technische universität dortmund

Detecting change-points and trends in the petrol consumption and petrol prices

Abdul Muqsit Farooqi

September 2, 2025

Outline

Basic Problem

- Two fuel logbooks: *grau* and *karriert*
- Five variables each: date, odometer, price (paid), currency, liters filled
- One extra variable in *karriert*: consumption (liters / 100 km)
- Goal: Analyze time series for consumption and petrol prices (€/liter)
- Detect change-points and trends in both series

Further Questions

- Is fuel consumption affected by car changes?
- Is the Euro introduction reflected in petrol prices?

Data Pre-Processing

Category	Action	Example(s)	Count
Date Corrections	Fixed incorrect / missing dates	2001-06-05 changed to 2014-05-26	7
Currency Standardization	Converted all prices to EUR using historical rates	DM, ATS, CZK, ITL, Zloty, Dinar	47
Price Corrections	Fixed typos or inconsistent values	Dinar price changed from 12.6 to 12600	2
Missing Data Handling	Interpolated or removed	Median of 1987-06-13 and 1987-06-09 is 1987-06-11	10

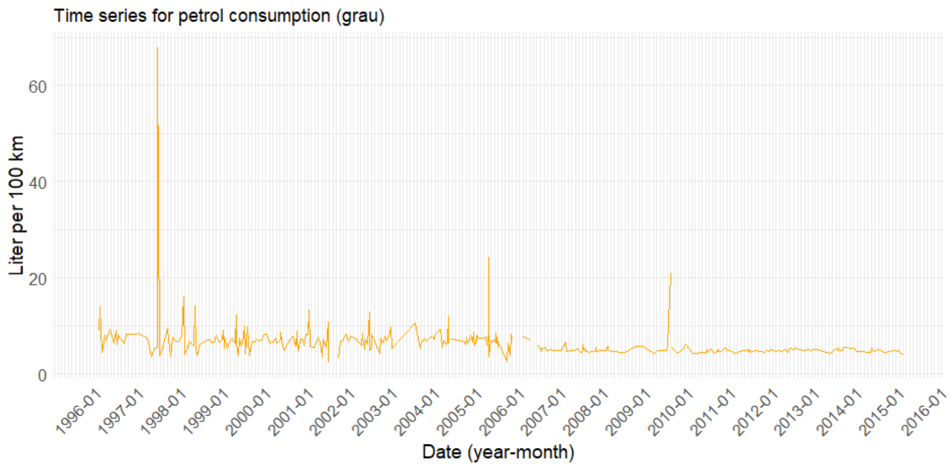

Figure 1: Time series of fuel consumption for *grau* in liters per 100 km.

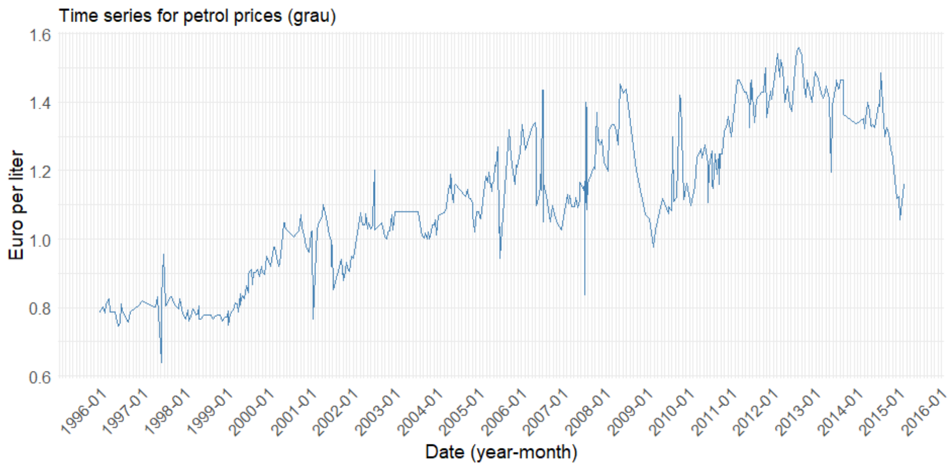

Figure 2: Time series of petrol prices for *grau* in Euro per liter.

Time series

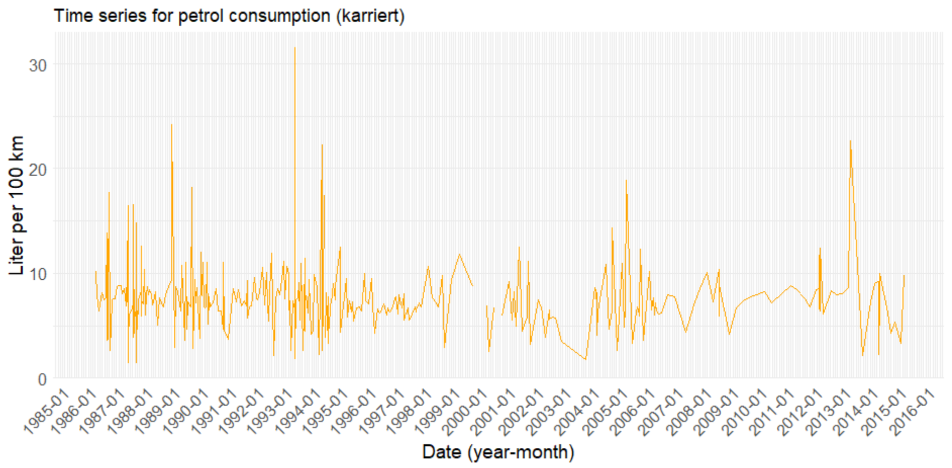

Figure 3: Time series of petrol consumption for *karriert* in liter per 100 km.

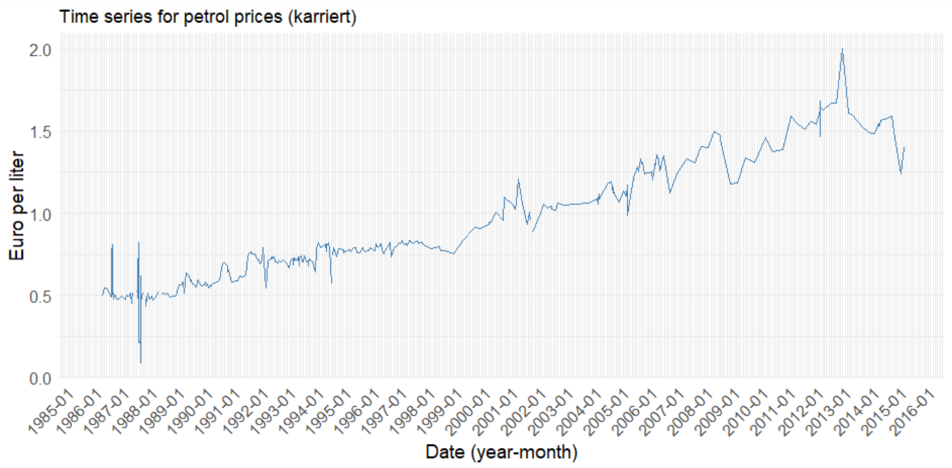

Figure 4: Time series of petrol prices for *karriert* in Euro per liter.

Changepoint Detection Method: PELT

PELT (Pruned Exact Linear Time): Identify structural changes (changepoints) in the time series.

- Detects multiple changepoints by minimizing a penalized cost function:

$$\min_{\{\tau_1,\dots,\tau_m\}} \left\{ \sum_{i=1}^{m+1} \mathcal{C}(y_{\tau_i - 1:\tau_i}) + \beta m \right\}$$

  where:

  - τ_i: changepoints (unknown time indices where the statistical properties change)
  - $\mathcal{C}(\cdot)$: cost function for each segment (e.g., sum of squared errors)
  - β: linear penalty to avoid overfitting (controls number of changepoints)
- Uses pruning to reduce computational cost.

Changepoint Detection Method: Binary Segmentation (BinSeg)

Binary Segmentation: fast and simple method to detect multiple changepoints in time series data. It works recursively:

- Start with the full data interval $[1, n]$
- $\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{(\tau+1):n}) + \beta < \mathcal{C}(y_{1:n}) \longrightarrow$ changepoint τ found
- If a changepoint is found, split the series into $[1, \tau]$ and $[\tau + 1, n]$ and repeat
- Stop when no further significant changes are detected

Computationally efficient but less accurate with smaller changes.

Changepoint Detection Results

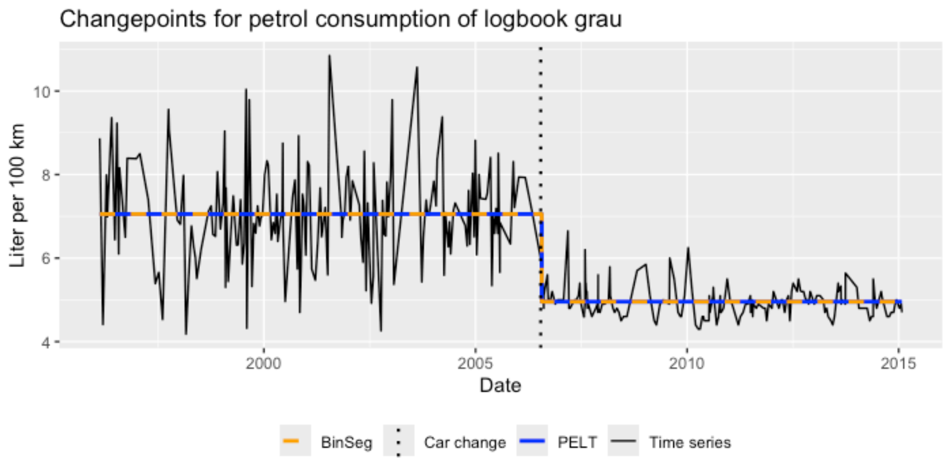

Figure 5: Changepoints for petrol consumption of logbook grau

Changepoint Detection Results

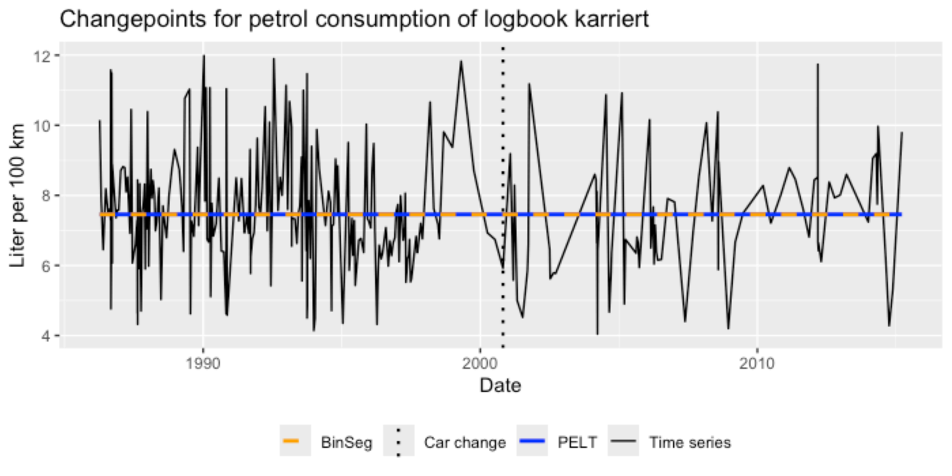

Figure 6: Changepoints for petrol consumption of logbook karriert

Changepoint Detection Results

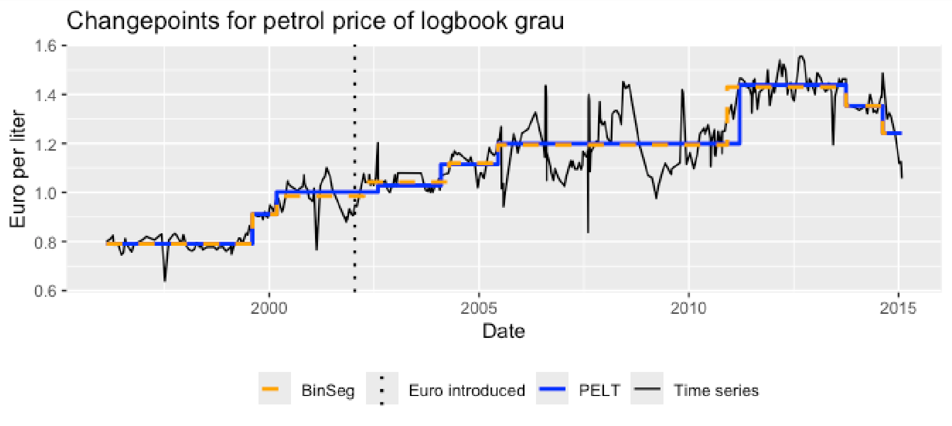

Figure 7: Changepoints for petrol price of logbook grau

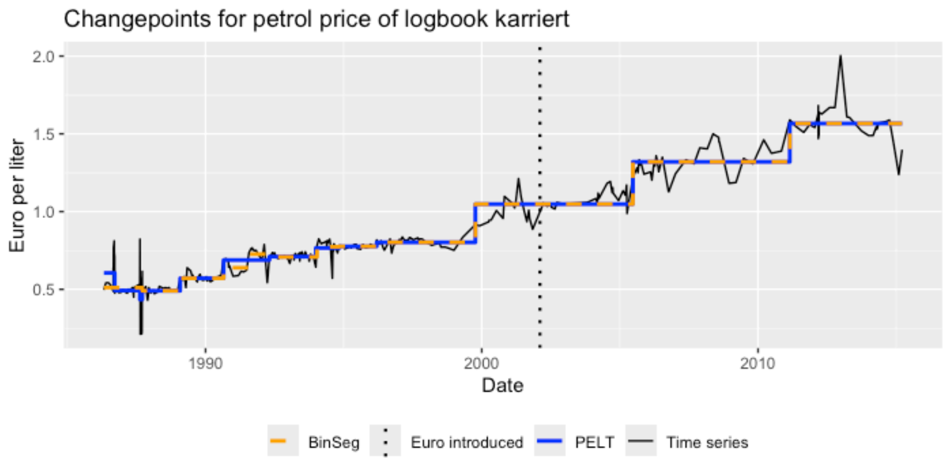

Figure 8: Changepoints for petrol price of logbook karriert

Trend Detection Methods: LOESS

LOESS (Locally Estimated Scatterplot Smoothing): A nonparametric method for visualizing and estimating smooth trends in time series.

Model:

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, \ldots, n$$

where $g(x)$ is a smooth function and ϵ_i are independent errors with $\mathbb{E}(\epsilon_i) = 0$ and constant variance.

Algorithm (for each x_i):

1. **Compute kernel weights** $w_k(x_i)$ using a weight function that satisfies certain conditions
2. **Fit weighted polynomial regression of degree** d to attain regression coefficients $\hat{\beta}_0(x_i), \ldots, \hat{\beta}_d(x_i)$ and the smoothed value $\hat{y}_i = \sum_{j=0}^{d} \hat{\beta}_j(x_i) x_i^j$

Robustness Step (optional):

1. Compute residuals: $r_i = y_i - \hat{y}_i$
2. Compute robustness weights δ_k using another weight function and the median of all residuals
3. Update weights:
$$\tilde{w}_k(x_i) = \delta_k \cdot w_k(x_i)$$
and refit the local polynomial regression using the updated weights.
4. Repeat the robustness step for a fixed number of iterations (typically $t = 2$).

LOESS Trends in petrol consumption

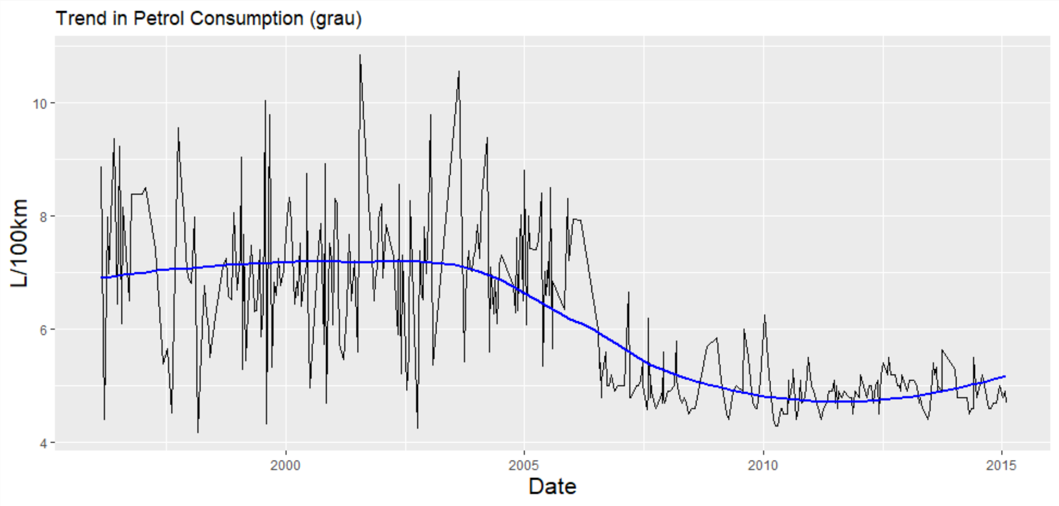

Figure 9: LOESS trend in petrol consumption for *grau*

LOESS Trends in petrol consumption

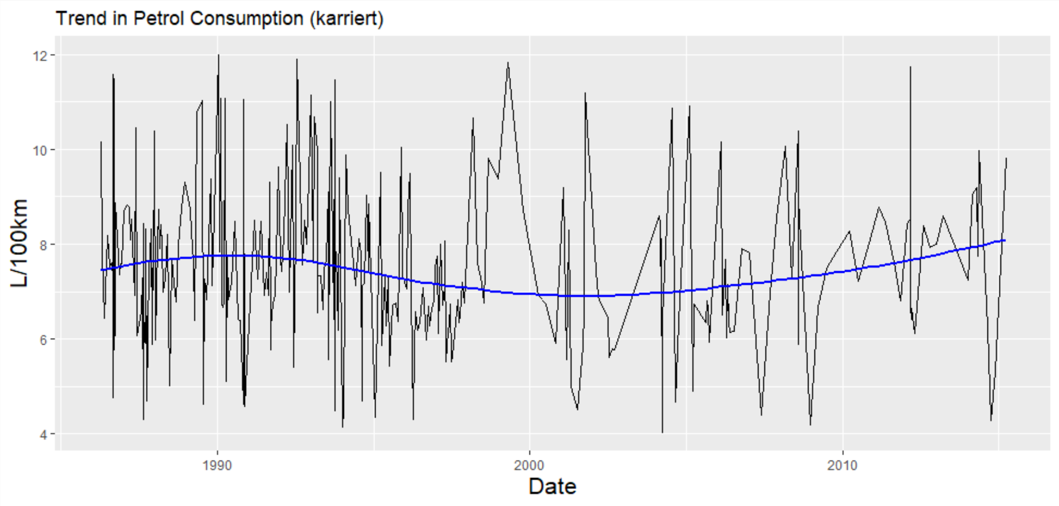

Figure 10: LOESS trend in petrol consumption for *karriert*

LOESS Trends in petrol price

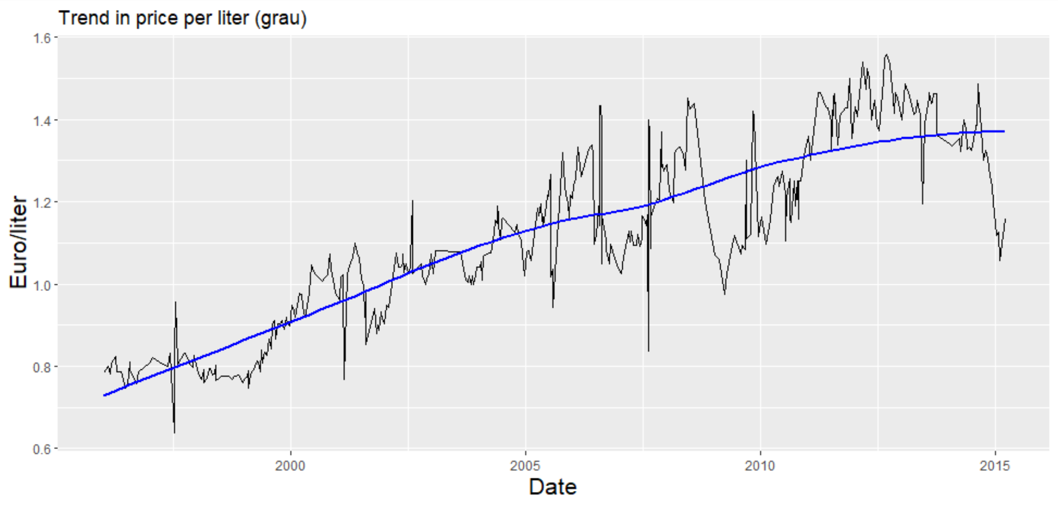

Figure 11: LOESS trend in petrol price for *grau*

LOESS Trends in petrol price

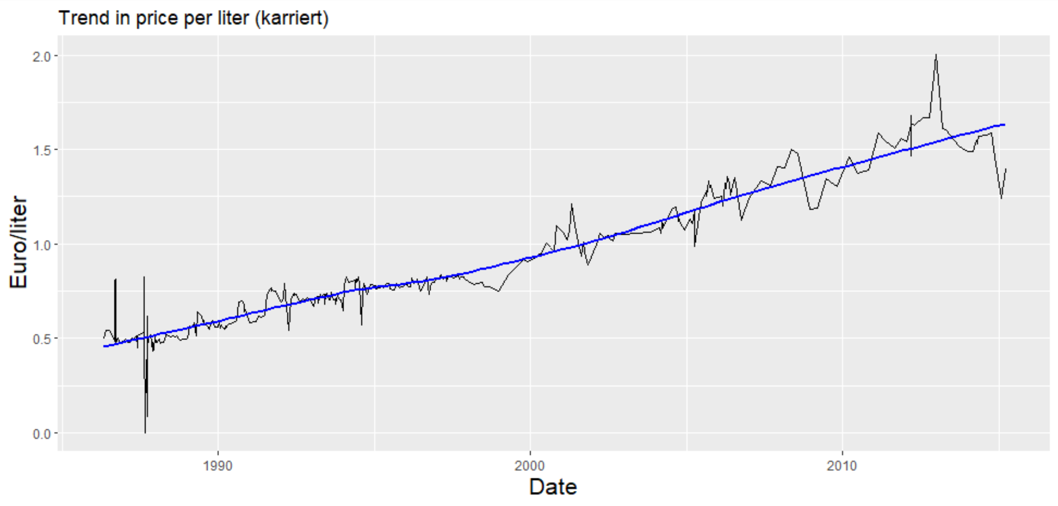

Figure 12: LOESS trend in petrol price for *karriert*

A non-parametric test used to detect a monotonic trend (increasing or decreasing) in a time series.

- **Two-sided hypotheses:**

$$H_0 : \mathbb{P}(X_j > X_k) = \mathbb{P}(X_j < X_k) \quad \text{vs.} \quad H_1 : \mathbb{P}(X_j > X_k) \neq \mathbb{P}(X_j < X_k) \quad \text{for all } j > k$$

- **Test statistic:** $S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(X_j - X_k)$ counts the number of increasing and decreasing pairs of observations.
- **Standardized test statistic:**

$$Z = \begin{cases} \frac{S-1}{\sqrt{\mathrm{Var}(S)}}, & S > 0 \\ 0, & S = 0 \\ \frac{S+1}{\sqrt{\mathrm{Var}(S)}}, & S < 0 \end{cases}$$

- **Decision rule:** Under H_0, $Z \sim \mathcal{N}(0, 1)$. Reject H_0 if $|Z| > z_{1-\alpha/2}$.

Modified Mann-Kendall Test

The standard Mann-Kendall test assumes that the data are **serially uncorrelated**. When **autocorrelation** is present, it inflates the Type I error rate. To address this, the **variance of the test statistic** is adjusted.

Variance Correction:

$$\mathrm{Var}^*(S) = \mathrm{Var}(S) \cdot \frac{n}{n^*}$$

- n is the actual sample size (ASS)
- n^* is the effective sample size (ESS), adjusted for autocorrelation

Effective Sample Size (ESS) n^* must be estimated from the data.

The R package `modifiedmk` provides the `mmkh` function.

Autocorrelation

The `acf()` function in R computes and plots the sample autocorrelation function (ACF).

Sample autocorrelation at lag t:

$$r_t = \frac{c_t}{c_0} \quad \text{where} \quad c_t = \frac{1}{n} \sum_{s=1}^{n-t} (X_{s+t} - \bar{X})(X_s - \bar{X})$$

- $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean
- c_0 is the sample variance (i.e. autocovariance at lag 0)

Autocorrelation in petrol consumption

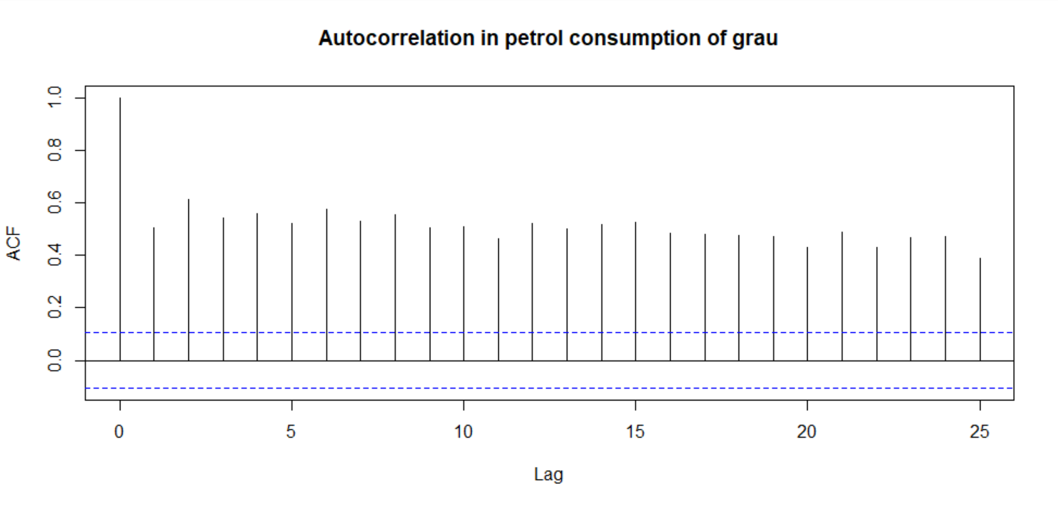

Figure 13: Autocorrelation in petrol consumption of *grau*

Autocorrelation in petrol consumption

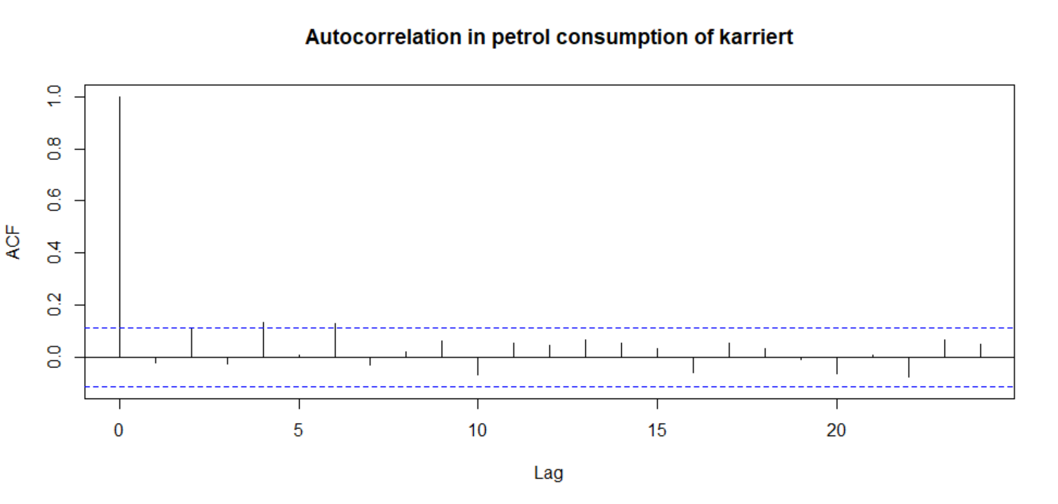

Figure 14: Autocorrelation in petrol consumption of *karriert*

Autocorrelation in petrol price

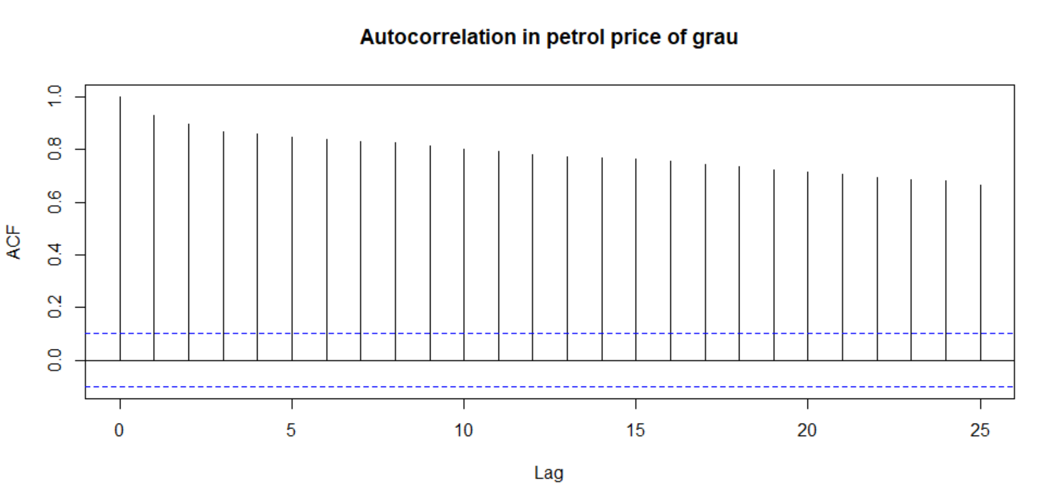

Figure 15: Autocorrelation in petrol price of *grau*

Autocorrelation in petrol price

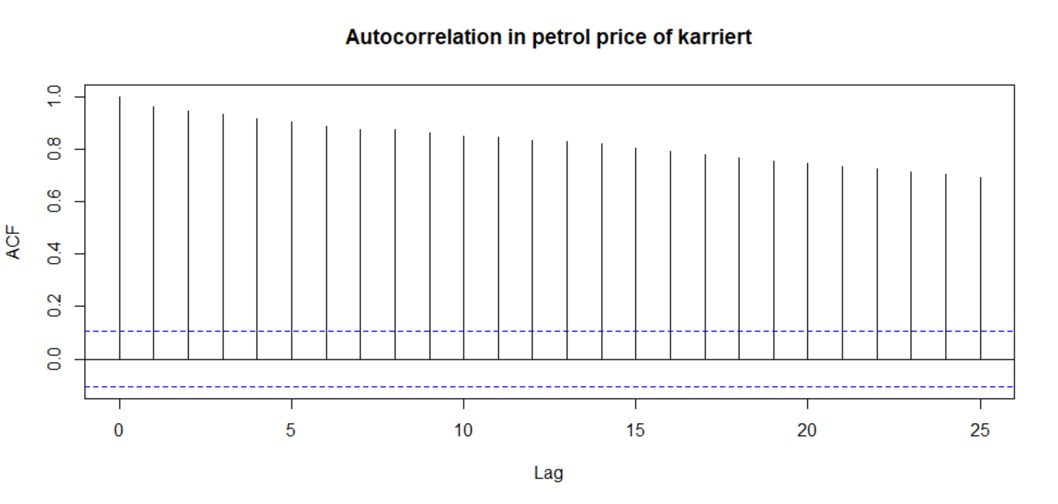

Figure 16: Autocorrelation in petrol price of *karriert*

Mann-Kendall Test Results

Time series	p-value	Test result
consumption grau	8.4655×10^{-7}	H_0 rejected $\Rightarrow$ trend
consumption karriert	0.1190	H_0 not rejected $\Rightarrow$ no trend
price grau	1.7500×10^{-17}	H_0 rejected $\Rightarrow$ trend
price karriert	5.0373×10^{-6}	H_0 rejected $\Rightarrow$ trend

Table 1: The p-values from the modified Mann-Kendall tests for all four time series.

Summary

- Extensively cleaned and standardized time-series data by correcting dates, converting currencies, and resolving missing or inconsistent values for accurate time-based analysis.
- One changepoint detected for petrol consumption in grau when car changes
- Many changepoints for petrol prices but not at the introduction of Euro
- The Mann-Kendall test detects a trend for 3 out of 4 time series (no trend in consumption of *karriert*)