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

— Project Report —
Case Studies II

Abdul Muqsit Farooqi

Project Supervisors:
Prof. Dr. Christine Müller
Dr. Mirko Jakubzik

September 2, 2025

TU Dortmund

Contents

1 Introduction **1**

2 Description of the Problem **2**
2.1 The Objectives of the Project 2
2.2 The Data Material 2

3 Statistical Method **4**
3.1 Pruned Exact Linear Time (PELT) Method 4
3.2 Binary Segmentation Method 5
3.3 LOESS for Trend Detection 6
3.4 Autocorrelation 7
3.5 Mann–Kendall Test 8

4 Statistical Analysis **9**
4.1 Preprocessing of Odometer Reading Data for Karriert and Grau 9
4.2 Preprocessing of Fuel Consumption for Karriert and Grau 11
4.3 Currency Conversion and Data Preprocessing 11
4.4 Time Series for Petrol Consumption and Petrol Prices 12
4.4.1 Exploratory Analysis of Petrol Price Time Series 12
4.4.2 Exploratory Analysis of Petrol Consumption Time Series 14
4.5 Change Point Detection in Fuel Consumption 15
4.6 Trend Detection in Petrol Consumption 17
4.6.1 Autocorrelation Structure in Petrol Consumption 18
4.7 Change Point Detection in Petrol Prices 20
4.8 Trend Detection in Petrol Prices 22
4.8.1 Autocorrelation Structure in Petrol Prices 24

5 Summary **26**

References **27**

A Currency Conversion Details **28**
A.1 Currency Codes 28
A.2 Conversion Rates to Euro 29

1 Introduction

Petrol prices and petrol consumption have long been of economic, public, and environmental interest. Tracking fuel prices can help detect economic shifts and policy effects, such as currency changes. This project aims to explore historical refueling data to uncover trends and detection in price and car change.

Two refueling logs, recorded between 1986 and 2015, that include datasets of the refueling date, volume filled (liters), odometer readings, currency, price, and fuel consumption. The goals of this project are as follows:

- Create time series for petrol consumption (liters per 100 km) and petrol prices (Euro per liter),
- Detect change-points and trends in the consumption data.
- Investigate whether vehicle changes can be identified through shifts in consumption behavior.
- Detect change-points and trends in petrol pricing,
- Evaluate whether the transition from Deutsche Mark (DM) to Euro (EUR) introduced a structural break in fuel prices.

In several stages, the analysis proceeded. Firstly, comprehensive data preprocessing was performed to correct wrong date entries, address missing values, and standardize price data across multiple currencies, e.g., converting all currencies to euro. Prices recorded in Deutsche Mark (DEM) and other foreign currencies were converted to Euros using historical exchange rates or fixed rates. Incorrect or missing odometer readings and fuel volumes were corrected or interpolated.

Following data cleaning, petrol consumption was calculated as liters per 100 kilometers driven, and petrol prices were normalized as euros per liter. Using the preprocessed time series, the Pruned Exact Linear Time (PELT) and Binary Segmentation algorithms were then applied to detect structural changes (the introduction of the euro and car changes) in both fuel consumption and price. Additionally, Mann–Kendall trend test was then used to assess overall trends. Locally Estimated Scatterplot Smoothing (LOESS) were employed to reduce noise and validate the robustness of detected patterns.

The statistical analysis revealed a significant downward trend in petrol consumption for the "Grau" dataset, with structural breakpoints indicating potential changes in driving behavior or a switch in vehicles. For the "Karriert" dataset, no statistically significant trend was observed. Petrol prices show a clear upward trend in both datasets. However, no abrupt changepoint was detected around the 2002 currency change from DM to EUR.

The report is structured as follows: Chapter 2 defines objectives of the project and details the data material. Chapter 3 presents the statistical methods used for trend analysis and change point detection. Chapter 4 provides an in-depth presentation of the results and interprets them in the context of the project goals. Chapter 5 concludes the report by summarizing the findings and suggesting directions for future analysis.

2 Description of the Problem

Understanding long-term trends in petrol consumption and fuel pricing is crucial for analyzing vehicle efficiency, driving behavior, and the effects of economic changes like currency transitions. This project utilizes two refueling logbooks to investigate whether time series patterns in fuel usage and prices reveal significant structural changes or trends.

By applying statistical methods for changepoint detection and trend analysis, is aimed at identifying economic shifts such as a car change or the switch from DM to Euro. The underlying data, derived from two detailed logbooks, allows us to explore how individual consumption and fuel pricing evolved over time and whether such changes are statistically detectable.

2.1 The Objectives of the Project

The aim of this project is to explore changes and trends in petrol consumption and petrol prices using detailed refueling records from two logbooks. The central questions are:

- How do petrol consumption and fuel prices evolve over time?
- Can changepoints in the data be detected, indicating shifts in usage behavior, pricing policies, or vehicle changes?
- Is there a significant structural break in petrol prices associated with the introduction of the euro currency?

Statistically, the project involves several steps: constructing time series from irregularly spaced entries, cleaning (removing outliers or missing data entries) and standardizing inconsistent data (e.g., currency conversion), detecting structural changes using changepoint detection methods such as PELT and Binary Segmentation, and evaluating trend significance through the modified Mann–Kendall trend test.

2.2 The Data Material

The data used in this project consists of two real-world fuel logbooks:

- **Tankbuch_grau.xlsx**, containing 383 entries from 1996 to 2015,
- **Tankbuch_karriert.xlsx** (karr), containing 357 entries from 1986 to 2015.

Each logbook consists of the following variables:

- `Datum` – Date of refueling (daily resolution),
- `Kilometerstand` – Odometer reading (in km),
- `Liter` – Volume of fuel refueled (in liters),
- `Preis` – Total amount paid for the refueling (in original currency),

- `Währung` – Currency in which the payment was made,
- `Verbrauch` – Fuel consumption (in L/100 km, only available for some entries in the Grau dataset but not available in "Karriert" dataset).

The datasets represent an **observational study** derived from personal vehicle usage. Data entries are irregular in time.

Several issues affected the quality of the raw data:

- **Missing values** occurred in odometer readings, prices, and currencies (e.g., 8 missing odometer entries and 4 missing currencies in the "Grau" data; 2 missing odometer entries, 10 missing prices, and 10 missing currencies in the "Karriert" data).
- **Inconsistent units** due to multiple currencies used across time and geography, including DM, Czech Koruna, Austrian Schilling, British Pound, and others. These were manually converted to euro using historical or fixed exchange rates.
- **Temporal errors** were detected in the form of non-monotonic date sequences. These were identified through differencing and corrected based on logical inference and interpolation.
- **Data entry anomalies**, such as extreme or illogical consumption values, were identified and removed (e.g., values outside the 4–12 L/100 km range).

Following cleaning, derived variables such as price per liter (Euro/liter) and petrol consumption (L/100 km) were computed. These variables form the basis of the time series analysis, changepoint detection, and trend detection in the later stages of the project.

3 Statistical Method

In this Section, various statistical methods are discussed for the analysis of the data. For the calculation and graphical representation of statistical measures R Software version, 4.2.3 (R Core Team 2023) is used with the package `dplyr` (Wickham 2022) for data manipulation and transformation, `changepoint` for detecting changepoints in data (Killick, Fearnhead, and I. Eckley 2014), `cpt.meanvar` a function to detect changepoints in both the mean and variance of a data sequence (Killick 2024), ggplot2 for data visualization (Wickham 2016), and `zoo` used for working with time series and ordered data (Zeileis and Grothendieck 2005).

3.1 Pruned Exact Linear Time (PELT) Method

The Pruned Exact Linear Time (PELT) method is a statistically robust algorithm for detecting multiple changepoints in a time series. It was introduced by (Killick, Fearnhead, and I. A. Eckley 2012) and is widely recognized for combining exact segmentation with computational efficiency.

Key Features

- **Exactness:** PELT returns the segmentation that minimizes a penalized cost function, ensuring statistically optimal changepoint detection.
- **Efficiency:** Under mild conditions, the algorithm achieves linear computational complexity in the number of observations.
- **Pruning:** By discarding candidate changepoints that cannot improve the objective function, PELT avoids unnecessary computations and reduces runtime.

Objective Function

PELT minimizes the following penalized cost function (Killick, Fearnhead, and I. A. Eckley 2012):

$$\sum_{i=1}^{m+1} \left[\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta \right]. \tag{1}$$

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)

Let $F(s)$ represent the minimum cost required to segment the series $y_{1:s}$ in accordance with (1), and define $\mathcal{T}_s = \{\tau \mid 0 = \tau_0 < \tau_1 < \cdots < \tau_m < \tau_{m+1} = s\}$ as the collection of valid change-point sequences up to time s (Killick, Fearnhead, and I. A. Eckley 2012, p. 1591). The initial condition is given by $F(0) = -\beta$. Using this, the following recursion

$$F(s) = \min_{t} \left\{ F(t) + \mathcal{C}(y_{(t+1):n}) + \beta \right\}$$

enables efficient computation of $F(s)$ by reusing previously calculated values of $F(t)$ (Killick, Fearnhead, and I. A. Eckley 2012, p. 1592).

Pruning Rule

The key innovation of PELT is a pruning condition that enables linear runtime. If a constant K exists such that for all $t < s < T$ (Killick, Fearnhead, and I. A. Eckley 2012):

$$\mathcal{C}(y_{(t+1):s}) + \mathcal{C}(y_{(s+1):T}) + K \leq \mathcal{C}(y_{(t+1):T}),$$

Furthermore, a candidate point t can be excluded from further evaluation if it satisfies (Killick, Fearnhead, and I. A. Eckley 2012)

$$F(t) + \mathcal{C}(y_{(t+1):s}) + K \geq F(s).$$

This pruning criterion is the key factor enabling the PELT algorithm to attain linear computational complexity, provided certain mild assumptions are met.

PELT's pruning mechanism ensures that only changepoint candidates with potential to improve the optimal cost are retained during computation, enabling it to scale efficiently even for long sequences.

3.2 Binary Segmentation Method

Binary Segmentation (BS) is one of the most widely used and well-established algorithms for detecting multiple changepoints in time series analysis. The approach builds upon a single changepoint detection method by recursively applying it to subsets of the data.

The process begins by applying the single changepoint detection method to the entire dataset. Specifically, the method tests whether there exists an index τ such that the cost of segmenting the data at τ, plus a penalty β, is less than the cost of modeling the entire data sequence without any changepoint (Killick, Fearnhead, and I. A. Eckley 2012):

$$\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{(\tau+1):n}) + \beta < \mathcal{C}(y_{1:n})$$

If this inequality is not satisfied, then no changepoint is detected, and the algorithm terminates. If a changepoint is found, the data is divided into two segments: one before and one after the detected changepoint. This recursive partitioning continues until no further changepoints are found in any segment.(Killick, Fearnhead, and I. A. Eckley 2012, p. 1591)

While Binary Segmentation is computationally efficient and straightforward to implement, it does not guarantee finding the global minimum of the overall cost function (1). However, it performs well in practice, particularly when changepoints are well-separated.

3.3 LOESS for Trend Detection

LOESS (Locally Estimated Scatterplot Smoothing) was applied to estimate and visualize smooth trends within the time series data. LOESS is a nonparametric regression technique that constructs a smooth curve through noisy observations without requiring a predefined global model. Instead, it locally approximates the underlying structure by fitting low-degree polynomials to subsets of the data.

At each data point x_i, LOESS carries out the following steps:

1. **Compute kernel weights:**

$$w_k(x_i) = W\left(\frac{x_k - x_i}{h_i}\right), \quad k = 1, \dots, n,$$

   Here, h_i denotes the distance from x_i to its r-th nearest neighbor (Cleveland 1979, p. 830). According to Cleveland (1979, p. 829), the weight function $W(x)$ must satisfy the following conditions:

   (a) $W(x) > 0$ for $|x| < 1$;

   (b) $W(-x) = W(x)$;

   (c) $W(x)$ is a nonincreasing function for $x \geq 0$;

   (d) $W(x) = 0$ for $|x| \geq 1$.

   A commonly used function that satisfies these conditions is the tricube kernel (Cleveland 1979, p. 831):

$$W(x) = \begin{cases} (1 - |x|^3)^3, & \text{for } |x| < 1, \\ 0, & \text{for } |x| \geq 1. \end{cases}$$

2. **Fit a locally weighted polynomial:** As described in (Cleveland 1979, p. 830), the next step involves fitting a polynomial of degree d to the locally weighted data by minimizing (Cleveland 1979, p. 830)

$$\sum_{k=1}^{n} w_k(x_i) \left(y_k - \beta_0 - \beta_1 x_k - \dots - \beta_d x_k^d\right)^2.$$

   The fitted value at x_i is then given by (Cleveland 1979, p. 830)

$$\hat{y}_i = \sum_{j=0}^{d} \hat{\beta}_j(x_i) {x_i}^j.$$

The model is expressed as (Cleveland 1979, p. 830):

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

where $g(x)$ is a smooth function, and the error terms ϵ_i are random variables with zero mean, $\mathbb{E}(\epsilon_i) = 0$, and constant variance (Cleveland 1979, p. 830).

Modifying LOESS with the following interactive re-weighting scheme applied to the observations enhances robustness against outliers:

1. **Calculate residuals** (Cleveland 1979, p. 831)**:**

$$e_i = y_i - \hat{y}_i.$$

2. **Determine robustness weights:**

$$\delta_k = B\left(\frac{e_k}{6s}\right), \quad B(x) = \begin{cases} (1 - x^2)^2, & \text{for } |x| < 1, \\ 0, & \text{for } |x| \geq 1. \end{cases}$$

   where s is the median of the absolute residuals $|e_1|, \ldots, |e_n|$ (Cleveland 1979, p. 831).

3. **Refit with new weights** (Cleveland 1979, p. 830) **:**

$$\tilde{w}_k(x_i) = \delta_k \cdot w_k(x_i),$$

4. **Iterate:** Repeat the robustness weighting step for a few iterations. However, Cleveland suggests that two iterations are generally sufficient (Cleveland 1979, p. 834).

3.4 Autocorrelation

Prior to implementing the modified Mann-Kendall test, it is important to verify whether the time series exhibits autocorrelation. The function `acf()` in R was used.

This involves computing the ratio r_t as described in (Venables and Ripley 2002, p. 390):

$$c_t = \frac{1}{n} \sum_{s=\max(1,-t)}^{\min(n-t,n)} [X_{s+t} - \overline{X}][X_s - \overline{X}], \quad r_t = \frac{c_t}{c_0}$$

where (Venables and Ripley 2002, p. 390):

1. $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ denotes the sample mean,
2. c_0 represents the sample variance, which corresponds to the autocovariance at lag 0,
3. c_t is the sample autocovariance at lag t,
4. r_t is the sample autocorrelation at lag t.

3.5 Mann–Kendall Test

The Mann–Kendall test is a nonparametric test. It is used to detect monotonic trends (increasing or decreasing) in time series even if the data is not normally distributed.

The test assesses the following hypotheses:

H_0: There is no monotonic trend in the data.
H_1: There is a monotonic increasing or decreasing trend.

Under the null hypothesis H_0, the standardized test statistic Z is assumed to follow a standard normal distribution, $Z \sim \mathcal{N}(0,1)$. The null hypothesis is rejected at a significance level α if:

$$|Z| > z_{1-\alpha/2}.$$

The Mann-Kendall statistic S is defined as (Yue and Wang 2004, p. 215):

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(Y_j - Y_k),$$

where the sgn (θ) function is (Yue and Wang 2004, p. 215):

$$\operatorname{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases}$$

Under the null hypothesis, there is no trend; the expected value of S is zero. For large sample sizes ($n >= 8$), the distribution of S can be approximated by a normal distribution with the following variance (Yue and Wang 2004, p. 216):

$$\operatorname{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{n} t_i i(i-1)(2i+5)}{18},$$

where t_i represents the number of tied observations of extent i (Yue and Wang 2004, p. 216).

The standardized test statistic Z is then computed as (Yue and Wang 2004, p. 216):

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

To account for autocorrelation, the variance is adjusted using the following expression (Yue and Wang 2004, p. 206):

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

where n is the total number of observations, and n^* represents the "effective sample size" (Yue and Wang 2004, p. 206), which incorporates the impact of serial correlation (Yue and Wang 2004, p. 206):

$$n^* = \frac{n}{1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k},$$

where ρ_k denotes the autocorrelation coefficient at lag k.

4 Statistical Analysis

4.1 Preprocessing of Odometer Reading Data for Karriert and Grau

To ensure the reliability of the odometer data for datasets "Grau" and "Karriert," thorough preprocessing was performed to identify and correct date entry errors as well as inconsistencies in the recorded readings. The initial step involved checking the chronological sequence of the data by computing successive time differences between odometer entries using the `diff()` function on the date values. Negative time differences indicated that an entry's timestamp was earlier than that of its predecessor, suggesting a likely date entry error. Upon further inspection, several anomalies were detected, including mistyped years. These specific erroneous entries were manually corrected to restore the proper chronological order of observations.

Subsequently, attention was directed toward ensuring the completeness and consistency of the odometer readings. For entries with missing or corrupted dates, linear interpolation was applied to reconstruct plausible date values. This procedure generated a continuous time series without temporal gaps, enabling consistent tracking of vehicle usage over time.

To facilitate detection of sudden or implausible changes in odometer values, a new variable, *kilodiff*, was introduced. This variable represents the distance in kilometers traveled between consecutive entries.

The effects of these preprocessing measures are illustrated in the figures. Figure 1 displays the odometer readings for dataset "Grau" before and after the corrections, while Figure 2 provides the corresponding comparison for dataset "Karriert." In both cases, the raw data (red line) shows incorrect timestamps. Following the correction procedure, the processed data (green line) shows the corrected version of the time series having dates in ascending order. These outcomes indicate that the preprocessing successfully resolved the chronological inconsistencies and prepared the data for subsequent statistical analysis.

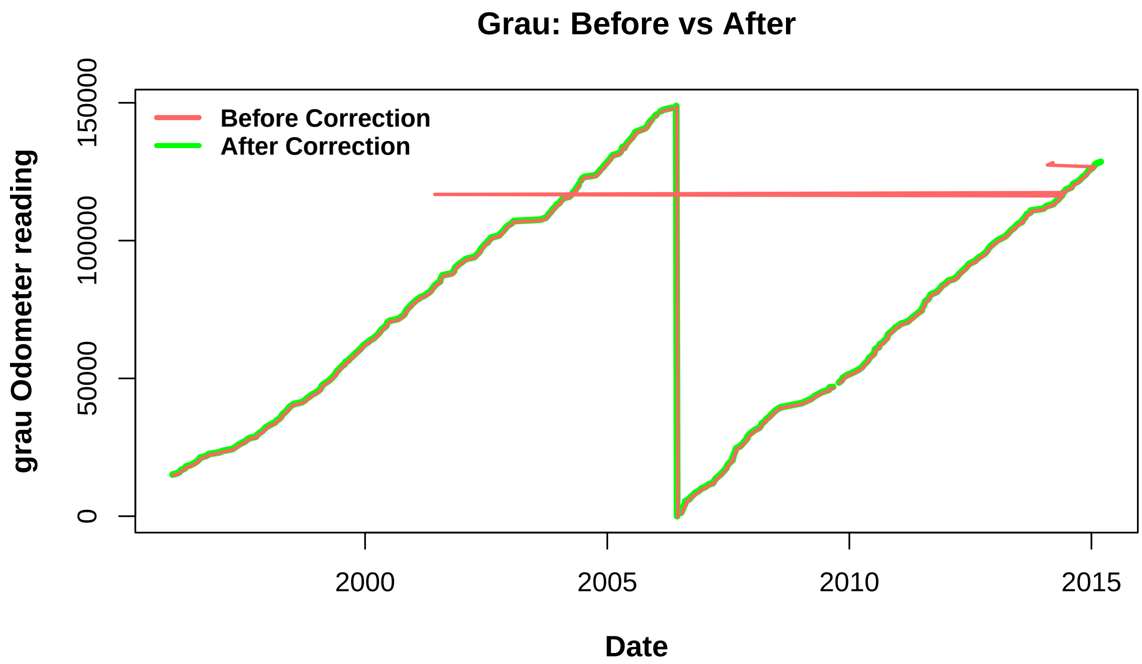

Figure 1: Odometer reading for dataset “Grau” before and after correction.

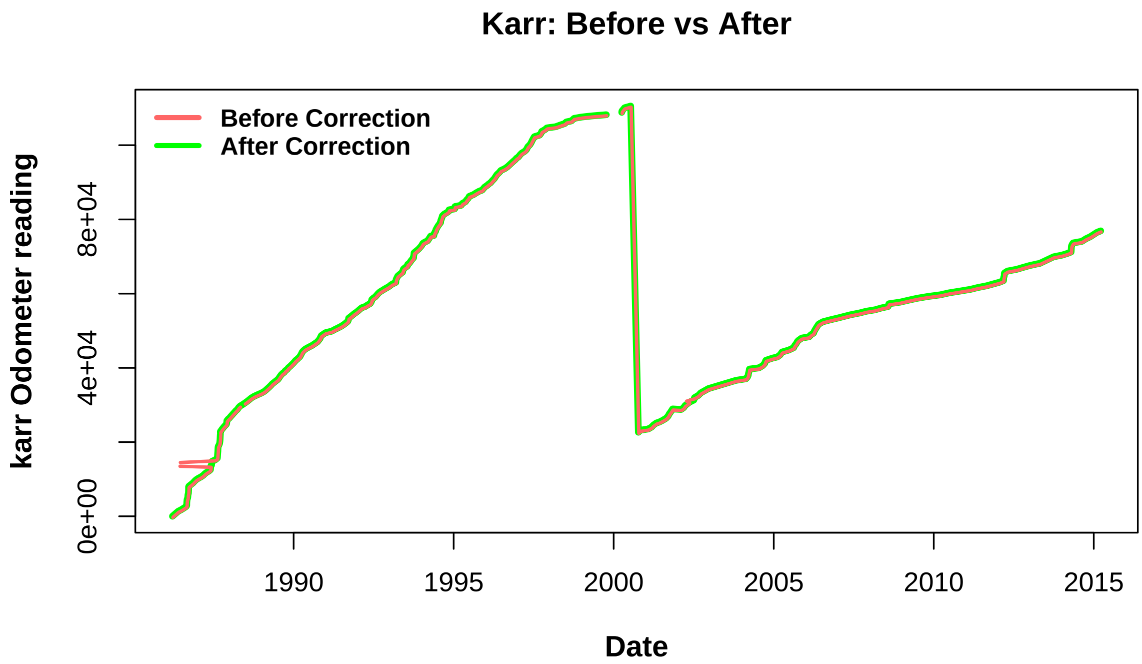

Figure 2: Odometer reading for dataset “Karriert” (karr) before and after correction.

4.2 Preprocessing of Fuel Consumption for Karriert and Grau

Fuel consumption for both “Karriert” and “Grau” was calculated. In the “Grau” dataset, a fuel consumption column already existed but contained only a few values. Therefore, fuel consumption was calculated for the remaining rows. But in the “Karriert” dataset, no such column existed, so a new fuel consumption column was first created and then populated by using the fuel consumption formula:

$$\text{L/100km} = \left(\frac{\text{Litres Used}}{\text{Kilometers Travelled}}\right) \times 100$$

where Kilometers Travelled $= K_n - K_{n-1}$. The first value in the dataset is left empty for fuel consumption if it did not already exist.

4.3 Currency Conversion and Data Preprocessing

Fuel price entries in the *grau* and *karriert* datasets were originally recorded in various currencies. To ensure consistency, all values were converted to euros (EUR) using fixed exchange rates or official reference rates from reliable sources such as the European Central Bank (European Central Bank 2024) and the historical dinar rates archive (kursna-lista.com 2024).

Initial screening identified all records with currencies other than Euro and Deutsche Mark (DM). For Deutsche Mark, a fixed conversion rate of 1 DM = 0.51129 EUR was applied. Similarly, fixed conversion factors were used for legacy currencies such as the Austrian Schilling (1 ATS = 0.07267 EUR), Czech Koruna (1 CZK ≈ 0.0295 EUR), and British Pound (1 GBP ≈ 1.41 EUR) for values recorded before the introduction of the euro. For currencies that continued to exist beyond 2001, such as the British Pound, Norwegian Krone, Danish Krone, and Polish Złoty conversion was performed using time-specific exchange rates based on the transaction date, rather than a single fixed value. This ensured an accurate representation of historical pricing fluctuations. Further details of the currency conversions and currency codes are given in Box A.2 and Table 4.

Manually entered corrections were necessary for rows where currency or price values were misrecorded. For example, entries with implausible Dinar values as shown in Table 1 were examined and corrected based on the other values in the dataset e.g. there were four rows in series, one price value was 13000 something and the other two were 9000 something but there was also 12.6 which was then corrected to 12600 based on other values of Dinar. For rows with missing currency labels, values were interpolated using surrounding entries (Table 1).

After currency conversion, values were then stored in the variable `PreisEU`. This standardization enabled comparison across time and currencies. Furthermore, prices per liter were computed using `PreisEU/Liter` and evaluated for the outliers. Those outliers and missing values were removed. The resulting cleaned and harmonized price data were then prepared for statistical modeling.

Table 1: Data Pre-Processing Steps Summary

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

4.4 Time Series for Petrol Consumption and Petrol Prices

4.4.1 Exploratory Analysis of Petrol Price Time Series

Figures 3 and 4 display the raw time series of petrol prices (in euros per liter) for the *grau* and *karriert* datasets, respectively. The plots span multiple decades, providing an initial visual assessment of the fuel price dynamics.

Before modeling, the original fuel consumption variable `verbrauch` was filtered to exclude values outside the interval $[4, 12]$ L/100km to remove extreme outliers. However, for the petrol price per liter (`ppl`), the complete historical record was retained for this exploratory stage to preserve the long-term structural information.

Both time series exhibit a gradual increase in price levels over the years, with several inflection points suggesting underlying structural changes. For example, in Figure 3, the *grau* series shows a steady rise beginning in the early 2000s, while Figure 4 reveals similar upward pressure on prices for *Karriert*, starting even earlier. This visual evidence suggests the presence of a positive trend.

Additionally, visual inspection reveals the missing values throughout both time series. These missing values were handled in subsequent analyses by filtering out incomplete observations before applying change point detection techniques, such as the pruned exact linear time (PELT) and Binary Segmentation (BinSeg) methods.

Overall, the exploratory plots reinforce the hypothesis of a sustained upward trend in petrol prices, which was further validated using Mann-Kendall test and autocorrelation diagnostics in the subsequent sections.

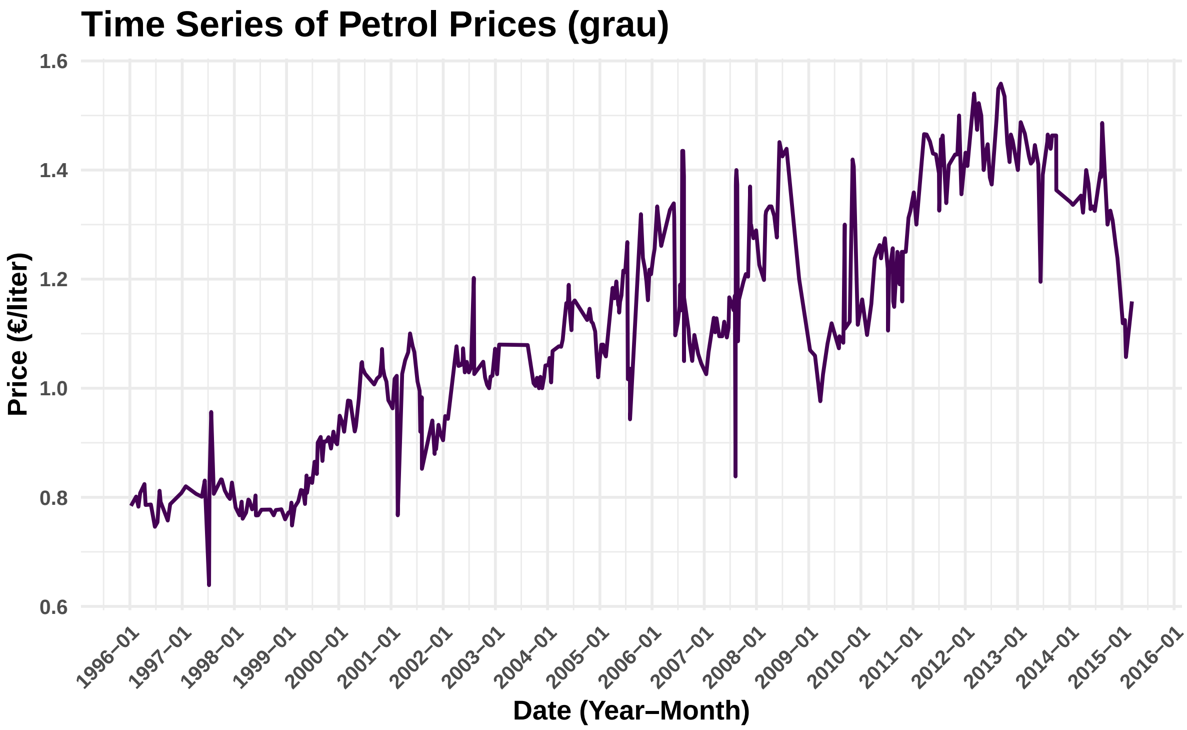

Figure 3: Time series of petrol prices (*Grau*).

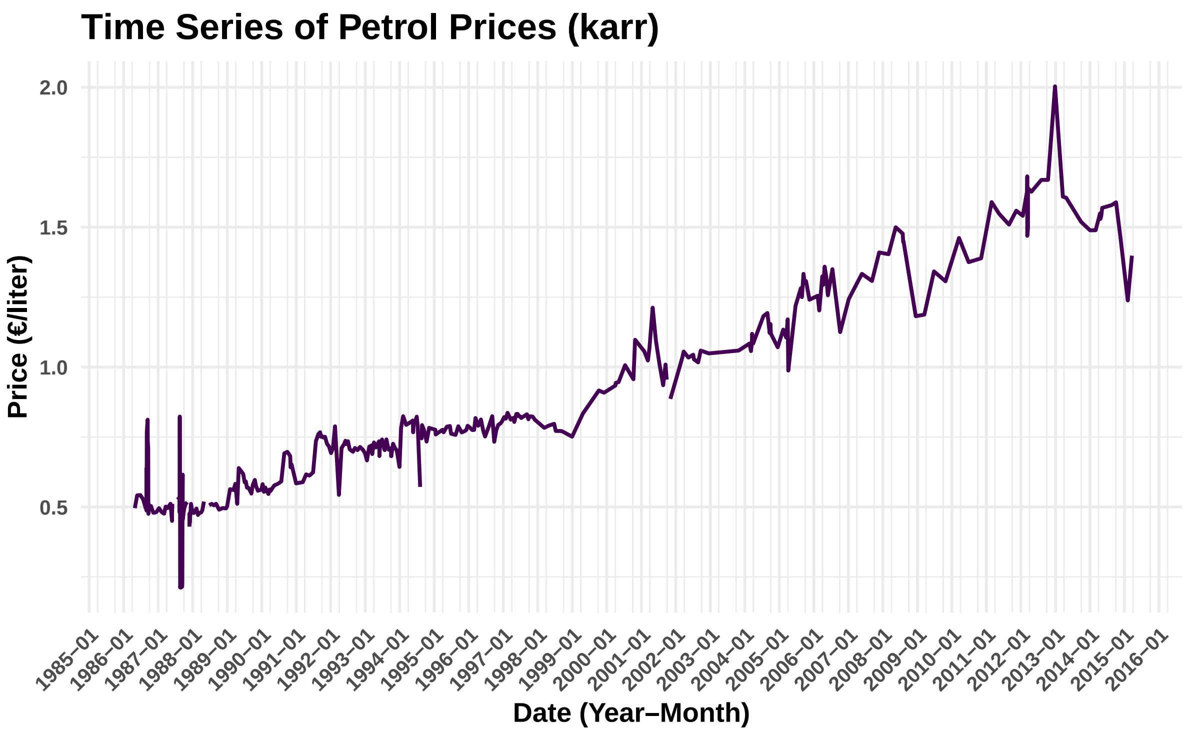

Figure 4: Time series of petrol prices *Karriert* (karr).

4.4.2 Exploratory Analysis of Petrol Consumption Time Series

Figures 5 and 6 display the raw time series of petrol consumption (in liters per 100 km) for the *grau* and *karriert* datasets, respectively. These plots provide an initial overview of temporal consumption patterns before any statistical modeling.

Outlier Filtering: Observations falling outside the interval $[4, 12]$ liters/100 km were considered implausible and removed for subsequent statistical analyses. In the case of the *grau* dataset, extreme values exceeding e.g. 60 L/100 km were excluded as clear outliers due to their unrealistic nature and the likelihood of recording or calculation errors.

Missing Values: The plots reveal the presence of gaps in the data series, corresponding to missing values. These may arise from incomplete refueling records, mechanical issues, or other reporting omissions. To apply change point detection method and the trend test, these missing entries were excluded.

Visual Insights: Despite data imperfections, both time series suggest possible patterns. The *grau* series exhibits a downward tendency in consumption over time, whereas *karriert* shows relatively more fluctuation but hints at a potential decline in earlier years. These qualitative patterns provide motivation to apply a test to check the trend and change point detection methods in the next section.

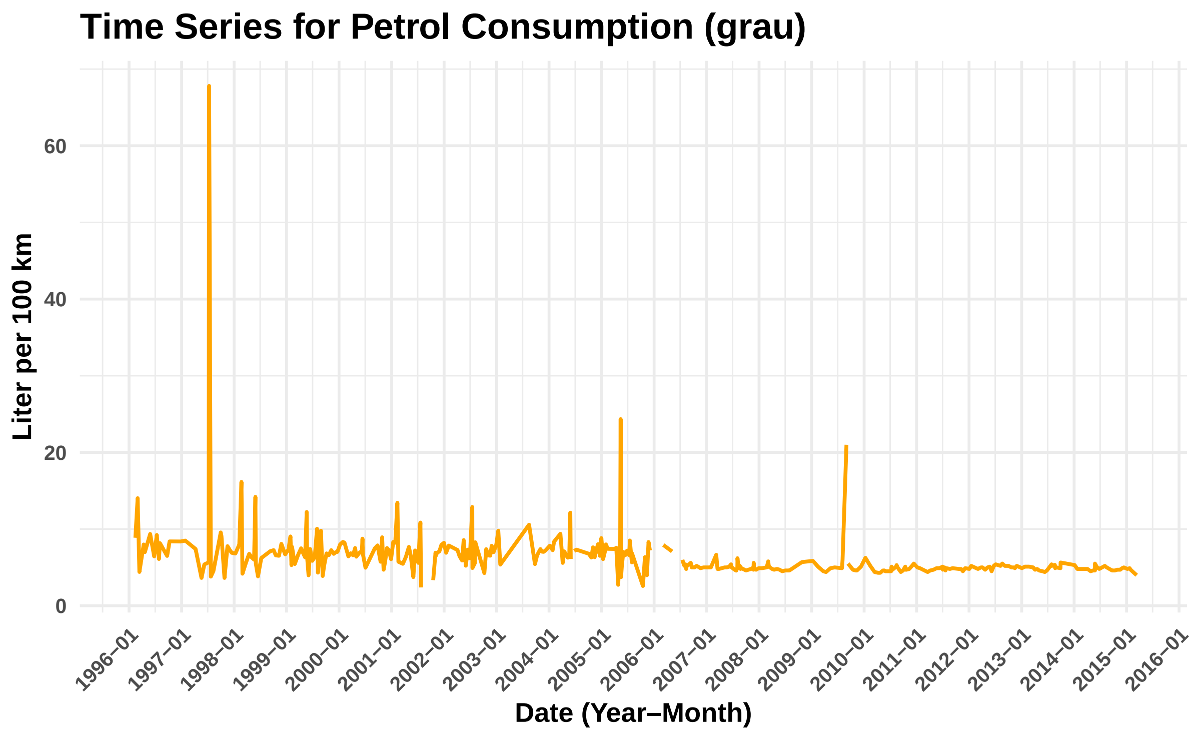

Figure 5: Time series of petrol consumption for *Grau*.

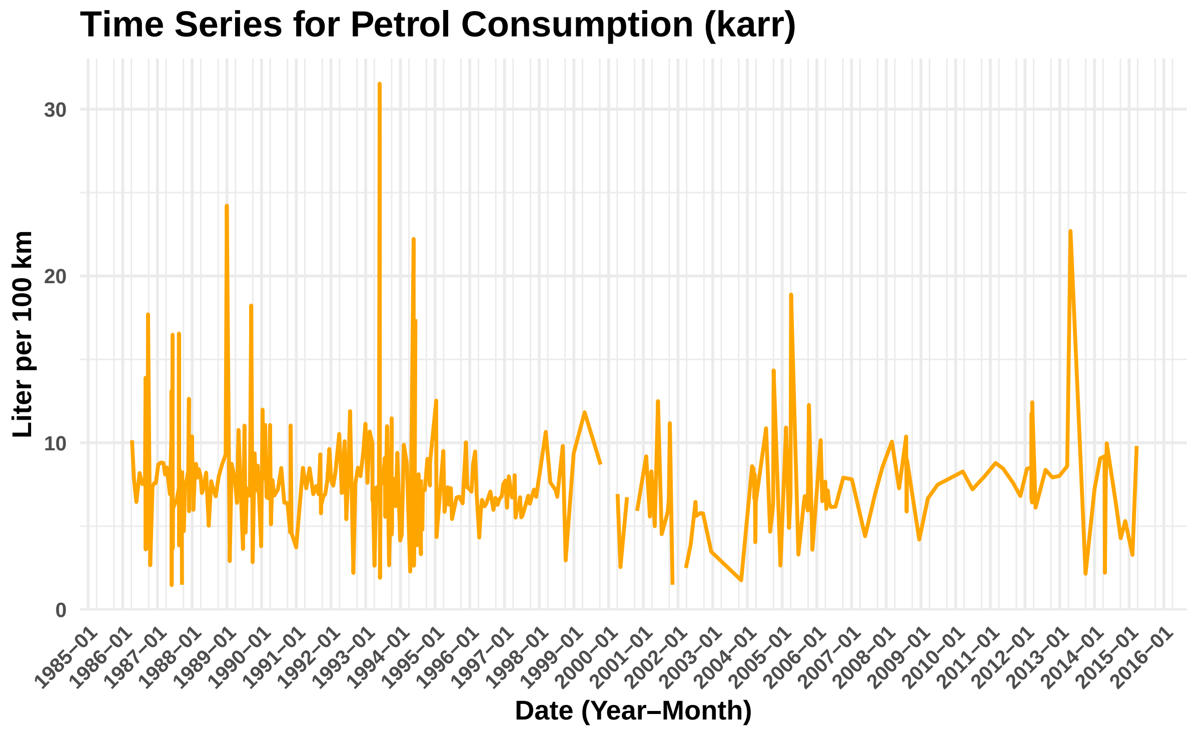

Figure 6: Time series of petrol consumption for *Karriert* (karr).

4.5 Change Point Detection in Fuel Consumption

Change point detection was applied to the cleaned fuel consumption data (`Verbrauch`) from the *grau* and *karriert* datasets. Observations outside the interval $[4, 12]$ liters/100 km were excluded as outliers. Two statistical methods were applied to detect structural breaks in the mean fuel consumption: the Pruned Exact Linear Time (PELT) algorithm and Binary Segmentation (BinSeg). These were used directly to the time series using `cpt.meanvar` function from the package (Killick 2024) identifying both mean and variance of a data series.

Figure 7 shows the detected change points in the **grau** dataset. A notable structural shift is observed around 2006, which corresponds exactly to the known car change (represented by the vertical dashed line). Both PELT (blue lines) and BinSeg (orange lines) reliably detected this transition, demonstrating high sensitivity to this major regime shift. As a result, the *grau* dataset consist of two cars.

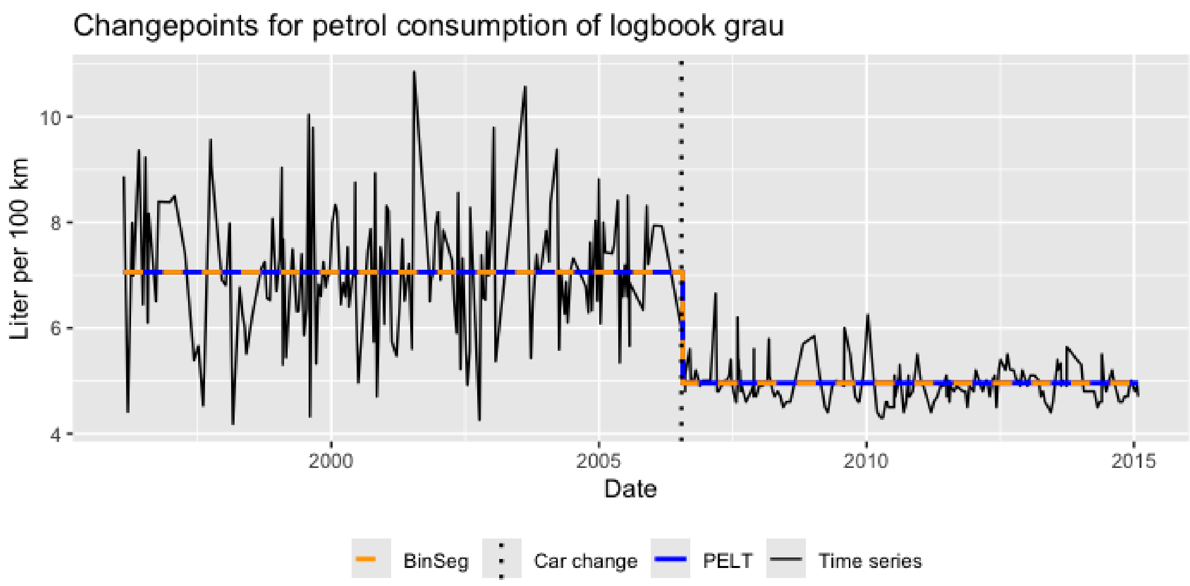

Figure 7: Change point detection in fuel consumption (Grau dataset) using PELT and BinSeg.

For the **karriert** dataset (Figure 8), both PELT and BinSeg are not able to detect a car change, while the dashed line indicates that there's some change, and by time series (black line) shows that the car is slowed gradually, (around 2001) but still cannot be concluded that it's a car change.

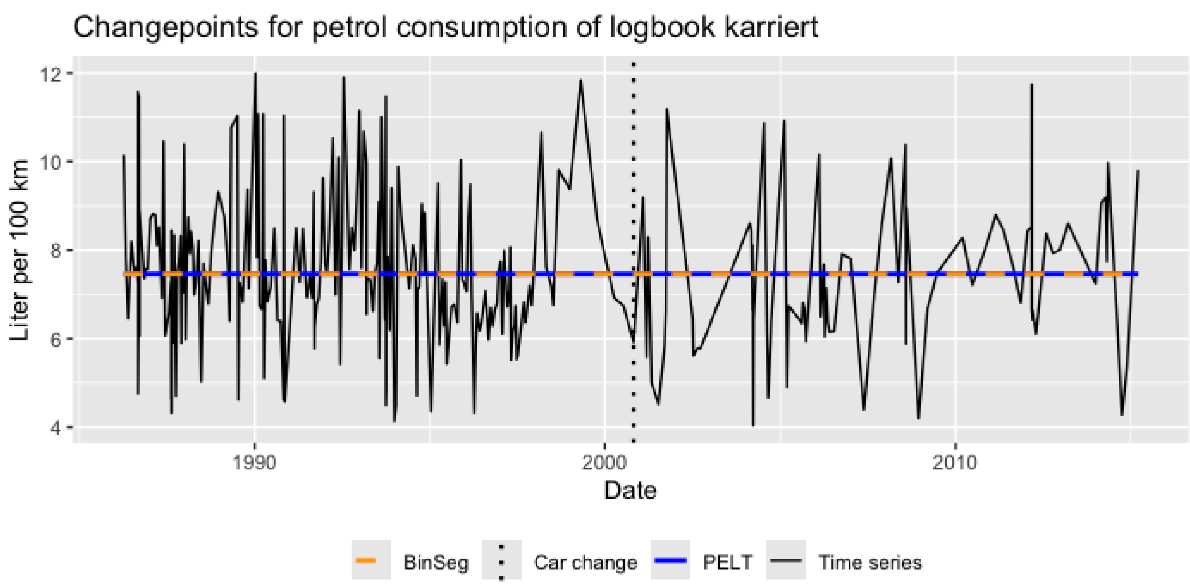

Figure 8: Change point detection in fuel consumption (Karriert dataset) using PELT and BinSeg.

4.6 Trend Detection in Petrol Consumption

To evaluate long-term changes in petrol consumption, trend analysis was conducted for both the *grau* and *karriert* datasets. The original consumption series (`Verbrauch`) was filtered to include only values within the range $[4, 12]$ to remove outliers and missing values were removed too. A LOESS (locally estimated scatterplot smoothing) curve was fitted to each time series to visualize gradual changes over time. Figures 9 and 10 display the raw consumption data along with the LOESS-smoothed trends for both vehicles.

To formally test the presence of a monotonic trend, the non-parametric Mann–Kendall trend test was applied. This method evaluates whether a significant upward or downward trend exists in a time series or even no trend. The results are summarized in Table 2.

For the **grau** dataset, the test yielded a p-value of 8.4655×10^{-7}, indicating a statistically significant trend. The trend direction was negative, indicating a gradual decrease in fuel consumption over time. In contrast, the test for the **karriert** dataset returned a p-value of 0.1914, which does not indicate statistical significance at the $\alpha = 0.05$ level. Therefore, no monotonic trend was detected in this case.

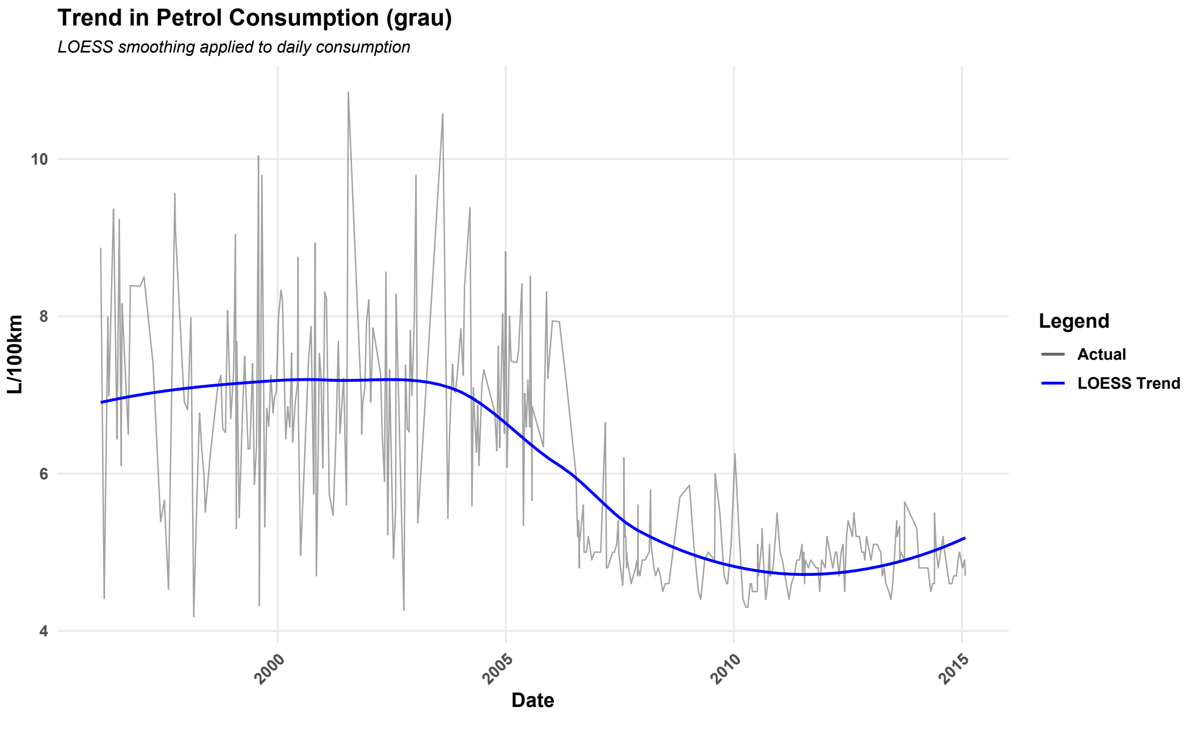

Figure 9: LOESS-smoothed consumption trend for dataset *Grau*.

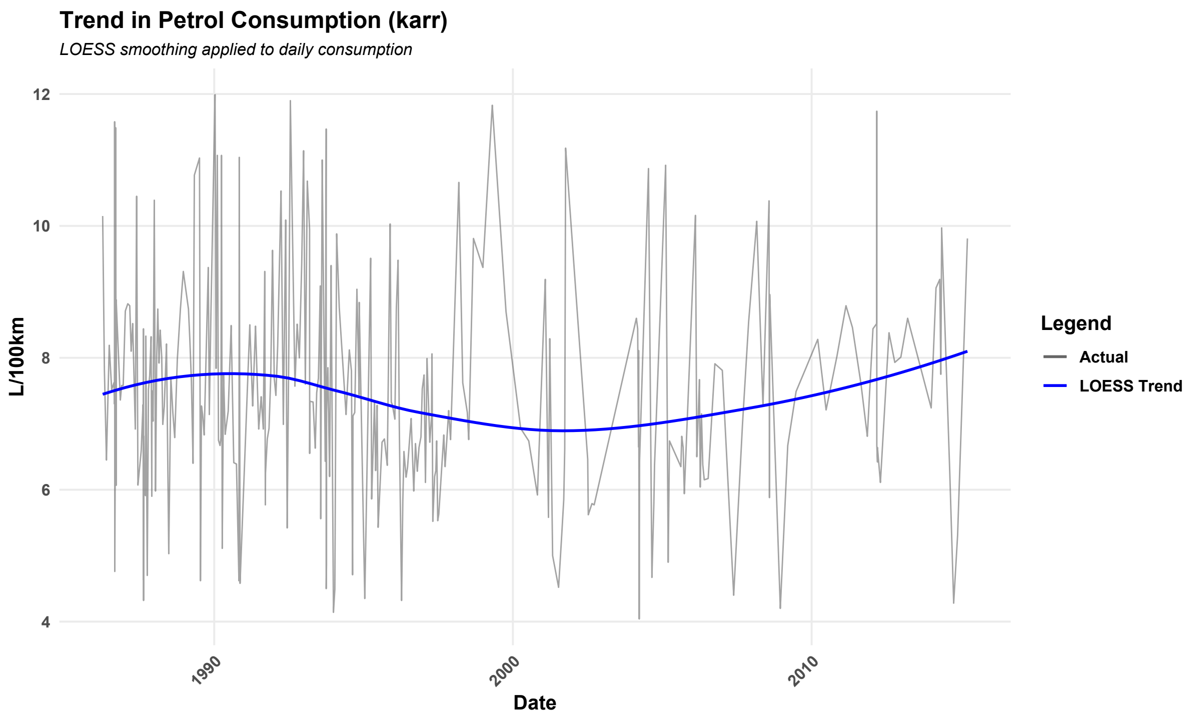

Figure 10: LOESS-smoothed consumption trend for dataset *Karriert* (karr).

Table 2: Results of the Mann–Kendall trend test for petrol consumption. Significance level $\alpha = 0.05$.

Dataset	p-value	Significant	Trend Direction
Grau	8.4655×10^{-7}	Yes	Decreasing
Karriert	0.1914	No	None

4.6.1 Autocorrelation Structure in Petrol Consumption

To assess the temporal dependency in petrol consumption data, autocorrelation functions (ACFs) were computed for both the *grau* and *karriert* time series. Figures 11 and 12 present the ACF plots, which display the correlation between observations at successive time lags.

In Figure 11, the **grau** dataset exhibits strong positive autocorrelation at lag 1, followed by a gradual decline. The slow decay pattern indicates a persistent time dependence, characteristic of a non-stationary or trending series. This supports the earlier Mann-Kendall test result indicating a significant downward trend. Importantly, several autocorrelation bars exceed the dashed horizontal confidence bounds, implying that the observed correlations are statistically significant and unlikely due to random noise.

Conversely, the ACF for **karriert** (Figure 12) also shows notable autocorrelation at lag 1, but with a slightly weaker magnitude and a quicker decay across subsequent lags. Most autocorrelations fall within the dashed confidence bounds, indicating that they are not statistically significant. This suggests short-term temporal dependence, but not a sustained trend over time. The absence of long-lasting autocorrelation aligns with the non-significant result from the Mann-Kendall test.

The dashed lines in both figures represent the 95% confidence bounds for testing the null hypothesis of no autocorrelation (white noise). Values that fall outside these lines indicate statistically significant autocorrelations at the corresponding lags.

Overall, the presence of statistically significant autocorrelation—especially in the *grau* series—justifies the use of trend detection methods that account for temporal structure.

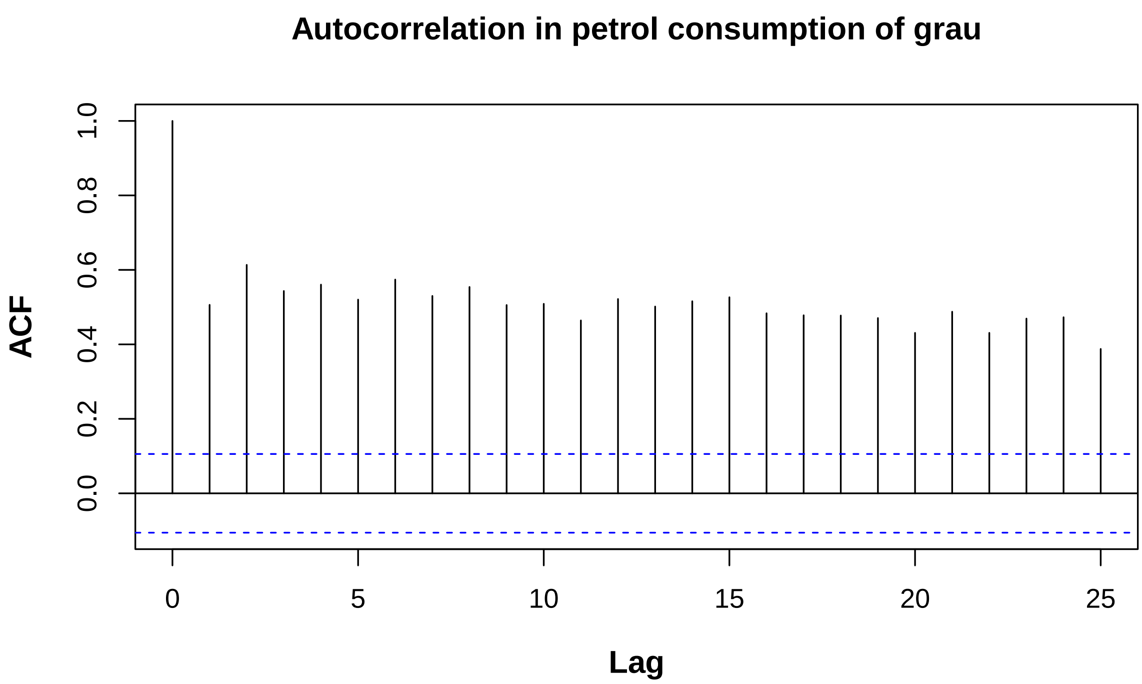

Figure 11: Autocorrelation in petrol consumption of *Grau*. Dashed lines indicate 95% confidence bounds.

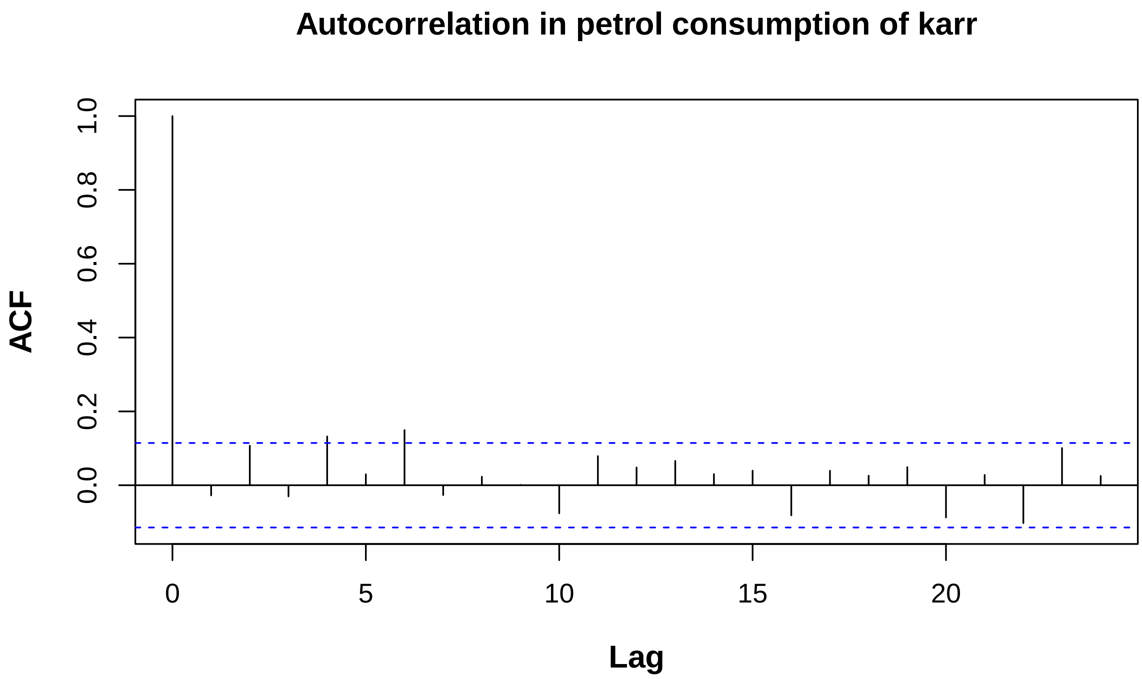

Figure 12: Autocorrelation in petrol consumption of *Karriert* (karr). Dashed lines indicate 95% confidence bounds.

4.7 Change Point Detection in Petrol Prices

To detect structural breaks in petrol price development, change point analysis was applied to the per-liter price time series `ppl` for both the *grau* and *karriert* datasets. Two prominent algorithms were employed: Pruned Exact Linear Time (PELT) and Binary Segmentation (BinSeg), both of which identify significant changes in the mean and variance of a data series. Change points were detected based on the unscaled, original series, with special interest in known historical events, such as the introduction of the Euro.

Grau Dataset: Figure 13 displays the detected changepoints in the petrol prices from the *grau* dataset. A critical reference point is the introduction of the Euro, marked with a vertical dashed line. Both PELT and BinSeg algorithms are not able to detect a changepoint. Although both algorithms cahngepoint look closer to the Euro introduced year, but still cannot have a final interpretation that they detect the changepoint where Euro was introduced. Additional changepoints were identified by both algorithms throughout the series, particularly during periods of price increase and stabilization phases (e.g., after 2004 and after 2011).

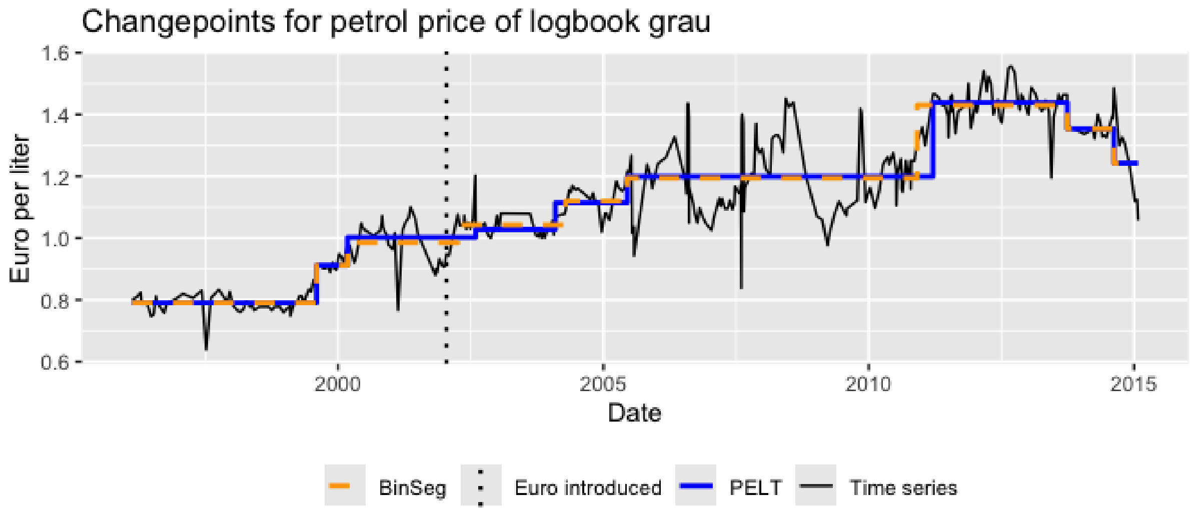

Figure 13: Changepoint detection in petrol prices (Grau dataset).

Karriert Dataset: In the karriert dataset (Figure 14) both methods not able to detect a changepoint corresponding to the Euro introduction. Subsequent changepoints were again clustered around rising price levels and plateaus, which likely reflect international oil price volatility, taxation policy updates, or fuel market adjustments.

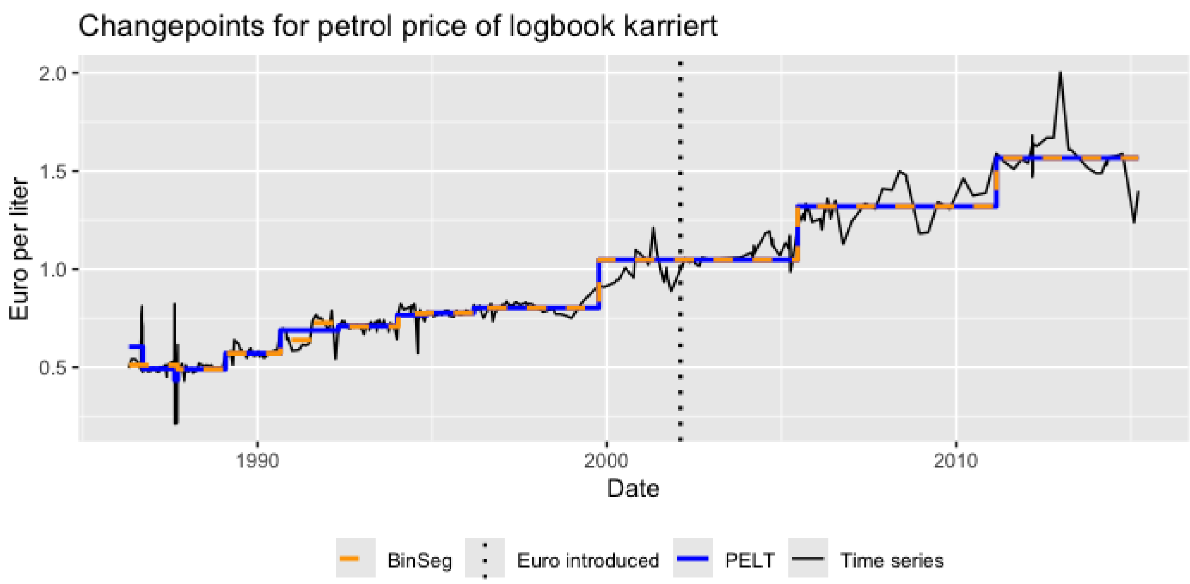

Figure 14: Changepoint detection in petrol prices (Karriert dataset).

4.8 Trend Detection in Petrol Prices

To investigate long-term trends in petrol prices, a trend analysis was performed for both the *grau* and *karriert* datasets. The original price series was assessed without outlier filtering to preserve the full historical price structure. LOESS (locally estimated scatterplot smoothing) was applied to each time series to visualize gradual changes in fuel prices over time.

While visual inspection through LOESS curves provides an intuitive understanding of price movements, a formal statistical test was necessary to confirm the presence of monotonic trends. For this purpose, the Modified Mann–Kendall test was employed. This non-parametric test accounts for autocorrelation in time series data and evaluates whether a significant upward or downward trend is present.

The results of the test are presented in Table 3. Both datasets showed statistically significant trends, with very small p-values indicating strong evidence against the null hypothesis of no trend. Specifically, the **grau** dataset returned a p-value of 1.7500×10^{-17}, and the **karriert** dataset returned a p-value of 7.296×10^{-6}. These results confirm the presence of significant upward trends in petrol prices for both datasets over the analyzed period.

Table 3: Results of Modified Mann–Kendall Test for Petrol Prices.

Dataset	**p-value**	**Significant**	**Trend Direction**
Grau	1.7500×10^{-17}	Yes	Increasing
Karriert	7.296×10^{-6}	Yes	Increasing

Graphical Representation. Figures 15 and 16 provide LOESS-smoothed trend curves superimposed on the raw time series data. The visualizations support the statistical findings by illustrating persistent upward movement in prices across both time series.

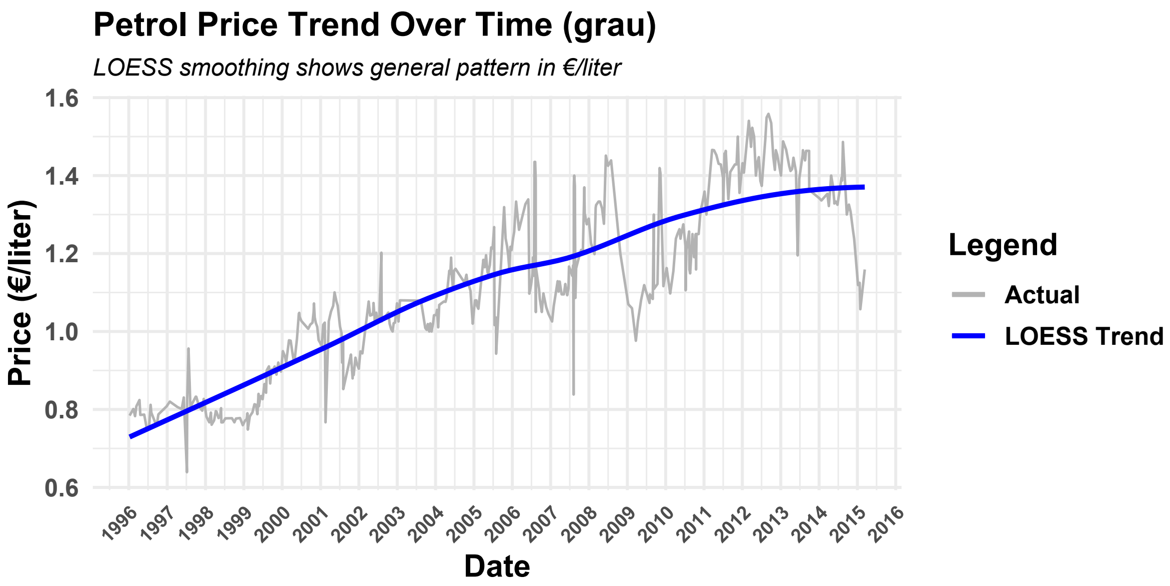

Figure 15: LOESS-smoothed petrol price trend over time for *Grau*.

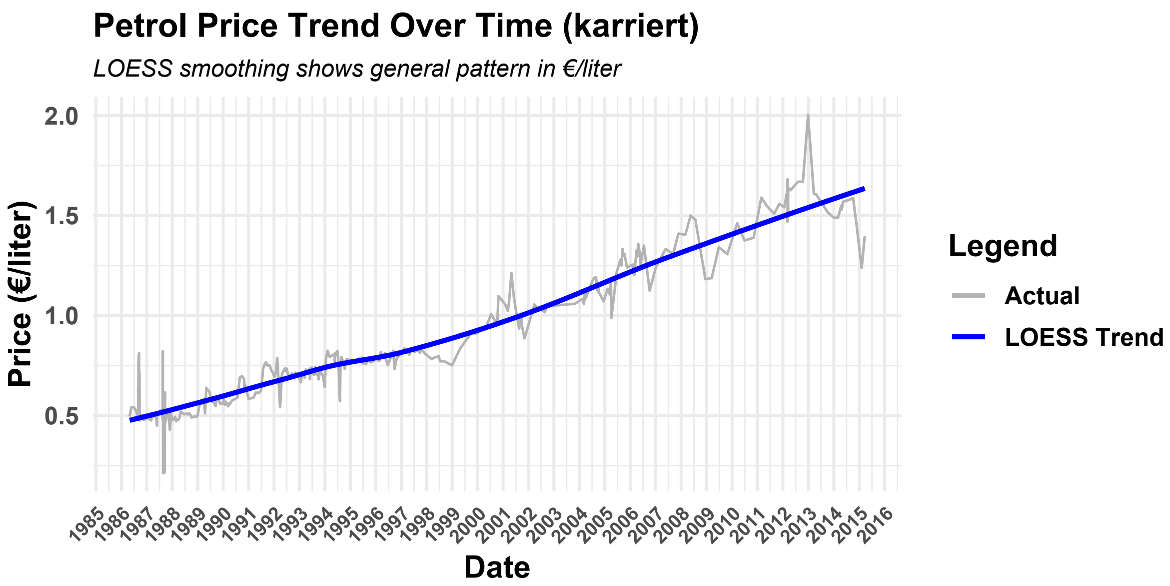

Figure 16: LOESS-smoothed petrol price trend over time for *Karriert*.

Interpretation. The strong statistical significance from trend test and the visual support from LOESS curve indicate that petrol prices have risen steadily over time in both datasets. These consistent upward trends reflect broader fuel price inflation and market changes and should be accounted for in any modeling of cost, consumption, or efficiency.

4.8.1 Autocorrelation Structure in Petrol Prices

To assess temporal dependencies in the evolution of petrol prices, autocorrelation functions (ACFs) were computed for both the *grau* and *karriert* price series. The ACF plots in Figures 17 and 18 display the correlation of the price series with lagged versions of itself.

In Figure 17, the **grau** petrol price series exhibits strong and persistent autocorrelation at multiple lags, beginning with a value near 1.0 at lag 1. The slow decay across successive lags indicates long memory and high temporal dependence in the series, which aligns with the trend detected earlier using the Modified Mann-Kendall test. The autocorrelations lie mostly outside the confidence bounds (indicated by dashed horizontal lines), suggesting statistical significance.

Similarly, the ACF in Figure 18 for the **karriert** petrol price series shows a high autocorrelation at lag 1 and gradual decay, though slightly less persistent than *grau*. This confirms the presence of temporal structure and supports the use of trend models that account for autocorrelation effects.

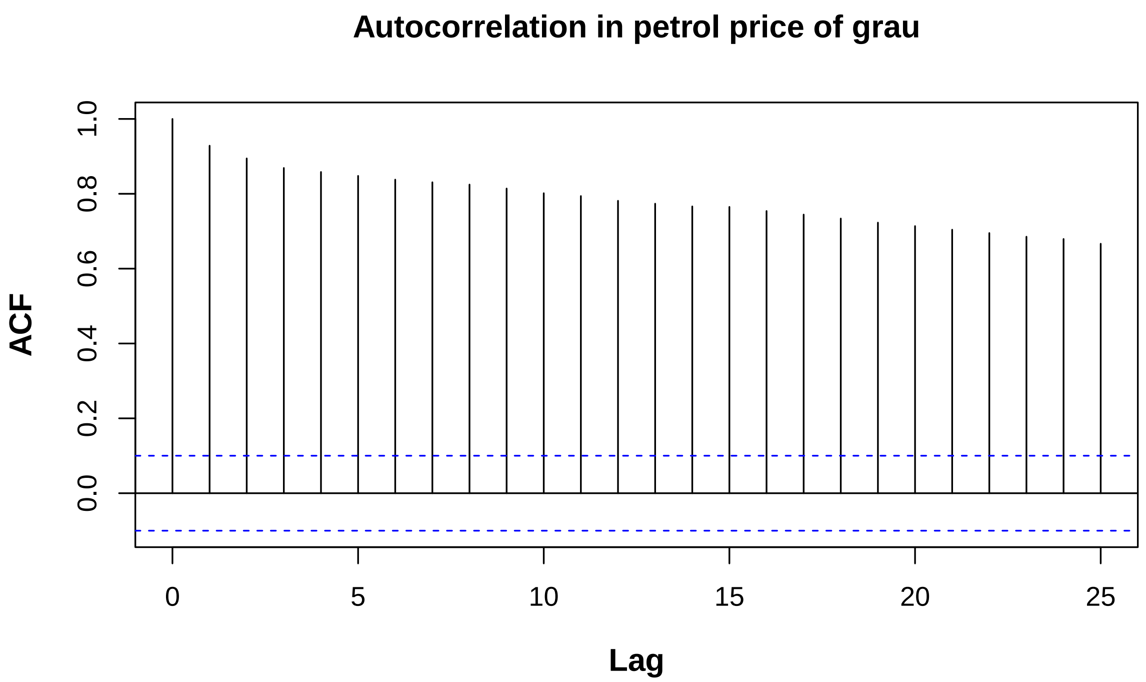

Figure 17: Autocorrelation in petrol price of *Grau*. Dashed lines indicate 95% confidence bounds.

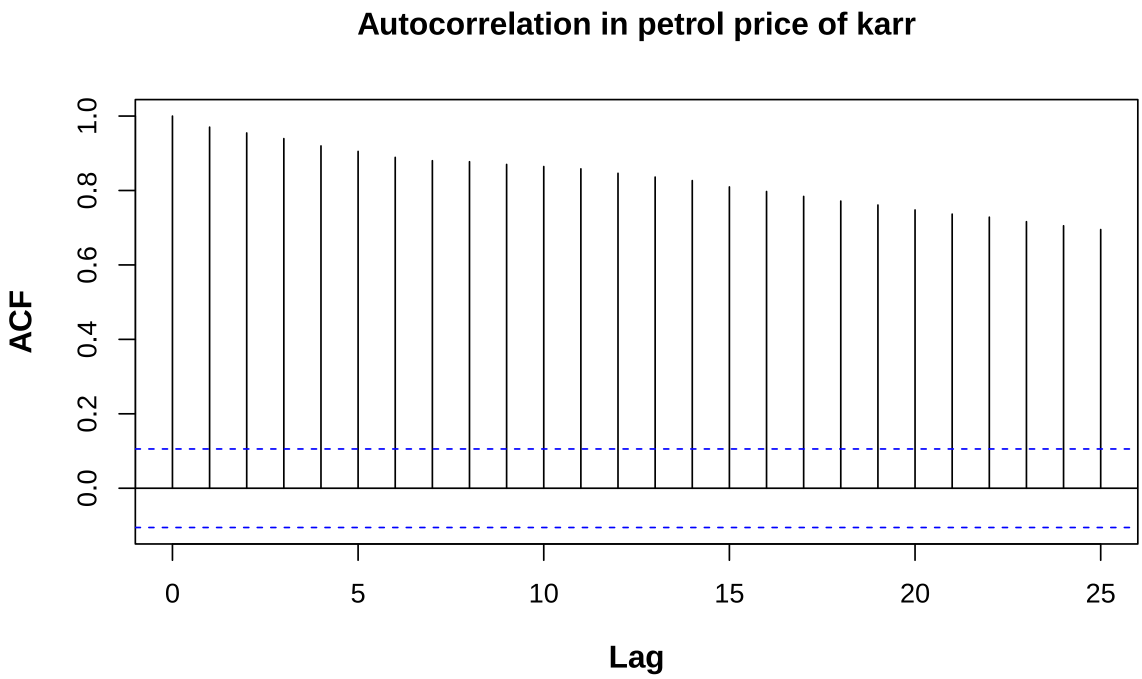

Figure 18: Autocorrelation in petrol price of *Karriert* (karr). Dashed lines indicate 95% confidence bounds.

5 Summary

This project set out to explore long-term trends and structural changes in petrol consumption and fuel prices based on two refueling logs recorded between 1986 and 2015. The goals included detecting consumption and price trends, identifying possible vehicle switches through shifts in consumption behavior, and evaluating the impact of the euro introduction on fuel pricing.

The analysis produced the following key findings:

- A statistically significant downward trend in petrol consumption was observed in the *grau* dataset (Mann-Kendall test $p < 0.084$), suggesting improvements in fuel efficiency or a behavioral/vehicle change.
- In contrast, the *karriert* dataset showed no statistically significant consumption trend ($p = 0.19$), though change point detection methods (e.g., Binary Segmentation) revealed a structural change around 1999.
- Petrol prices, standardized in EUR per liter, exhibited clear upward trends in both *grau* and *karriert* datasets. Mann-Kendall tests confirmed statistically significant positive trends ($p \approx 1.75 \times 10^{-17}$ for *grau*, $p \approx 7.29 \times 10^{-6}$ for *karriert*).
- However, despite the transition from Deutsche Mark to Euro in January 2002, no structural break in price series was detected using the PELT or BinSeg algorithms.
- A clear indication of a vehicle change was identified in the *grau* dataset, implying that two different cars were used over the observation period.
- No such vehicle change could be detected in the *karriert* dataset, indicating consistent usage of the same vehicle.

These results suggest gradual increases in petrol costs over time, improved fuel efficiency in at least one vehicle. However, there is no strong evidence to suggest that the Euro introduction caused a price spike in fuel cost.

Caution is advised when interpreting these findings. Despite careful preprocessing, the data originates from hand-recorded logs and may contain residual measurement or entry errors. The temporal resolution is irregular, and some price conversions rely on approximate historical exchange rates.

Future work could extend this analysis in several directions:

- Incorporate weather or seasonal effects to examine external influences on consumption.
- Apply advanced time series modeling (e.g., ARIMA, Bayesian changepoint models) to refine inference.
- Validate findings using official national fuel pricing datasets or vehicle inspection records.

Overall, this case study demonstrates the importance of rigorous statistical preprocessing and changepoint analysis in deriving insights from long-term, manually collected observational data.

References

Cleveland, William S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". In: *Journal of the American Statistical Association* 74.368, pp. 829–836. DOI: 10.2307/2286407.

European Central Bank (2024). *Euro Reference Exchange Rates*. Accessed July 15, 2025. URL: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates.

Killick, Rebecca (2024). *cpt.meanvar: Identifying Changes in Mean and Variance*. R function in the "changepoint" package version 2.3. URL: https://search.r-project.org/CRAN/refmans/changepoint/html/cpt.meanvar.html.

Killick, Rebecca, Paul Fearnhead, and Idris Eckley (2014). *changepoint: An R Package for Changepoint Analysis*. R package version 2.2.4. URL: https://cran.r-project.org/package=changepoint.

Killick, Rebecca, Paul Fearnhead, and Idris A. Eckley (2012). "Optimal Detection of Changepoints with a Linear Computational Cost". In: *Journal of the American Statistical Association* 107.500, pp. 1590–1598.

kursna-lista.com (2024). *Historical Serbian Dinar Rates*. Accessed July 15, 2025. URL: https://www.kursna-lista.com/en/historical-serbian-dinar-rates.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. 4th ed. Statistics and Computing. New York: Springer. ISBN: 978-0-387-95457-8.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

– (2022). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.10. URL: https://CRAN.R-project.org/package=dplyr.

Yue, Sheng and Chunyuan Wang (2004). "The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series". In: *Water Resources Management* 18.3, pp. 201–218. DOI: 10.1023/B:WARM.0000043140.61082.60.

Zeileis, Achim and Gabor Grothendieck (2005). *zoo: S3 Infrastructure for Regular and Irregular Time Series*. R package version 1.8-12. URL: https://cran.r-project.org/package=zoo.

A Currency Conversion Details

A.1 Currency Codes

Table 4: Currency Codes used in the Report

Code	Currency
ATS	Austrian Schilling
CZK	Czech Crown
DEM	Deutsche Mark
DKK	Danish Krone
EUR	Euro
GBP	British Pound
GRD	Greek Drachma
HUF	Hungarian Forint
IEP	Irish Pound
ISK	Icelandic Króna
ITL	Italian Lira
NLG	Dutch Guilder
NOK	Norwegian Krone
PLN	Polish Zloty
YUD	Yugoslav Dinar

A.2 Conversion Rates to Euro

Fixed and specific dates exchange rates used to convert all prices into Euro:

Euro Conversion Rates

- **ATS**: 1 ATS = 0.07267 EUR
- **CZK**: 2001-02-17: 1 CZK = 0.0295 EUR
- **DEM**: 1 DEM = 0.51129 EUR
- **DKK (pre-Euro)**: 1 DKK = 0.1342 EUR (from 1999-01-04)
- **DKK**: 2007-08-25/26: 1 DKK = 0.134 EUR
- **GBP (pre-Euro)**: 1 GBP = 1.41 EUR (from 1999-01-04)
- **GRD**: 1 GRD = 0.00293 EUR
- **HUF (pre-Euro)**: 1 HUF = 0.003976 EUR (from 1999-01-04)
- **IEP**: 1 IEP = 1.27 EUR
- **ISK**:
  - ▷ 2007-08-10/11/20: 1 ISK = 0.0110 EUR
  - ▷ 2007-08-14: 1 ISK = 0.0112 EUR
- **ITL**: 1 ITL = 0.00051646 EUR
- **NLG**: 1 NLG = 0.454 EUR
- **NOK**:
  - ▷ 2006-07-31, 2006-08-07: 1 NOK = 0.127 EUR
  - ▷ 2006-08-12/13, 2007-08-04/08: 1 NOK = 0.126 EUR
- **PLN**:
  - ▷ 2005-07-16: 1 PLN = 0.242 EUR
  - ▷ 2005-07-30/31: 1 PLN = 0.246 EUR
  - ▷ 2010-07-10/11: 1 PLN = 0.2457 EUR
- **YUD (via DEM)**:
  - ▷ 1987-08-18: 1 YUD = 0.001266322 EUR
  - ▷ 1987-08-19: 1 YUD = 0.001263786 EUR
  - ▷ 1987-08-20: 1 YUD = 0.001256241 EUR
  - ▷ 1987-08-21: 1 YUD = 0.001248754 EUR
  - ▷ 1987-09-16: 1 YUD = 0.001108127 EUR