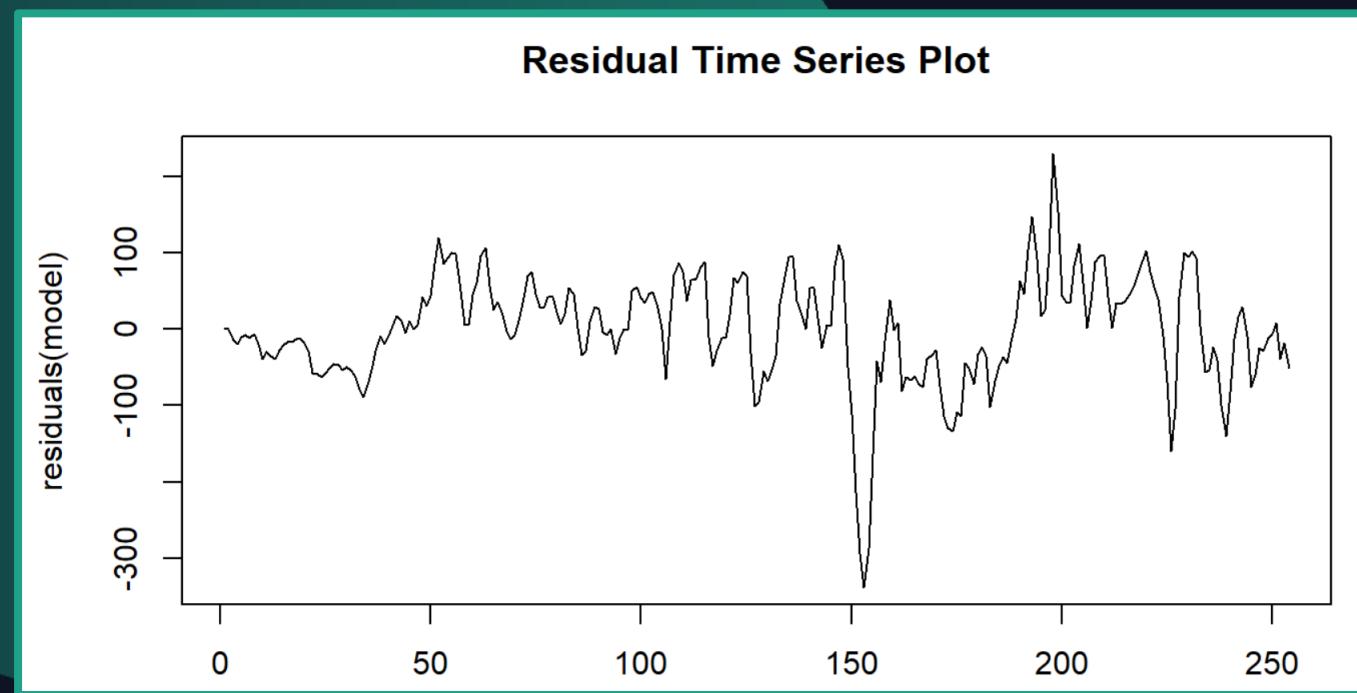
### Motivation

- The relationship may not be linear

Test RMSE = 0.169 €

### Results

- s(Month)
- s(Oil's price)
- s(Vehicle emissions)
- Employment rate
- s(EUR-USD exchange rate)



# Data Modeling

## Generalized Additive Model (smooth | spline)

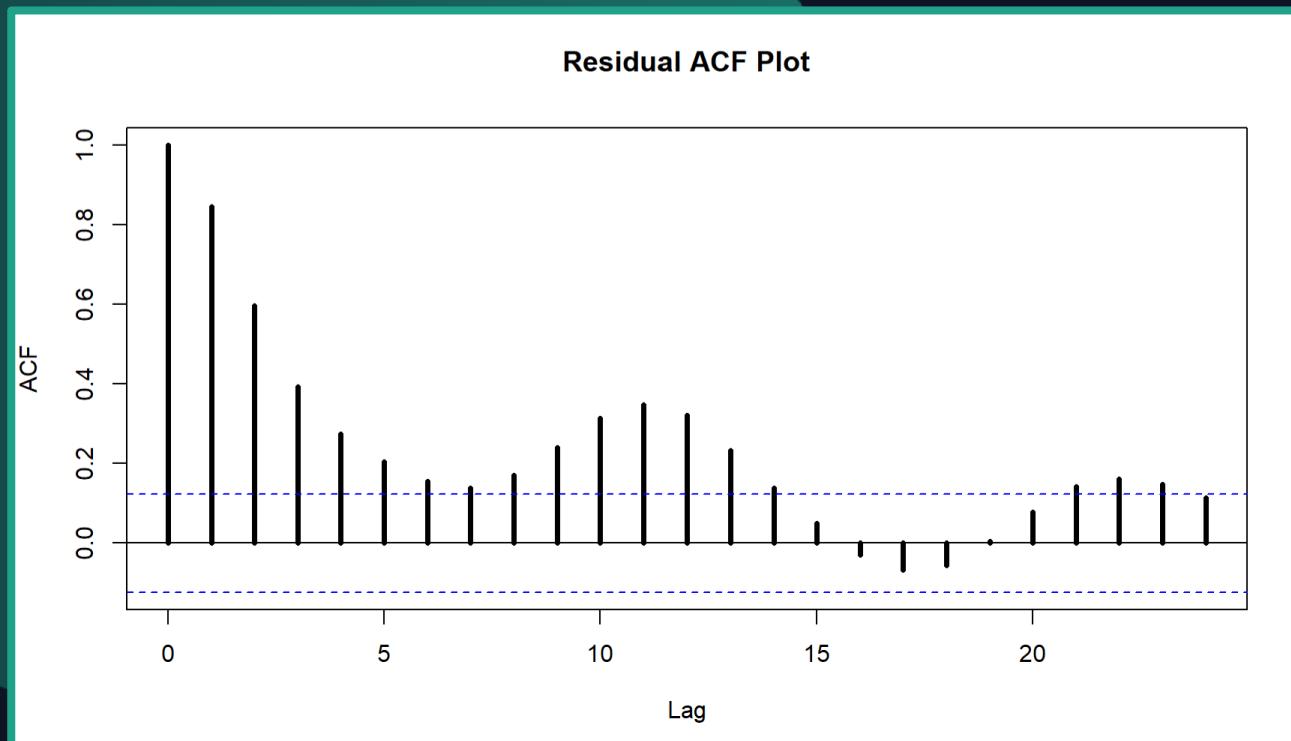
### Motivation

- The relationships may not be linear

Test RMSE = 0.169 €

### Results

- s(Month)
- s(Oil's price)
- s(Vehicle emissions)
- Employment rate
- s(EUR-USD exchange rate)



# Data Modeling

## Generalized Additive Model (smooth | loess)

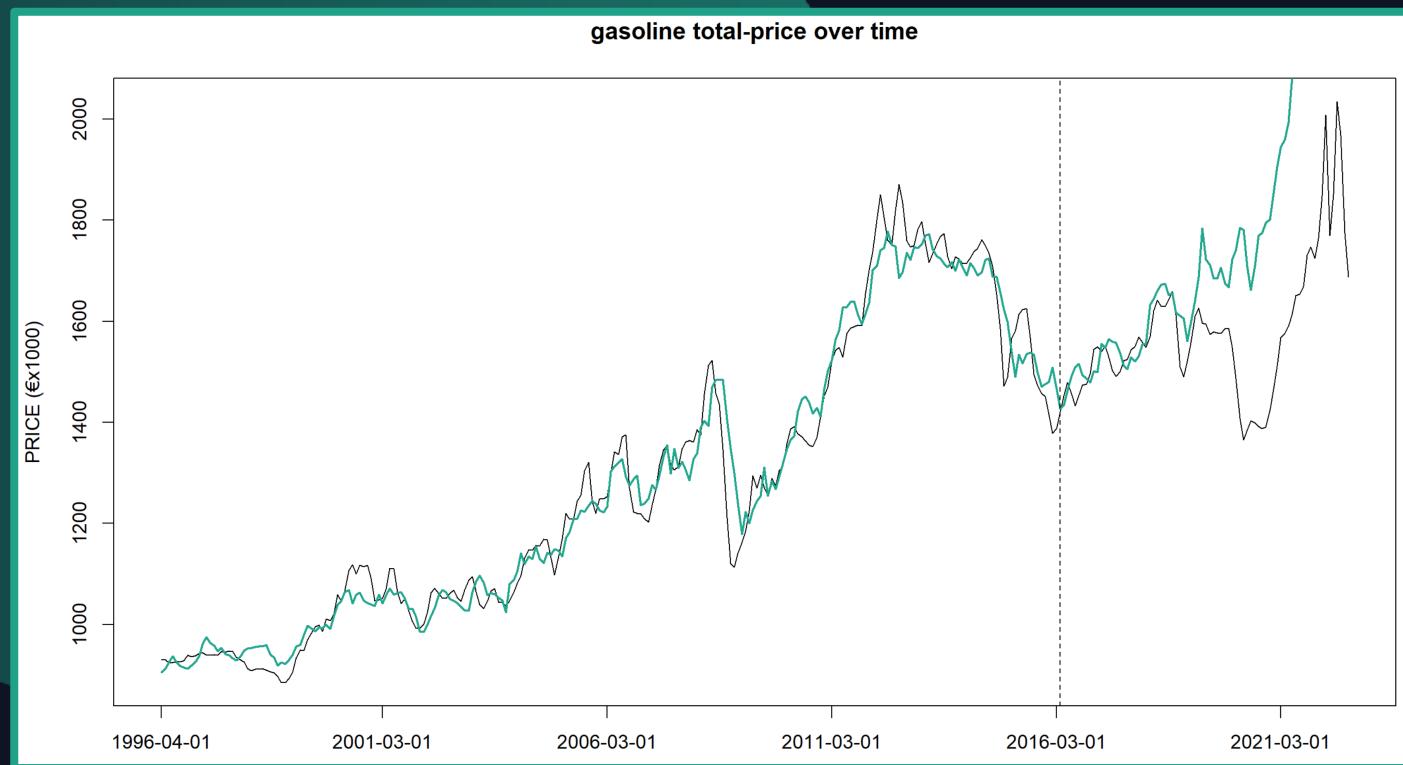
### Motivation

- The relationships may not be linear

Test RMSE = 0.429€

### Results

- s(Month)
- s(Oil's price)
- s(Vehicle emissions)
- Employment rate
- s(EUR-USD exchange rate)



# Data Modeling

## Generalized Additive Model (smooth | loess)

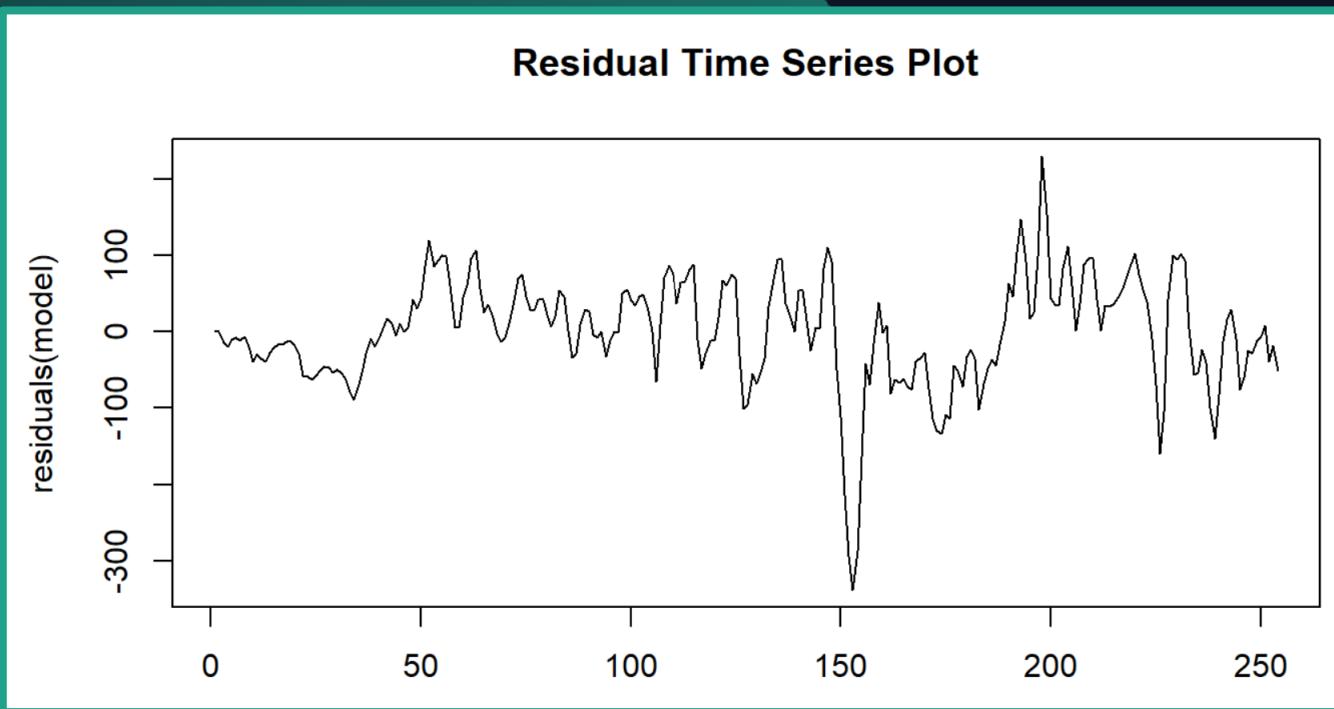
### Motivation

- The relationships may not be linear

Test RMSE = 0.429€

### Results

- s(Month)
- s(Oil's price)
- s(Vehicle emissions)
- Employment rate
- s(EUR-USD exchange rate)



# Data Modeling

## Generalized Additive Model (smooth | loess)

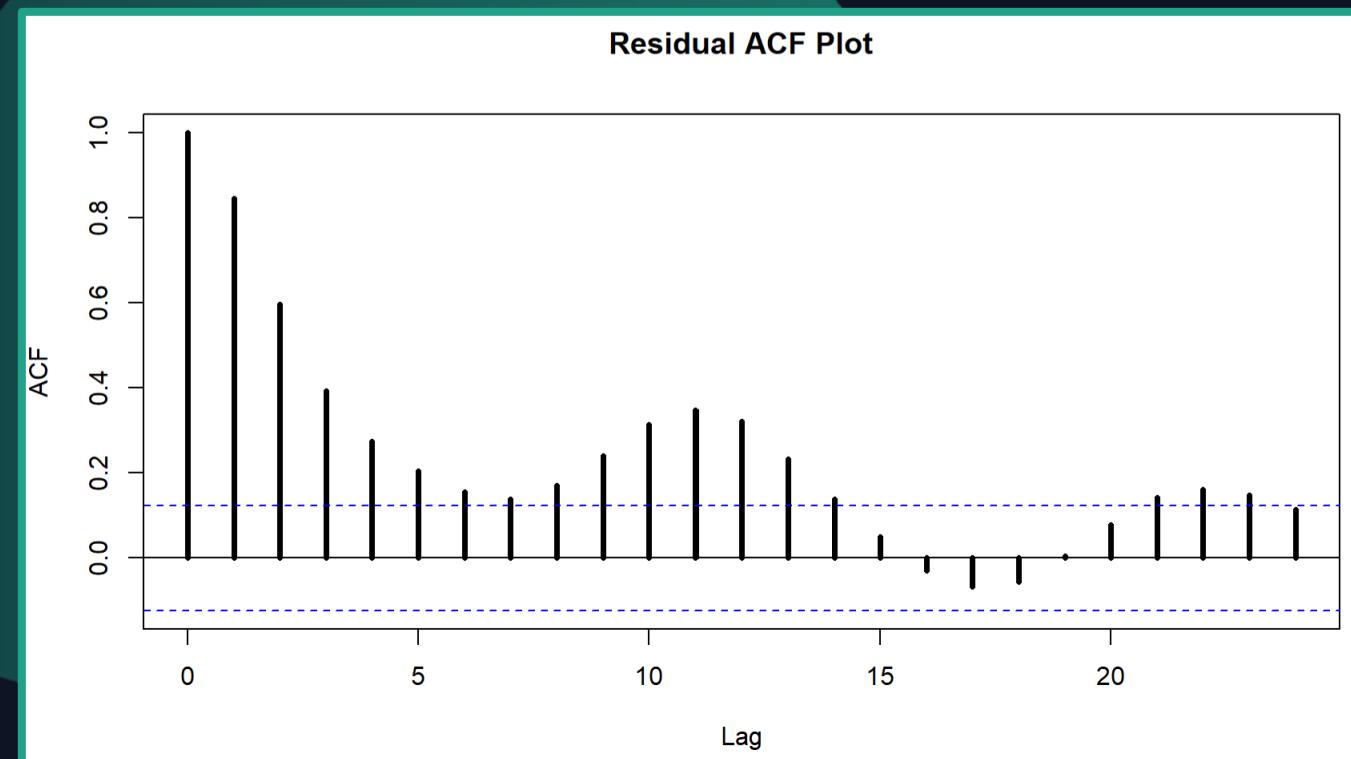
### Motivation

- The relationships may not be linear

Test RMSE = 0.429€

### Results

- s(Month)
- s(Oil's price)
- s(Vehicle emissions)
- Employment rate
- s(EUR-USD exchange rate)



# Data Modeling

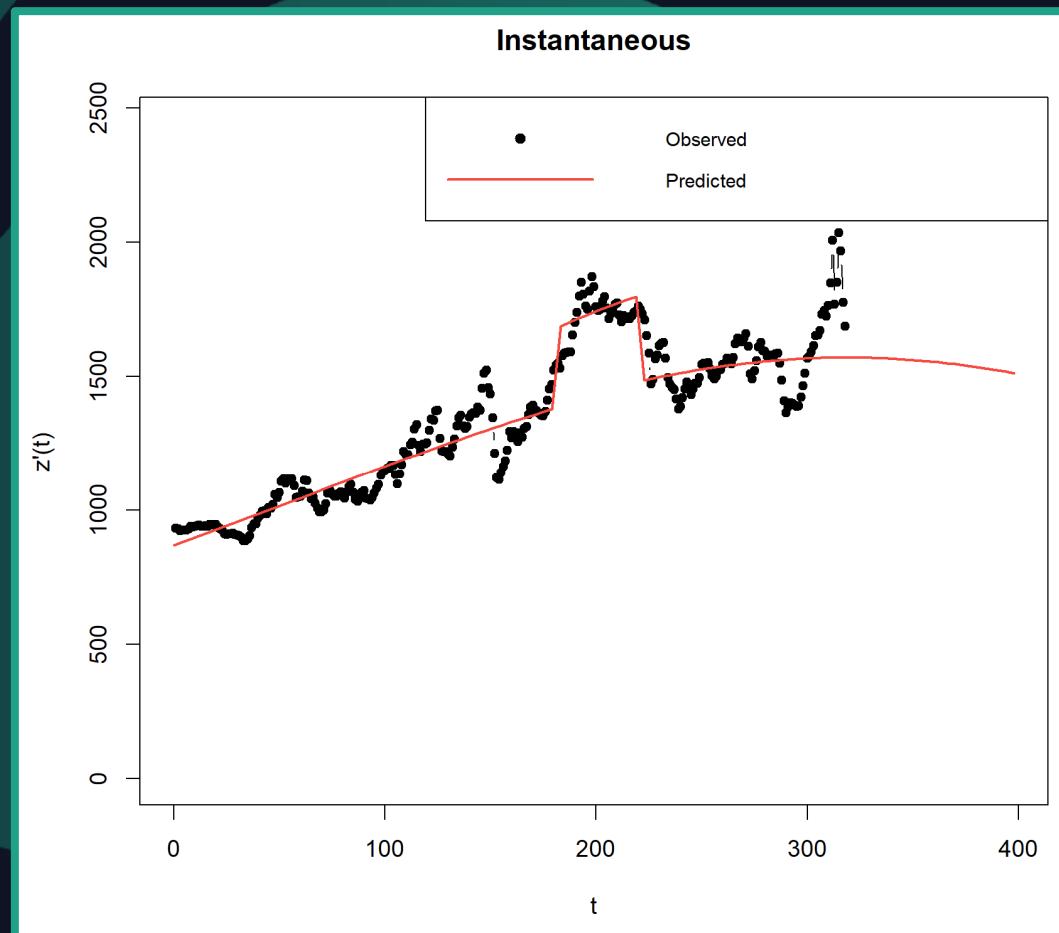
## Generalized Bass Model

### Motivation

- Check if there is shock after the 2008 economic crisis

### Results

- Gasoline's price



# Data Modeling

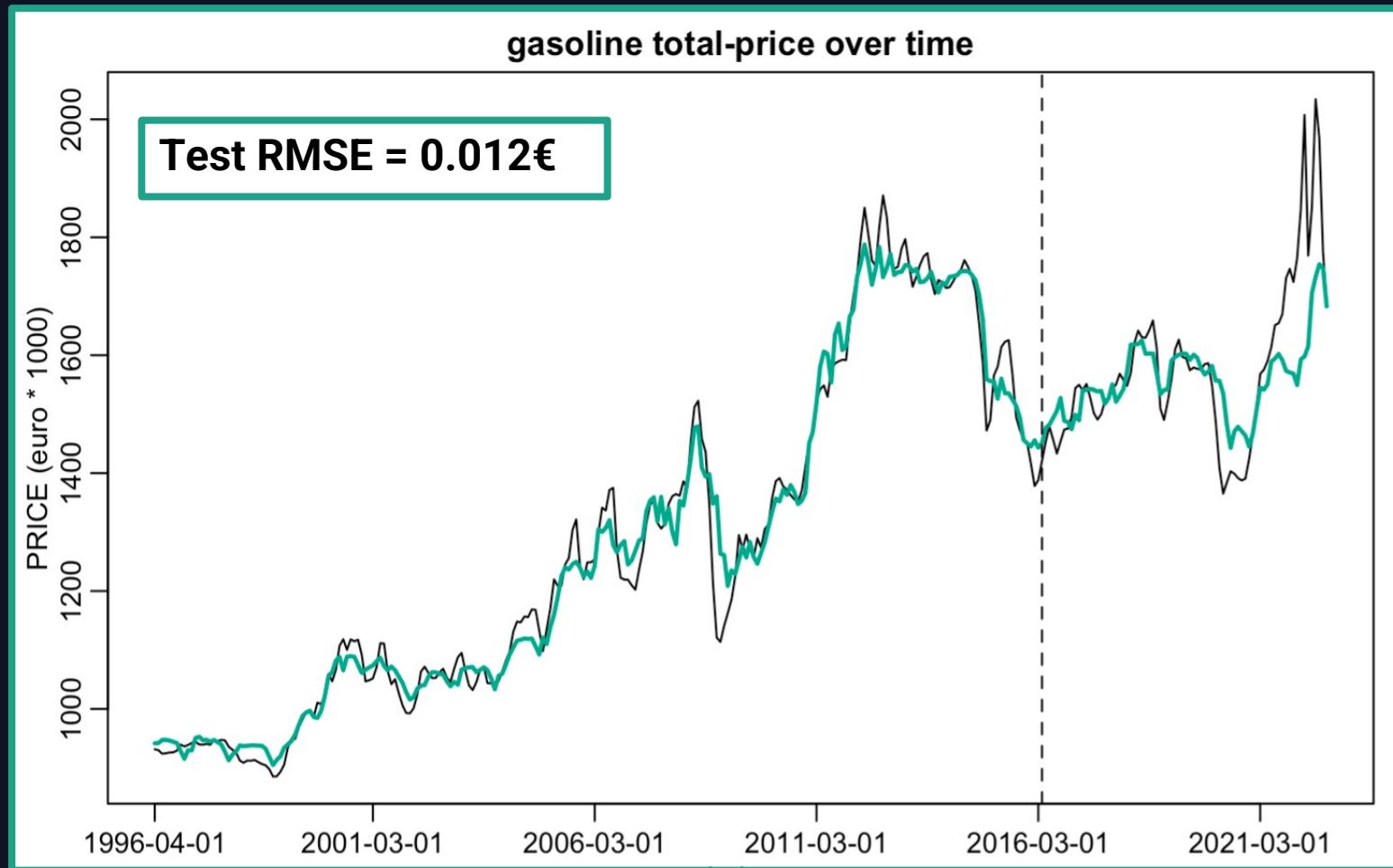
## Gradient Boosting

### Motivation

- Combine multiple weaker models
- Try to learn the seasonal component thanks to the predictors
- Very different approach compared to the previous ones

# Data Modeling

## Gradient Boosting Fit and Forecasts



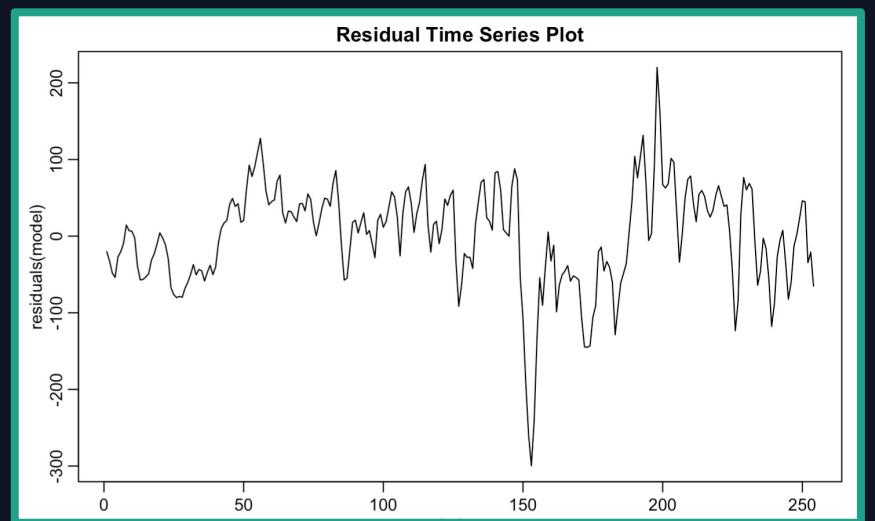
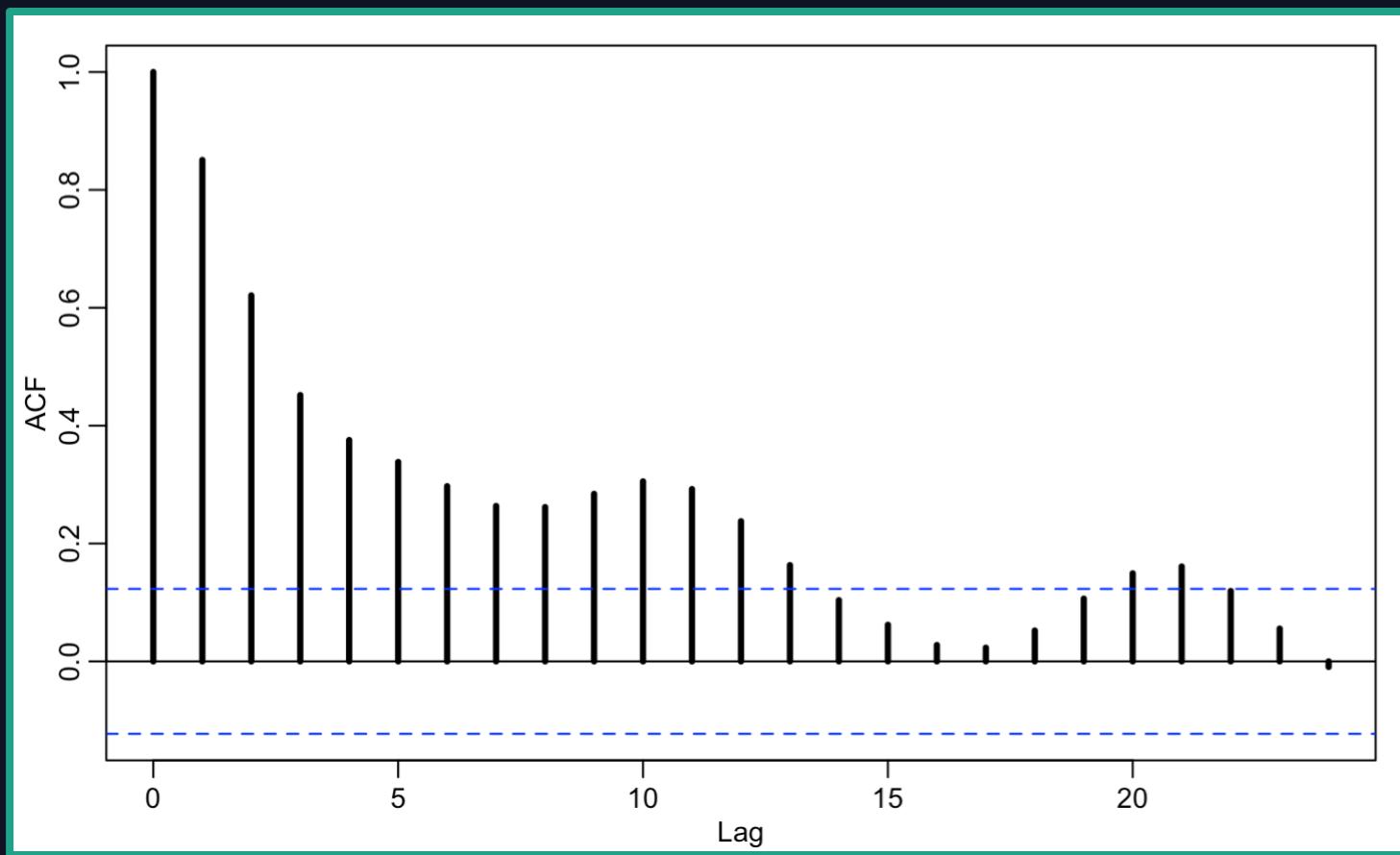
### Parameters:

- 1500 trees
- No cross validation
- Leaves with at least 2 elements

# Data Modeling

## Gradient Boosting

### Analysis of residuals

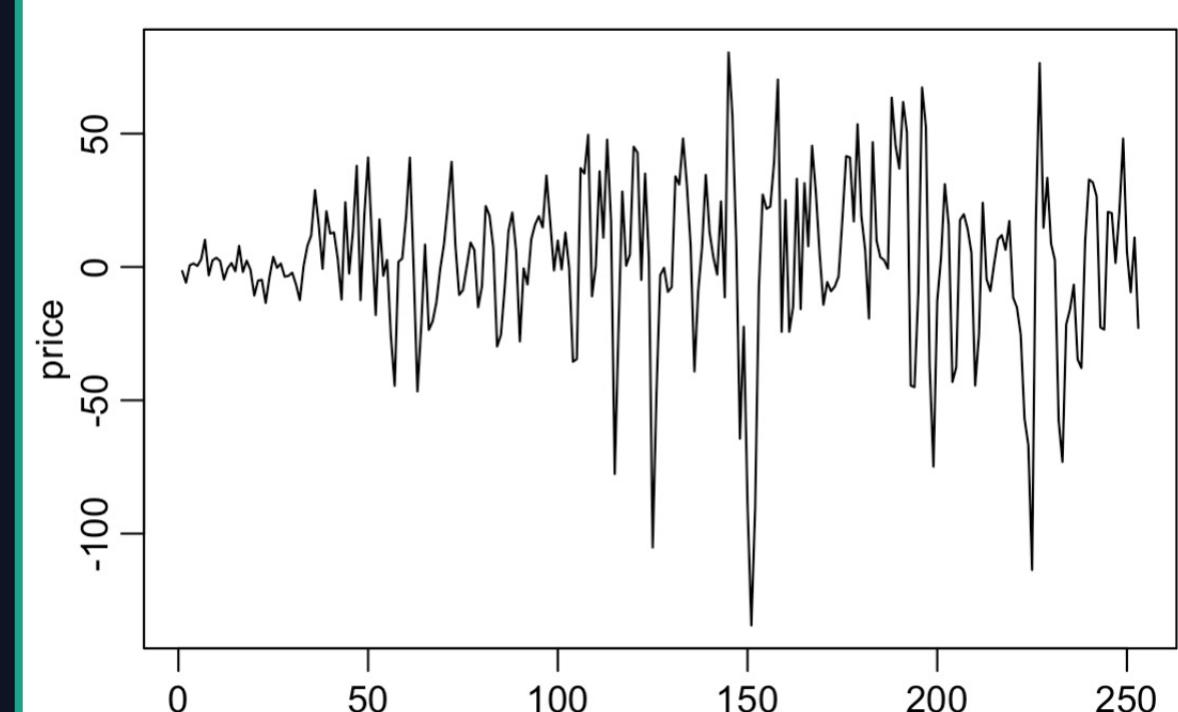


# Data Modeling ARIMA

## Motivation

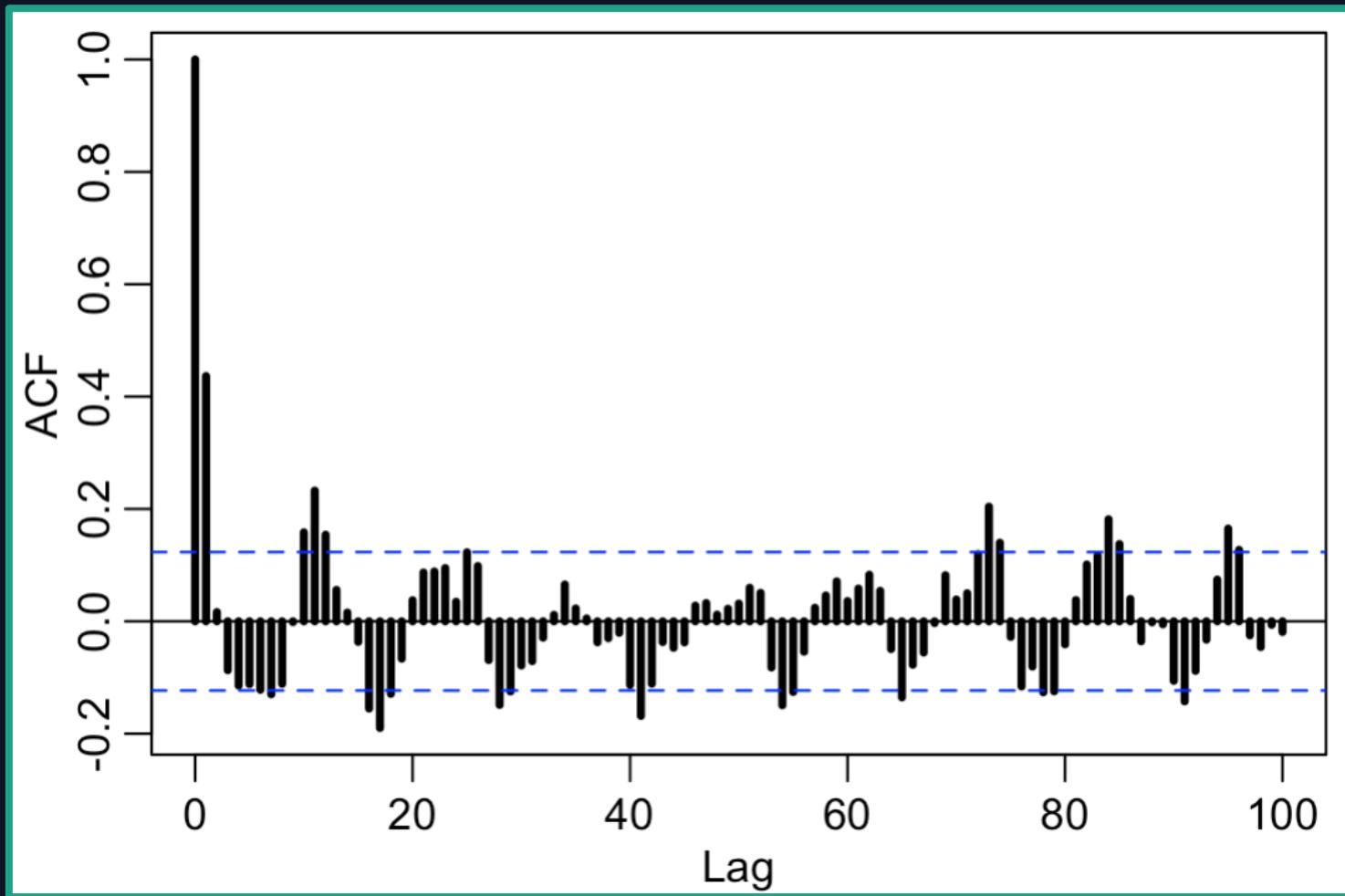
- Is it possible to better model the time-dependent component?

### Differentiated Series



# Data Modeling ARIMA

ACF of differentiated Series



*SARIMA(2, 1, 2)(0, 1, 1)<sub>12</sub>*

# Data Modeling ARIMA Best model

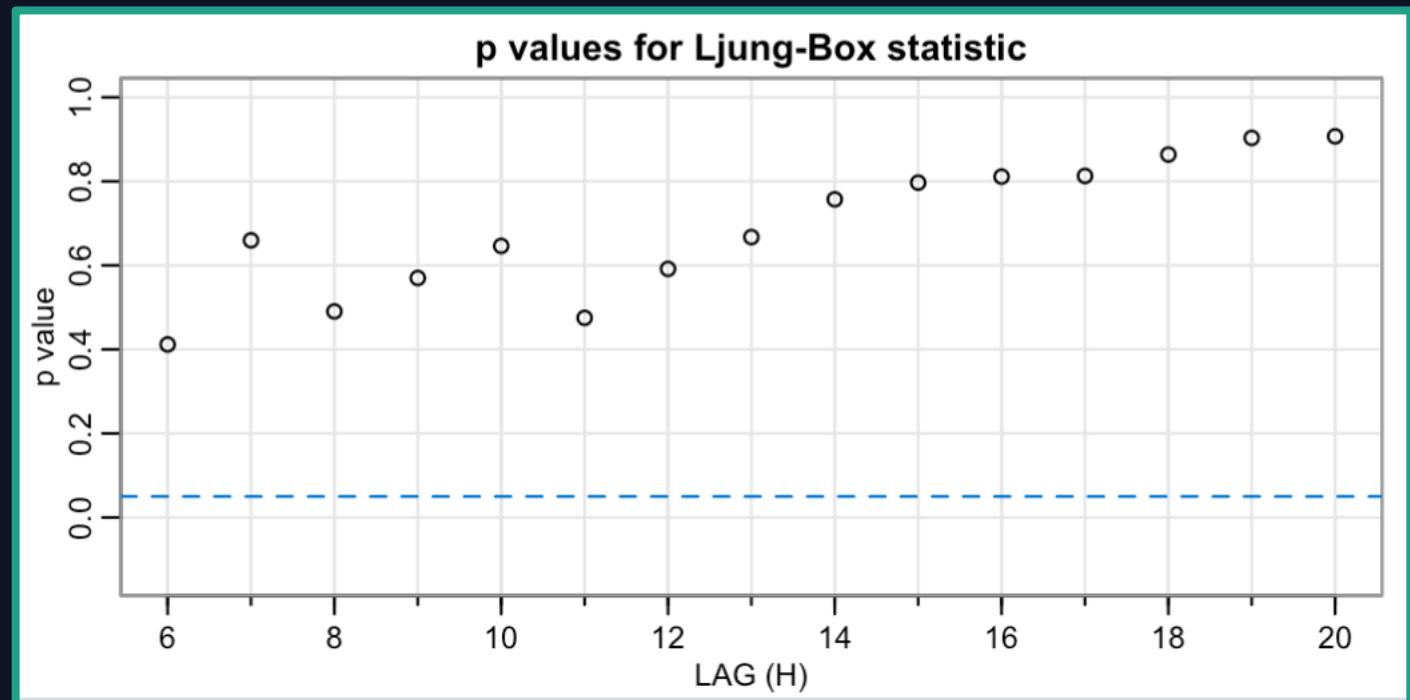
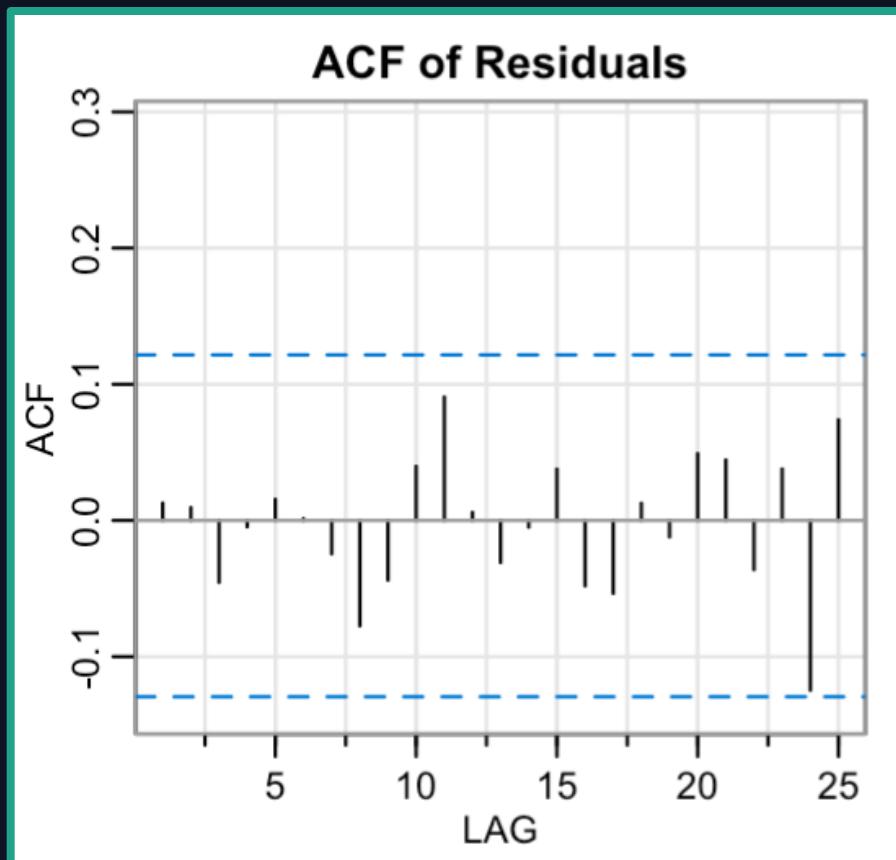
- The series needs to be differentiated
- Observations of the previous two months are useful in the autoregressive part as well as in the moving average one
- There is not seasonality

# Data Modeling

ARIMA

Best model

$SARIMA(2, 1, 2)(0, 1, 1)_{12}$



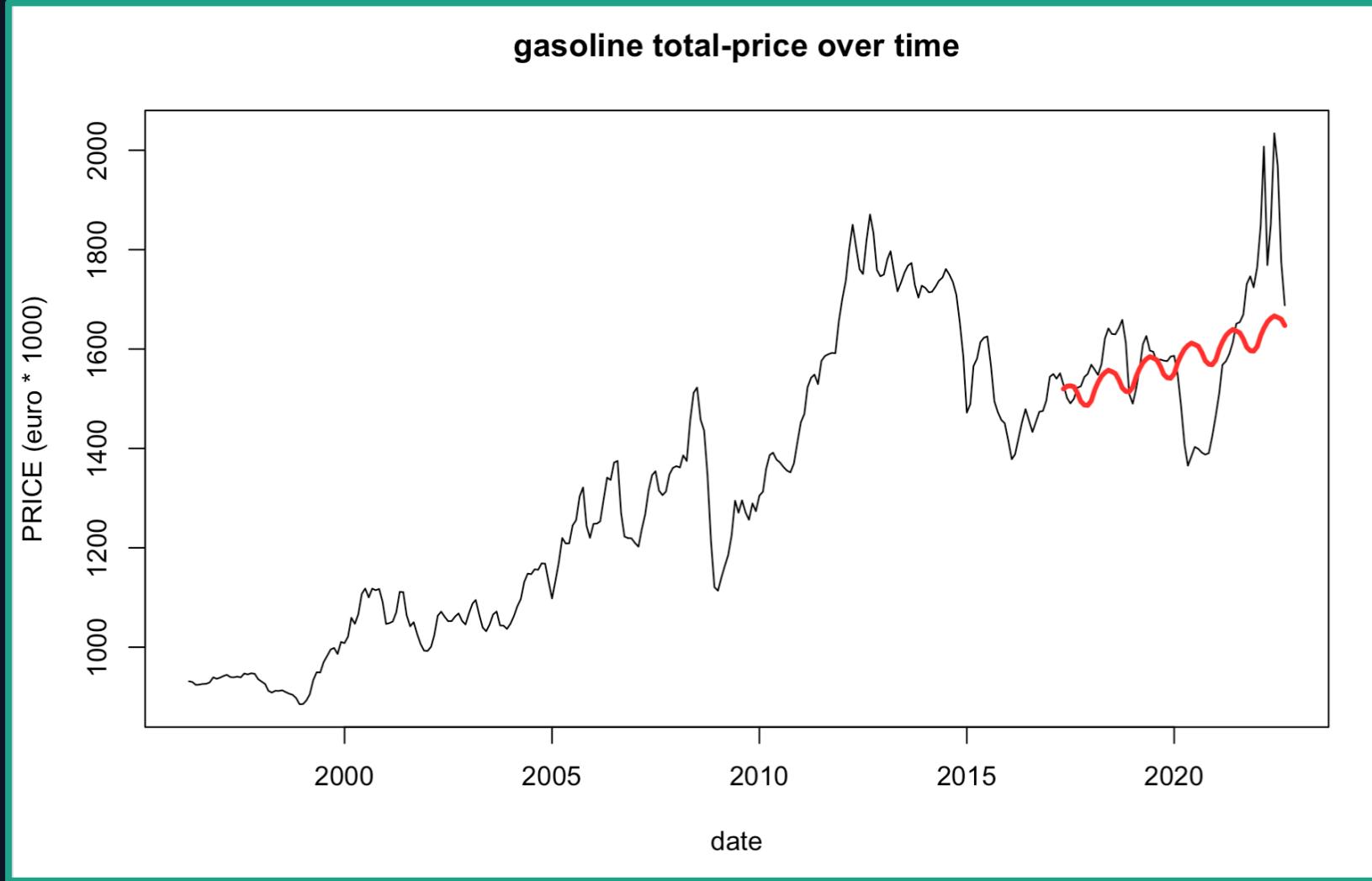
# Data

# Modeling

## ARIMA

## Best model

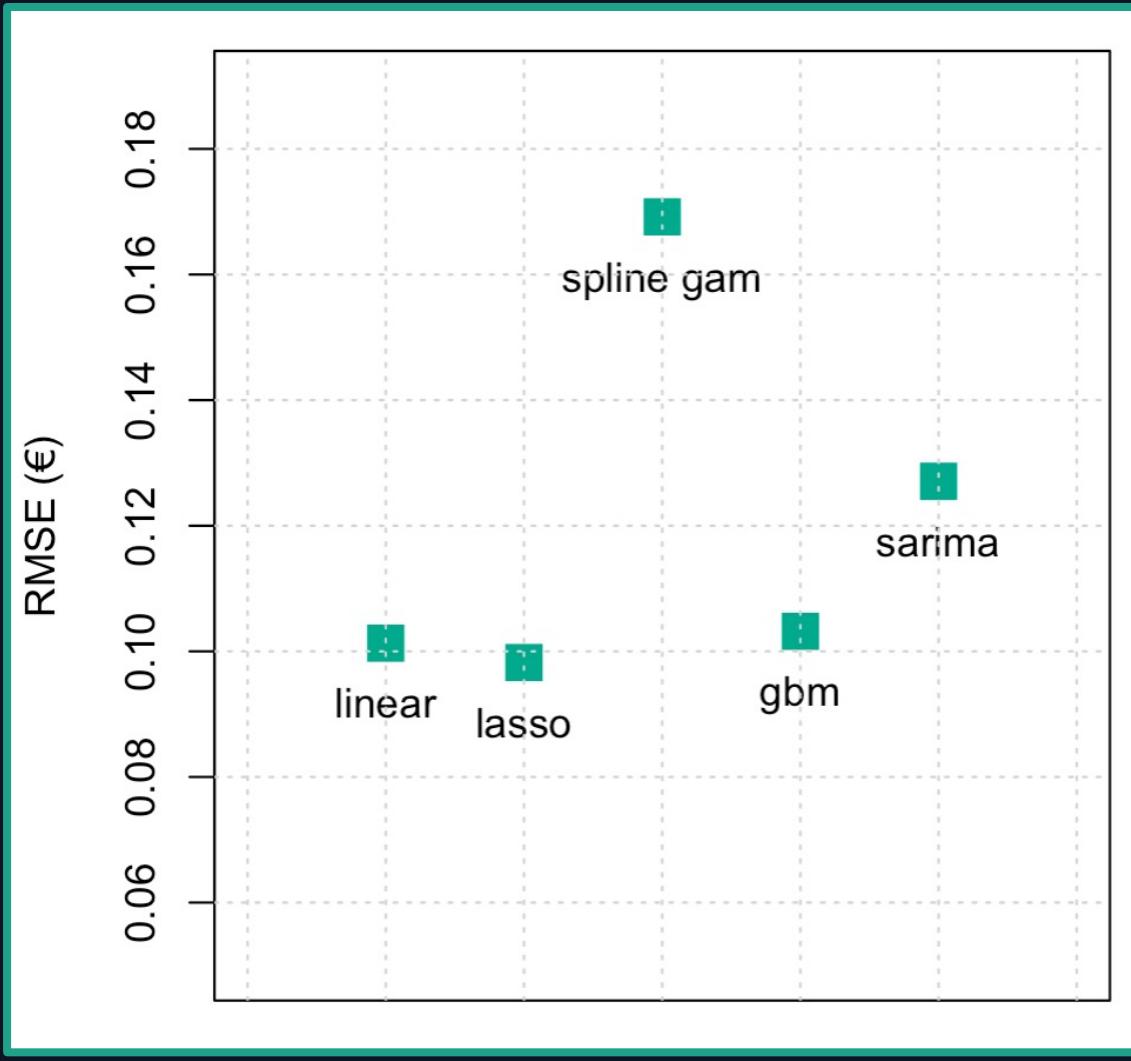
Test RMSE = 0.127 €



# Data Modeling

Find the best  
Model

RMSE  
(€)

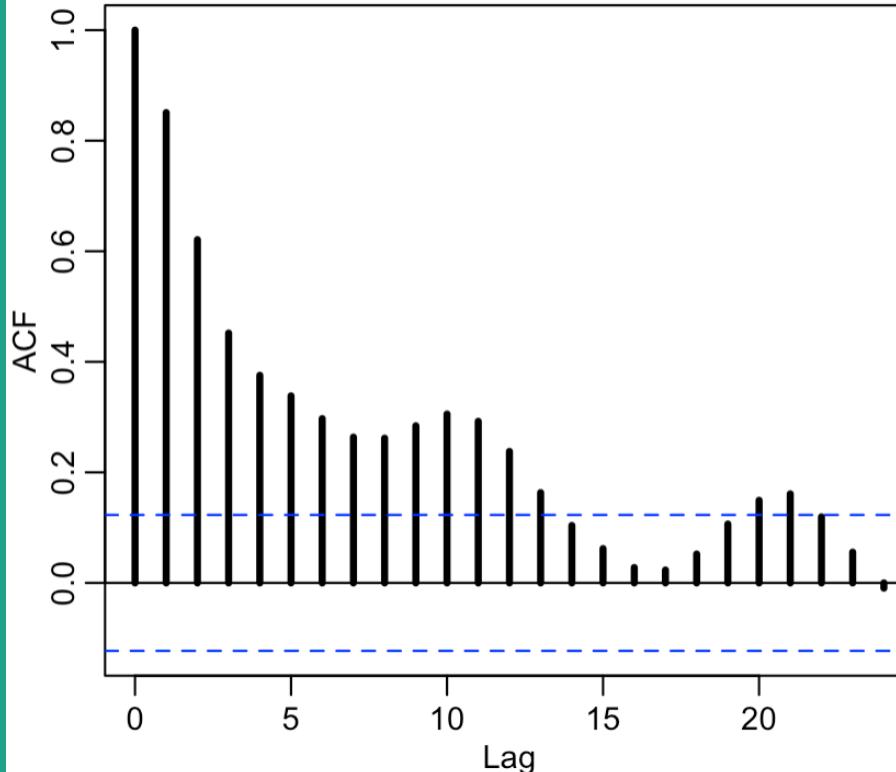


# Data Modeling Improving

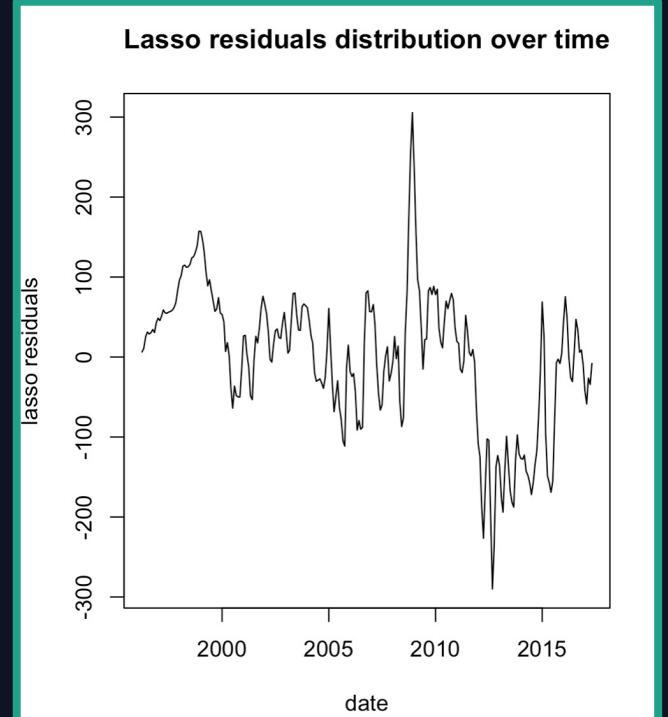
The best model in terms of RMSE  
is the lasso regression

Is it enough?

ACF on the LASSO residuals



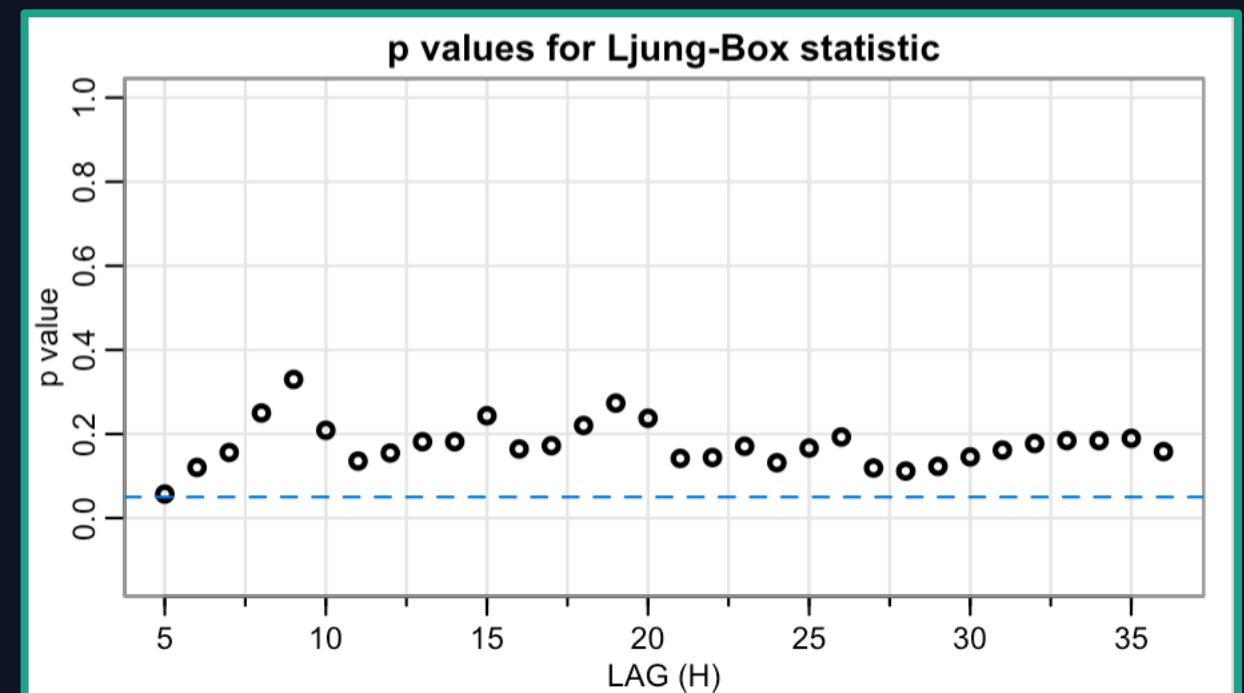
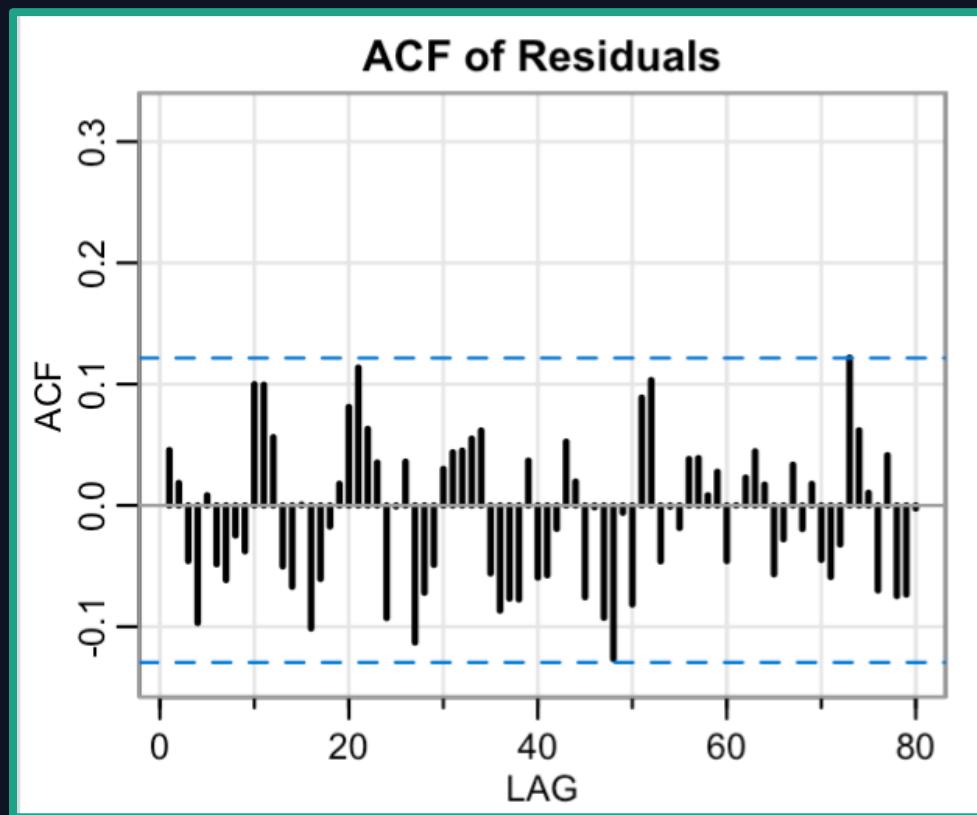
Lasso residuals distribution over time



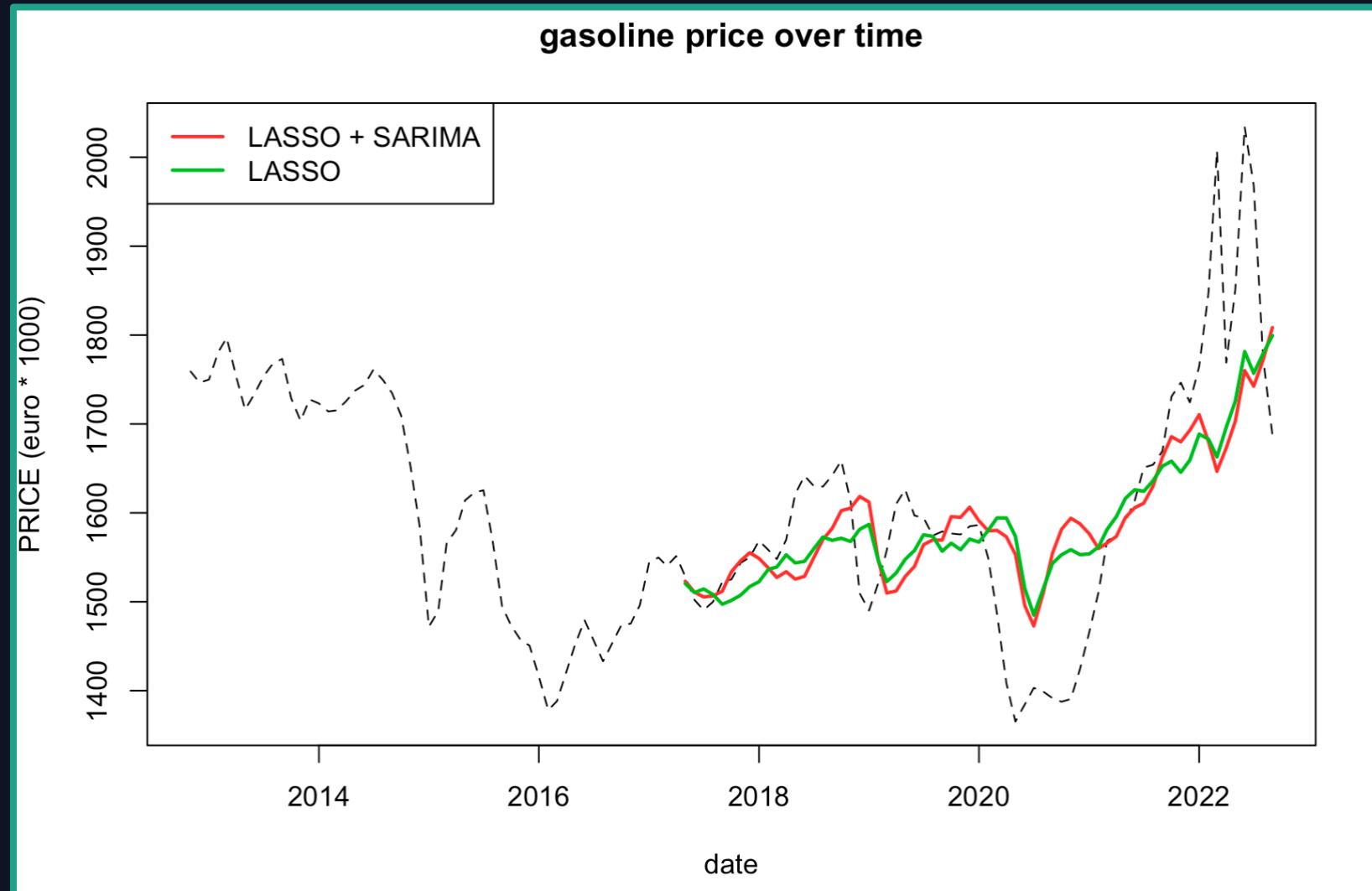
# Data Modeling Improving

The residuals can be modeled by  
a SARIMA MODEL

SARIMA(2, 1, 2)(0, 1, 1)<sub>12</sub>



# Data Modeling Improving



LASSO RMSE = 0.098 €

LASSO + SARIMA  
RMSE = 0.103 €

# Conclusions

- The best model in terms of RMSE is the **LASSO** regression
- On average, the error is of **9.8 cents** (~5%)
- The variables that are significant are: the **trend**, the **vehicles' emission efficiency**, the **employment rate** and the **oil price**
- The SARIMA on the LASSO gives worse results -> the autoregressive information (the **previous observations**) **do not help** to forecast the price

# Conclusions

The variables that are significant are: the **trend**, the **vehicles' emission efficiency**, the **employment rate** and the **oil price**

Trend: + 1 month → + 0.053 cents of €

Emission efficiency: + 1 CO2 kg / km → - 2.8 cents of €

Employment rate: + 1-point percentage → - 11 cent of €

Oil price: + 1\$ → + 29 cents of €

# Future work

A significant percentage of the price is given by the **taxes**. We have not considered this information in our analysis.

**Qualitative** analysis over the **future taxes' value** can be added to our forecasts to better **adapt to real scenarios**.

L'opposizione critica la premier per gli interventi promessi prima del voto. La replica: necessarie più entrate

## Benzina, il caso delle accise

Meloni: «Tagli impossibili ora, finanziati altri aiuti». Bolzano è la più cara

## Meloni, scivolone sui carburanti

*La premier spiega: «Scelte di giustizia sociale che rivendico. Mai promesso a queste elezioni il taglio delle accise». Misiani la smentisce citando il programma elettorale di FdI, opposizioni scatenate. E lei rettifica: sì, ma solo con più entrate. I gestori ipotizzano lo sciopero*

## Scontro sul taglio delle accise

Accise, il piano degli alleati per piegare la premier e la strategia di Meloni per spostare l'attenzione | La bugia sulla benzina