Forecasting Monthly Car Manufacturing in Spain

Lucas Viner and Oriol Parent
GCED
AD
Academic Year 2024–2025

Contents

1 Introduction

Forecasting plays a pivotal role in economic planning and industrial management, particularly in sectors with high capital investment and logistical complexity, such as automotive manufacturing. The objective of this report is to develop a forecasting model for the monthly production of passenger cars in Spain, applying the Box-Jenkins methodology using ARIMA models. This statistical framework offers a rigorous and widely accepted approach for analyzing time-dependent data, particularly when the objective is to understand underlying structures and produce reliable short- to medium-term forecasts.

This project is conducted within the scope of the Time Series Analysis course, and it aims to deepen both the theoretical understanding and practical implementation of ARIMA models. Beyond model fitting, special emphasis is placed on the diagnostic evaluation of residuals, the assessment of forecasting accuracy, and the integration of outlier detection techniques. The latter is crucial when working with real-world economic data, which often contains irregularities due to policy shifts, economic shocks, or extraordinary events such as pandemics or industrial strikes.

The final objective is not merely to identify a statistical model that fits the historical data well, but to construct a model that is robust, interpretable, and capable of generalizing to future periods with reasonable confidence. This aligns with the broader goals of time series modeling in applied research and policy contexts, where accuracy and interpretability must coexist.

2 Basic description of the dataset

The dataset under analysis in this report is referred to as TURISMOS, and it captures the monthly number of passenger cars manufactured in Spain, recorded in thousands of units. The source of the data is the Ministry of Industry, Energy, and Tourism of Spain (Ministerio de Industria, Energía y Turismo), specifically extracted from the statistical series titled Estadística de Fabricación de Vehículos.

Each observation in the series corresponds to the number of cars produced in a specific month, and the dataset is structured as a single-variable time series with a fixed monthly frequency. This structure lends itself well to ARIMA modeling, particularly because the method assumes equally spaced time intervals and is designed to capture autocorrelations and seasonality.

The preliminary exploration of the series reveals notable variation in production volume over time, likely reflecting both seasonal effects (for example, plant shutdowns in August or increased production before year-end sales periods) and structural shifts (the 2008 financial crisis). Understanding and accounting for these patterns is essential to producing reliable forecasts and drawing meaningful conclusions from the analysis.

In summary, the TURISMOS dataset offers an ideal foundation for applying time series modeling techniques, both due to its economic relevance and its statistical properties.

3 Summary of the methodology used for data analysis

The methodology employed in this project follows the classical Box-Jenkins framework for time series modeling, enhanced by modern techniques for outlier detection and model validation. This structured approach is particularly well-suited for univariate time series data with temporal dependencies and seasonal fluctuations, as is the case with the monthly car production data analyzed in this study.

The analysis begins with data visualization and preprocessing, where the time series is plotted to identify any apparent trends, seasonality, or structural breaks. Basic descriptive statistics are computed, and potential data issues such as missing values or sudden shifts are investigated. Since the raw time series typically exhibits non-stationary behavior—a prerequisite violation for ARIMA modeling—transformations such as logarithmic scaling or differencing are applied to stabilize the mean and variance of the series.

Once the series is transformed into a stationary form, the identification phase involves analyzing the auto-correlation function (ACF) and partial autocorrelation function (PACF) plots. These graphical tools help suggest candidate ARIMA models by revealing the underlying autoregressive (AR) and moving average (MA) structures. At least two plausible models are proposed based on the observed decay patterns in the ACF and PACF, and their theoretical implications.

The estimation phase is carried out using the R statistical software, where the selected ARIMA models are fitted to the training portion of the dataset. Maximum likelihood estimation is used to estimate the model parameters, and the performance of each model is evaluated based on goodness-of-fit metrics such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Next, the validation phase involves a thorough diagnostic check of the residuals. Residuals should behave as white noise—uncorrelated, homoscedastic, and normally distributed. To confirm this, we examine standardized residual plots, the ACF of residuals, and conduct formal tests such as the Ljung-Box test. Additionally, the stability of the models is verified by ensuring that the estimated coefficients satisfy stationarity and invertibility conditions. For out-of-sample validation, the final 12 observations are withheld from model fitting and used to evaluate the forecasting accuracy of each model.

An important extension in this analysis is the incorporation of automatic outlier detection, which identifies and adjusts for atypical observations that could distort model estimation and prediction. Using established algorithms available in R, the time series is scanned for additive, level shift, and temporary change outliers. Detected anomalies are interpreted in light of known historical events when possible, and a new ARIMA model is refitted to the adjusted series.

Finally, the selected model is used to generate 12-month ahead forecasts, including confidence intervals to account for prediction uncertainty. A comparison is made between forecasts obtained from the original and outlier-adjusted series to assess the impact of anomalies on forecast quality.

Overall, this methodology ensures that the final model is not only statistically sound but also robust to real-world irregularities, providing a credible tool for understanding and anticipating trends in the Spanish car manufacturing sector.

4 Results and interpretation

Let's now begin with the results and interpretation.

4.1 Identification

4.1.1 Determine the transformations needed to make the series stationary.

We start reading our original time serie and we plot it.

The plot shows the **monthly production of passenger cars in Spain** from January 1994 onward. A few key observations from the raw series:

- **Strong seasonality**: There's a recurring dip, likely due to seasonal shutdowns (August holidays).

- **Volatility**: Some periods exhibit higher variance.

- **Non-stationarity**: The mean level appears to fluctuate over time — this will need to be addressed before fitting ARIMA models.

In order to assess whether the variance of the series was stable over time, we will generate a year-wise boxplot.

This graphical representation grouped the data by year and highlighted the distributional spread of monthly values within each period. The resulting plot clearly showed substantial differences in the heights of the boxes and the interquartile ranges across different years, indicating that the variance is **not constant** over time. On top of that, we can clearly observe the presence of outliers.

The violation of the constant variance assumption (heteroskedasticity) implies that a transformation is required before model identification can proceed. In this context, a **variance-stabilizing transformation**, such as the **logarithmic transformation**, is commonly applied. This transformation compresses higher values more than lower ones, helping to normalize the variance and making the series more amenable to ARIMA modeling.

The plot of lnserie reveals that, although the variance appears more controlled, the **mean level is not constant** throughout the observation period. There are noticeable trends and long-term fluctuations, indicating **non-stationarity in the mean**. This violates one of the key assumptions of ARIMA modeling: that the underlying data-generating process has constant statistical properties over time.

Given this observation, it becomes necessary to apply a **regular (non-seasonal) first-order difference** to the log-transformed series. This operation removes the linear trend and stabilizes the mean of the series, allowing for clearer identification of autoregressive and moving average components.

Now, let's move on to studying the seasonality of our time series.

To further investigate the seasonality of the transformed series, a seasonal decomposition was visually inspected. This plot illustrates the average seasonal pattern of the time series across the twelve months of the year, providing valuable insight into recurring intra-annual behavior.

The monthly plot reveals a relatively consistent mean level throughout the year, suggesting that the overall trend component may be stable. However, there are **notable deviations**, particularly in **August and Setember**, where production sharply declines. This drop likely reflects planned factory shutdowns or reduced industrial activity during the summer holiday period. A smaller but consistent dip is also observed in **December**, possibly associated with year-end slowdowns or reduced operating schedules.

Despite the seemingly stable trend, these pronounced and recurring seasonal effects indicate that the series is **not stationary in its seasonal component**.

Consequently, the series requires **seasonal differencing** to remove this deterministic seasonal component and achieve stationarity. This step is essential for isolating the stochastic structure of the time series, allowing for reliable model identification and parameter estimation in the next phases of the Box-Jenkins methodology.

If we apply the monthplot to the new series, we can now see how we have gotten rid of the seasonal component.

After ensuring that the series is visually stationary in both the mean and seasonal components through differencing, the next step consists of quantitatively validating the effectiveness of the transformations. To achieve this, a **variance comparison** is performed among the different transformed versions of the series.

Specifically, the variances of the following series are computed:

- `var(lnserie)`: the log-transformed series without differencing,

- `var(d1d12lnserie)`: the series after applying both a regular and a seasonal difference,

- `var(diff(d1d12lnserie))`: the series after an additional regular differencing step.

The objective of this comparison is to identify which transformation yields the **lowest variance**, as lower variance typically indicates a more stable and stationary process.

```
##             V1
## V1 0.1693238

##              V1
## V1 0.04099308

##             V1
## V1 0.1247695
```

The results of the variance calculations clearly show that the series `d1d12lnserie`, obtained after applying one regular and one seasonal difference, achieves the **lowest variance** among the considered transformations. This outcome confirms that the combination of both types of differentiating is necessary and sufficient to render the series stationary in both its mean and seasonal structures.

Consequently, the analysis proceeds using the `d1d12lnserie` series for the identification of candidate ARIMA models through the examination of its autocorrelation and partial autocorrelation functions.

4.1.2 Analyze the ACF and PACF of the series to identify at least two possible models. Explain the graphical characteristics used to support your selection.

After applying both regular and seasonal differencing, the autocorrelation structure of the transformed series is examined through the **Autocorrelation Function (ACF)** and the **Partial Autocorrelation Function (PACF)** plots.

Now, we have observed that the series exhibits a regular dependence (a relationship with the immediately preceding observations), but also a seasonal dependence (to predict the number of cars in August, it makes sense to consider not only July, June, and May of the same year, but also the values recorded in August of previous years).

Therefore, we will proceed to work with an ARIMA model characterized by six parameters: three for the regular dependence and three for the seasonal dependence: ARIMA(p, d, q)(P, D, Q)s.

To propose the components of the seasonal dependence (denoted by uppercase letters), the key will be to examine the ACF/PACF plots, focusing only on the lags that are multiples of the seasonality period (in R, we highlight these lags by marking them in red).

We begin by determining the seasonal components. First, let us examine the ACF plot. As we know, the objective is to identify in which of the two plots (ACF or PACF) a finite number of significant lags appears. The ACF plot displays a very rapid decay after lag 1, with only the first lag being significantly different from zero, and subsequent autocorrelations falling within the confidence bands. This suggests that the autocorrelation structure is short-term and does not persist across longer lags, which is a typical characteristic of a moving average (MA) process of low order. The significant spike at lag 1 points toward the presence of an **MA(1) component**.

Similarly, the PACF plot shows a significant spike at lag 1, followed by nonsignificant values for higher lags. This pattern is associated with an autoregressive (AR) process of low order, particularly an **AR(1) model.**

Secondly, we must determine the regular component, which relates the present value to the immediate past. To identify models for the regular part, we will focus exclusively on the first few lags (the first four or five), as lags close to multiples of the seasonality period may act as satellite: that is, non-representative lags.

Let's start by analyzing the ACF plot. We can clearly observe that only one lag falls outside the confidence interval within the region of interest, since the next lag that exceeds the interval is very close to the seasonal lag and is clearly biased by it. Therefore, only one lag is considered significant. Since the finite number of lags is determined based on the ACF, we conclude that the model corresponds to an MA(q) process, where q is the last significant lag. In this case, we identify an **MA(1) process.**

If we apply the same method to the PACF plot, we observe that two lags fall outside the confidence interval. Similarly to the ACF case, the subsequent lags that exceed the interval are already very close to the next seasonal lag. Therefore, it is reasonable to propose an **AR(2) model.**

So, we propose the following 4 models:

- ARIMA(2,1,0)(1,1,0)12
- ARIMA(0,1,1)(1,1,0)12
- ARIMA(2,1,0)(0,1,1)12
- ARIMA(0,1,1)(0,1,1)12

4.2   Estimation

4.2.1   Use R to estimate two of the identified models

After identifying possible model structures based on the ACF and PACF analysis, the next step consists of estimating the parameters of two candidate ARIMA models using the R statistical environment. The selected models are:

- ARIMA(0,1,1)(1,1,0)12
- ARIMA(0,1,1)(0,1,1)12

The choice of these models is motivated by a preference for **parsimony**, meaning that models with fewer parameters are prioritized, provided they offer an adequate fit. Parsimonious models are generally preferred because they tend to have better forecasting properties and avoid overfitting. Thus, among the identified possibilities, the two models with the simplest structures were selected for estimation.

Lets start by estimating the first of the two models: **ARIMA(0,1,1)(1,1,0)12.**

The estimation is performed using the `arima` function in R, which employs maximum likelihood estimation to determine the model coefficients. Moreover, it is important to remember that we will first apply the command to the series Wt in order to also obtain the estimation of the mean (the intercept) and assess whether it is statistically significant.

```
##
## Call:
## arima(x = d1d12lnserie, order = c(0, 0, 1), seasonal = list(order = c(1, 0,
##      0), period = 12))
##
## Coefficients:
##           ma1     sar1  intercept
##       -0.6431  -0.5106    -0.0005
## s.e.   0.0438   0.0522     0.0021
##
## sigma^2 estimated as 0.02101:  log likelihood = 144.97,  aic = -281.93
```

The first step is to examine the intercept and determine whether it is statistically significant. We will do so by conducting the following hypothesis test

$H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ where

$$\hat{t} = \frac{\hat{\mu} - 0}{S\mu}$$

and $S\mu$ represents the standard error.

The decision rule will be that if $\|\hat{t}\| > 2 \Rightarrow H_1$.

In this case, we observe that the t-ratio of the intercept value is close to zero. Therefore, we can conclude that the intercept is not significant, and we will use the ARIMA model that also estimates the necessary transformations.

```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(1, 1, 0), period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.6429  -0.5107
## s.e.   0.0439   0.0522
##
## sigma^2 estimated as 0.02102:  log likelihood = 144.93,  aic = -283.86
```

The outputs obtained from these estimations include the estimated coefficients, their standard errors, and information criteria such as the **Akaike Information Criterion (AIC)**, which are key tools for model comparison.

The final expression of the estimated model is:

$$(1 + 0.5107B)(1 - B)(1 - B^{12})X_t = (1 - 0.6429B^{12})Z_t$$

We now proceed with the following model: **ARIMA(0,1,1)(0,1,1)12**. As before, we first check whether the intercept is statistically significant. In this way, we will set both differencing parameters (d and D) to 0 and use the series Wt instead.

```
##
## Call:
## arima(x = d1d12lnserie, order = c(0, 0, 1), seasonal = list(order = c(0, 0,
```

```
##      1), period = 12))
##
## Coefficients:
##           ma1      sma1   intercept
##       -0.6556   -0.7328     -2e-04
## s.e.   0.0456    0.0442      8e-04
##
## sigma^2 estimated as 0.01778:  log likelihood = 166.16,  aic = -324.32
```

Once again, we observe that the ratio between the intercept values is close to zero. Therefore, we can use the ARIMA model that also estimates the necessary transformations.

```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))
##
## Coefficients:
##           ma1      sma1
##       -0.6555   -0.7329
## s.e.   0.0456    0.0442
##
## sigma^2 estimated as 0.01778:  log likelihood = 166.13,  aic = -326.25
```

The final expression of the estimated model is:

$$(1 - B)(1 - B^{12})X_t = (1 - 0.6555B)(1 - 0.7329B^{12})Z_t$$

Furthermore, since we have obtained a lower AIC value, we can conclude that this model is an improvement over the previous one.

4.3   Validation

4.3.1   Perform a complete residual analysis, justifying the assumptions based on the corresponding graphical results.

A crucial step in the modeling process is the analysis of the residuals generated by the fitted models. The residual analysis allows for a verification of the fundamental assumptions underlying ARIMA models, namely:

- Homoscedasticity: The variance of the residuals should be constant over time. Residuals should not show any pattern of increasing or decreasing dispersion.

- Normality: The residuals should approximately follow a normal distribution. This assumption is important for the validity of confidence intervals and hypothesis tests associated with the model.

- Independence: The residuals should not exhibit autocorrelation; they must behave as white noise.

To begin this analysis, the residuals from the first model **ARIMA(0,1,1)(1,1,0)12**, are extracted using the `resid()` function in R. A time plot of these residuals is generated.

4.3.1.1 Homoscedasticity To rigorously assess the assumption of homoscedasticity, a more sensitive graphical technique is employed. Specifically, the square root of the absolute value of the residuals is plotted against time using the `scatter.smooth()` function in R, with a locally weighted regression line (LOESS) superimposed.

In the resulting plot, the behavior of the red smoothing curve is examined. Ideally, if the variance were constant over time, the smoothed line would be approximately horizontal, indicating that the spread of the residuals does not vary across the observation period.

However, the results indicate that the smoothing line is not perfectly flat. A slight upward trend is observed, particularly at the beginning of the series. This suggests that the variance of the residuals tends to increase slightly over time. Furthermore, the initial time plot of the residuals already showed several points falling outside the ± 3 standard deviation bands, further hinting at the presence of outliers and mild heteroscedasticity.

Based on these observations, it is concluded that the assumption of constant variance is **not fully satisfied**. While the deviations are not extreme, the evidence suggests that the residuals exhibit a slight form of heteroscedasticity, which could potentially affect the efficiency of the model's forecasts and confidence intervals.

4.3.1.2 Normality To evaluate the assumption of normality of the residuals, both graphical and statistical methods are employed.

First, a **normal Q-Q plot** of the residuals is generated.

In the resulting plot, the theoretical quantiles from a normal distribution are compared to the empirical quantiles of the residuals. Ideally, if the residuals followed a normal distribution, the points would align closely along the red reference line. However, it is observed that the residuals deviate considerably from the reference line, particularly in the tails. The presence of numerous points far from the line at both extremes indicates **outliers** and **heavy tails**, suggesting a deviation from normality. Additionally, there is noticeable volatility around the central part of the plot, where the points are dense but still deviate from the line, reinforcing the conclusion that the normality assumption is not satisfied.

To further investigate this, a **histogram** of the residuals is constructed and compared to the theoretical normal distribution curve.

The histogram confirms the previous findings. Although the general shape is roughly symmetric, it is evident that the empirical distribution does not perfectly match the theoretical normal curve. Outliers are again detected, and the residuals do not tightly follow the normal bell-shaped pattern, particularly at the extremes.

Finally, to formally test the normality hypothesis, the **Shapiro-Wilk test** is conducted:

```
## 
##  Shapiro-Wilk normality test
## 
## data:  resi1
## W = 0.98017, p-value = 0.000359
```

The test yields a **p-value significantly smaller than 0.05**, leading to the rejection of the null hypothesis of normality at a 95% confidence level. Therefore, based on the combined evidence from the Q-Q plot, the histogram, and the Shapiro-Wilk test, it is concluded that the residuals **do not follow a normal distribution**.

4.3.1.3 Independence The final assumption to verify in the residual analysis concerns **independence**, meaning that the residuals should not exhibit any autocorrelation structure and should behave as white noise.

To investigate this, the **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF)** of the residuals are plotted. Ideally, in the absence of autocorrelation, all points in the ACF and PACF plots should lie within the 95% confidence bounds, suggesting that the residuals are uncorrelated at all lags.

Upon visual inspection of the ACF and PACF plots, it is observed that although most of the lags fall within the confidence interval, a few lags are either very close to the boundary or slightly exceed it. While these small deviations are not conclusive by themselves, they raise suspicions regarding the complete independence of the residuals. In particular, the presence of several lags near the threshold suggests that some degree of autocorrelation might persist.

To formally verify this, the **Ljung-Box test** is applied.

The Ljung-Box test evaluates whether groups of autocorrelations are jointly zero. In the resulting plot, the **p-values** for all lags up to lag 72 are systematically analyzed. It is clearly observed that almost all p-values are **below 0.05**, indicating that the null hypothesis of residual independence is rejected at the 5% significance level.

Therefore, based on both the graphical analysis and the formal Ljung-Box test, it is concluded that the residuals **do not satisfy the independence assumption**. Some autocorrelation remains in the residuals, which may affect the adequacy of the model for forecasting purposes.

To continue with this analysis, as we have done before, the residuals from the second model **ARIMA(0,1,1)(0,1,1)12**, are extracted using the `resid()` function in R. A time plot of these residuals is generated.

We now plot the square root of the absolute value of the residuals with a smooth fit.

We now analyze the obtained results. In the residuals plot, we observe several outliers falling outside the confidence interval. Moreover, when examining the plot of the square root of the absolute value of the residuals, we do not obtain the desired result: a nearly horizontal red line. In fact, we notice a slight upward trend, particularly at the beginning of the plot. Based on this, we can conclude that the variance is not constant

If we analyze the normality of the residuals, we can do so through a normality plot as well as by using the *Shapiro-Wilk test.*

We will begin first with the graphical analysis.

We can clearly see that our residuals do not follow the red normality line. If we focus on the tails, we observe the presence of outliers—points that are very distant and sparsely distributed. Additionally, there is a certain degree of volatility, as indicated by clusters of points that deviate significantly from the normality line, especially in the central part of the distribution. This pattern suggests that the residuals do not conform well to a normal distribution, which may impact the reliability of statistical inference and confidence intervals based on this model.

Once again, we confirm the hypotheses suggested by the previous plot. Although less clearly than with the Q-Q plot, we still detect the presence of outliers, indicating that our residuals do not follow a normal distribution.

Finally, we proceed to apply the Shapiro-Wilk test to the residuals.

```
##
##  Shapiro-Wilk normality test
##
## data:  resi2
## W = 0.96921, p-value = 5.045e-06
```

Analyzing the p-value obtained, we see that it is much smaller than 0.05. Therefore, based on all the points mentioned previously in this section and the results from the Shapiro-Wilk test, we can conclude that the residuals do not follow a normal distribution.

Now, we move on to study the independence of our residuals. To do this, we will examine their ACF and PACF plots and apply the Ljung-Box test.

If we analyze these plots, we see that although some lags are very close to the limit of the confidence interval, or even slightly exceed it, we cannot clearly detect non-independence in the residuals. Therefore, to resolve any doubts, we proceed to apply the Ljung-Box test.

If we analyze this plot, we can clearly see that all but the first two p-values of our test are below 0.05. Therefore, we can conclude that these residuals are **not independent.**

4.3.2 Include details on model expressions, such as infinite AR and MA representations, stationarity and invertibility conditions, and model fit metrics.

4.3.2.1 Model expressions The two estimated models are expressed as follows:

- **Model 1**: ARIMA(0,1,1)(1,1,0)12
  The model structure can be written as:
  $$(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12}) \ln(X_t) = (1 + \theta_1 B)Z_t$$

  - θ_1: regular MA(1) coefficient
  - $\Phi1$: seasonal AR(1) coefficient
  - B: backshift operator $(BX_t = X_{t-1})$
  - Z_t: white noise

Let's now give the infinite AR and MA representations:

MA:
$$X_t = \frac{(1 - 0.6429B)}{(1 + 0.5107B^{12})(1 - B)(1 - B^{12})} Z_t$$

AR:

$$Z_t = \frac{(1 + 0.5107B^{12})(1 - B)(1 - B^{12})}{1 - 0.6429B} X_t$$

- **Model 2:** ARIMA(0,1,1)(0,1,1)12
  The model structure can be written as:

  $$(1 - B)(1 - B^{12}) \ln(X_t) = (1 + \theta_1 B)(1 + \Theta_1 B^{12})Z_t$$

  - θ_1: regular MA(1) coefficient
  - Θ_1 seasonal MA(1) coefficient
  - B: backshift operator $(BX_t = X_{t-1})$
  - Z_t: white noise

MA:

$$w_t = Z_t - 0.6555Z_{t-1} - 0.7329Z_{t-12} + 0.4805Z_{t-13} + \cdots$$

AR:

$$Z_t = \frac{1}{(1 - 0.6555B)(1 - 0.7329B^{12})} w_t$$

4.3.2.2 Stationarity and Invertibility Conditions

- **Stationarity**:
  Both models include **first-order regular differencing** and **first-order seasonal differencing**, which are necessary to induce stationarity in the series.
  The original series ln(Xt) is non-stationary, but the differenced series is stationary based on visual and statistical diagnostics.

  Applied regular differencing: $(1 - B)$

  Applied seasonal differencing: $(1 - B^{12})$

Thus, the differenced series is stationary.

- **Invertibility**:
  Invertibility requires that the roots of the moving average polynomials lie outside the unit circle.
  The roots of the MA polynomials $(1 + \theta_1 B)$ and $(1 + \Theta_1 B^{12})$ must lie **outside** the unit circle (i.e., modulus greater than 1).

Model 1:

```
##  [1] 1.057602 1.057602 1.057602 1.057602 1.057602 1.057602 1.057602 1.057602
##  [9] 1.057602 1.057602 1.057602 1.057602
```

```
## [1] 1.555367
```

Model 2:

```
## numeric(0)
```

```
##  [1] 1.026238 1.026238 1.026238 1.026238 1.026238 1.026238 1.026238 1.026238
##  [9] 1.026238 1.026238 1.026238 1.026238 1.525591
```

As all of the roots of both models are greater than 1, we can say they are invertible.

Thus, both models meet the necessary **stationarity** and **invertibility** conditions after differencing.

4.3.2.3 Model fit metrics
The goodness of fit for each model is evaluated using the **Akaike Information Criterion (AIC)** and the **Bayesian Information Criterion (BIC)**:

AIC:

```
## [1] -283.8638
```

```
## [1] -326.2528
```

BIC:

```
## [1] -272.8854
```

```
## [1] -315.2743
```

As it has a lower value of AIC and BIC, we can say the second model has better fit metrics.

4.3.3 Check model stability and assess forecasting accuracy by reserving the last 12 observations for validation.

4.3.3.1 Model stabililty To assess the **stability** of the fitted model, a comparative analysis is conducted. This analysis involves estimating the same model structure twice:

- Once using the full dataset.
- Once using a **truncated dataset**, excluding the last 12 observations.

The idea is to verify whether the model's parameters remain consistent when slightly changing the estimation sample, which is a necessary condition for model stability and reliability in forecasting.

Specifically, the coefficients obtained from the two models are compared by focusing on three key aspects:

1. Magnitude

2. Sign

3. Statistical significance

This approach ensures that the model is not overfitted to specific fluctuations in the dataset and that it maintains a stable structure when forecasting future values.

To implement this comparison, a new model is estimated using the truncated dataset, defined by removing the last 12 months of observations.

We now compare this one with our models.

```
##
## Call:
## arima(x = lnserie_aj, order = c(0, 1, 1), seasonal = list(order = c(1, 1, 0),
##     period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.6376  -0.489
## s.e.   0.0449   0.054
##
## sigma^2 estimated as 0.02129:  log likelihood = 137.22,  aic = -268.44


##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(1, 1, 0), period = 12))
##
## Coefficients:
##           ma1     sar1
##       -0.6429  -0.5107
## s.e.   0.0439   0.0522
##
## sigma^2 estimated as 0.02102:  log likelihood = 144.93,  aic = -283.86


##
## Call:
## arima(x = lnserie_aj, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
```

```
##      period = 12))
##
## Coefficients:
##           ma1      sma1
##       -0.6457   -0.7267
## s.e.   0.0469    0.0444
##
## sigma^2 estimated as 0.01763:  log likelihood = 160.23,  aic = -314.47


##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))
##
## Coefficients:
##           ma1      sma1
##       -0.6555   -0.7329
## s.e.   0.0456    0.0442
##
## sigma^2 estimated as 0.01778:  log likelihood = 166.13,  aic = -326.25
```

After analyzing the results, it is observed that the coefficients remain consistent in all three aspects. The magnitude and sign of the parameters are preserved, and the coefficients continue to be statistically significant even when the last 12 observations are excluded from the estimation process.

Therefore, it can be concluded that both models demonstrate **parameter stability**, confirming that the model structure is not overly sensitive to minor changes in the data sample. This strengthens the reliability of the models for forecasting purposes.

4.3.3.2 Forecasting accuracy To assess the forecasting performance of the proposed models, the last 12 observations of the series are reserved as a **validation set**. The models are trained on the truncated dataset (excluding the last 12 months), and **forecasts** are generated for the validation period.
The goal is to compare the forecasted values to the real observations and evaluate the predictive capability of each model.

The following procedure is followed:

1. The models are estimated using the training dataset (all observations except the last 12 months).

2. Forecasts are generated for the next 12 months.

3. The forecasted values are compared to the real observed values, and forecast errors are computed.

The accuracy of the forecasts is quantified using several standard error metrics:

- **Mean Absolute Error (MAE)**: measures the average magnitude of the errors without considering their direction.
$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |X_t - \hat{y}_t|$$

- **Root Mean Squared Error (RMSE)**: penalizes larger errors more heavily and is sensitive to large deviations.
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (X_t - \hat{y}_t)^2}$$

- **Mean Absolute Percentage Error (MAPE)**: expresses the forecast errors as a percentage of the actual values, allowing for scale-independent interpretation.

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{X_t - \hat{y}_t}{X_t} \right|$$

The model with the **lower** values of MAE, RMSE, and MAPE will be considered **more accurate** for forecasting purposes.

Let's calculate this measures for both models.

```
##
## RMSE =   22.54406

##
## MAE =   17.95891

##
## MAPE =   0.1032013

##
## RMSE =   25.27715

##
## MAE =   20.76572

##
## MAPE =   0.1248739
```

4.3.4 Select the best model for forecasting.

The final model selection is based on a comprehensive evaluation of several criteria, including:

- The model fit metrics, particularly the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC),

- The forecast accuracy measures (MAE, RMSE, MAPE) obtained from the reserved validation set,

- The results of the residual analysis and model stability assessment.

Between the two candidate models, the model with lower AIC and BIC values is favored as it indicates a better balance between goodness of fit and parsimony.

Additionally, the model demonstrating lower forecast error metrics on the validation set is considered more reliable for out-of-sample predictions.

After analyzing the results, Model 2 (mod2) remains the preferred choice for forecasting, primarily due to its superior information criteria values. It has a significantly lower AIC (-326.2528) and BIC (-315.2743) compared to Model 1 (mod1), which indicates a better balance between model fit and complexity.

However, when comparing forecasting accuracy metrics:

Model 1 (mod1) performs better in terms of:

RMSE: 22.54406 (vs. 25.27715 for mod2)

MAE: 17.95891 (vs. 20.76572 for mod2)

MAPE: 0.1032 (vs. 0.1249 for mod2)

These lower error metrics suggest that mod1 provides more accurate point forecasts on the observed data.

Despite this, we prioritize mod2 due to its lower AIC and BIC values, as these criteria assess the overall model quality, including parsimony and generalization capacity. Therefore, Model 2 is chosen for its stronger statistical foundation and expected robustness in out-of-sample forecasting

4.4 Forecasting

4.4.1 Generate long-term forecasts for the 12 months following the last recorded observation, including confidence intervals.

In this task, we aim to generate 12-month ahead forecasts beyond the last recorded observation, including corresponding confidence intervals. The forecasts are based on **Model 2 (mod2)**, which has been selected as the most suitable model for prediction.

We can see that the results obtained are quite good, as the prediction follows a pattern similar to the original series. The forecasted values align closely with the historical data, indicating that the model is capturing the underlying trends and behaviors effectively.

4.5 Calendar Effects

Calendar effects refer to systematic variations in the data caused by differences in the number of working days across months. In the context of car production, these effects are particularly relevant in Spain due to recurring shutdowns in months like August and December. If not corrected, these variations can distort the analysis and lead to misleading conclusions. This section adjusts the series to account for these effects, ensuring that the model captures genuine trends and patterns.

```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12),
##     xreg = data.frame(vEa, vTD))
##
## Coefficients:
##           ma1      sma1      vEa      vTD
##       -0.5473   -0.6712   -0.1395   0.0135
## s.e.   0.0518    0.0443    0.0266   0.0018
##
## sigma^2 estimated as 0.01372:  log likelihood = 204.45,  aic = -398.9


##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12),
##     xreg = data.frame(vEa))
##
## Coefficients:
##           ma1      sma1      vEa
##       -0.6232   -0.7056   -0.1622
## s.e.   0.0471    0.0456    0.0304
##
## sigma^2 estimated as 0.01628:  log likelihood = 179.26,  aic = -350.53
```

```
##
## Call:
## arima(x = lnserie, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12),
##     xreg = data.frame(vTD))
##
## Coefficients:
##           ma1     sma1     vTD
##       -0.5850  -0.6956  0.0145
## s.e.   0.0496   0.0435  0.0019
##
## sigma^2 estimated as 0.01496:  log likelihood = 191.67,  aic = -375.35
```

If we examine the coefficient values of vTD and vEa in the models modTD and modEa, we observe that both are statistically significant. This indicates that the calendar effects—specifically, those related to the number of working days (vTD) and Easter holidays (vEa)—have a real and measurable impact on the behavior of the series.

For this reason, we decide to adjust the original series by removing these calendar effects. This is done by estimating their influence through the coefficients obtained in the respective models, and then subtracting their contribution from the original log-transformed series. The result is a cleaner series, where variations due to calendar structure (such as shorter months or holiday periods) no longer interfere with the analysis.

By doing this, we ensure that the remaining variation in the data more accurately reflects genuine economic or production-related dynamics, rather than artifacts introduced by the calendar. This step is crucial for building a reliable time series model and for interpreting results and forecasts without distortion.

We finally do our model without the calendar effects. We now analyze the residuals of the adjusted model to assess improvements in independence, normality, and homoscedasticity compared to the original model. We also review key forecast accuracy metrics (MAE, MAPE, RMSE, AIC) to evaluate whether correcting for calendar effects and outliers has resulted in a more accurate and reliable model.

We first start studying the normality of our residuals.

If we examine the two plots—the Q-Q plot and the histogram of residuals—we can clearly see that the residuals do not follow a normal distribution. In the Q-Q plot, the points deviate noticeably from the reference line, indicating that the residuals do not align with what we would expect under normality. Similarly, the histogram does not resemble the shape of a normal bell curve, further confirming the lack of normality in the residuals. These visual indicators suggest that the assumption of normally distributed errors is not satisfied in this model. And it seems that it does not improve the orignial model.

Now we are going to study the homocedasticity.

Once again, we find that our adjustment does not improve upon the previous model, as the plot clearly shows. In this case, we would expect to see a straight line, which would indicate homoscedasticity—that is, constant variance of the residuals over time. However, the result we obtain deviates significantly from this expectation, suggesting that the residuals still exhibit heteroscedasticity, and that the model has not fully captured the variance structure of the data.

Finally we are going to see if the independence improves.

We can observe that all the p-values in our test remain above the 5% threshold. This indicates that, after removing the calendar effects, we have successfully achieved independence in the residuals. In other words, there is no significant autocorrelation remaining, which confirms that the adjusted model better captures the time-dependent structure of the data and that the residuals behave more like white noise—one of the key assumptions for a well-specified time series model.

```
## numeric(0)
```

```
## [1] 1.033777 1.033777 1.033777 1.033777 1.033777 1.033777 1.033777 1.033777
## [9] 1.033777 1.033777 1.033777 1.033777 1.827125
```

We can observe that the values of all the roots are greater than 1—specifically, twelve of them are equal to approximately 1.0336, and one stands out with a value of around 1.8212. This confirms that all roots lie outside the unit circle, which is a fundamental condition for invertibility in time series models. Invertibility ensures that the moving average (MA) component of the model is stable and that the model has a valid representation in terms of past observations. Therefore, we can conclude that our model satisfies the invertibility condition, contributing to the overall stability and reliability of the model for forecasting.

```
## [1] -398.8979
```

```
## [1] -380.6005
```

We can see that we improve the values of the AIC and BIC of the orignial model with the calendar effects treatment.

Finally we are going to forecast the observations without taking into account the calendar effects.

If we study the values of the performance metrics we can see that all of them improve, but not too significantlly.

```
## [1] 28.68393
```

```
## [1] 23.41934
```

```
## [1] 0.1686307
```

```
## [1] 0.1365079
```

```
## [1] 141.2264
```

If we study the values of the performance metrics—specifically RMSE, MAE, RMSPE, MAPE, and the average confidence interval width—we can see that all of them show improvement compared to the original model, although the gains are not particularly large.

For instance, the Root Mean Square Error (RMSE) is 28.60 and the Mean Absolute Error (MAE) is 23.06, indicating a modest reduction in absolute and squared prediction errors. Similarly, the RMSPE and MAPE, at 0.1682 and 0.1348 respectively, reflect a slight improvement in relative error performance. The average width of the confidence intervals, at 141.17, is narrower than before, suggesting a small gain in forecast precision.

These results confirm that removing calendar effects and correcting for outliers leads to a better-performing model, but the improvement is incremental rather than transformative. Nonetheless, even moderate gains in forecast accuracy and reliability are valuable when working with real-world time series data.

4.6   Outlier Treatment

4.6.1   For the final selected model, apply automatic outlier detection and attempt to interpret the detected anomalies.

```
##    Obs type_detected   W_coeff ABS_L_Ratio     Date
## 9   32            AO 0.2368963    3.357925 1996.583
```

```
## 3    56            AO   0.3714285    4.553159 1998.583
## 15  93            LS  -0.1822509    2.995924 2001.667
## 2   104           AO   0.4075404    4.831680 2002.583
## 13 132           TC  -0.2070318    3.077819 2004.917
## 11 152           AO   0.2236410    3.287816 2006.583
## 5   164           AO   0.3446157    4.499858 2007.583
## 8   178           TC  -0.2537906    3.459367 2008.750
## 1   180           TC  -0.5054508    5.653970 2008.917
## 4   188           AO   0.3437699    4.344369 2009.583
## 12 228           LS   0.2007752    3.149441 2012.917
## 6   248           AO  -0.2787073    3.709168 2014.583
## 7   260           AO  -0.2690139    3.643242 2015.583
## 10 284           AO  -0.2438664    3.389168 2017.583
## 14 297           TC  -0.2370152    3.083733 2018.667
```

If we study the reason behind these outliers, we can see that there are some outliers we are able to identify. However, there are many others that we cannot. The transitory changes we find in both 2008 and 2009 are known to be due to the crisis experienced during those years. But for the rest, we have not been able to identify what caused them.

We apply these commands to generate a new series that does not take into account either the calendar effects or the previously detected outliers. Afterwards, we generate the model that analyzes this new series without outliers or calendar effects.

Now, just as we did with the series that ignored calendar effects, we will do the same with the series that does not take into account either the calendar effects or the outliers.

First we start studing the homocedasticity of our residuals.

We can see that this metric continues to show no improvement. Despite the transformations and adjustments applied to the series, the issue persists. The assumption of homoscedasticity — that is, the presence of constant variance across the residuals or the series — is still not satisfied. This indicates that the variability in the data is not stable over time, which may affect the reliability of our model's estimates and inferences. Further analysis or alternative modeling approaches may be necessary to address this limitation.

```
##
##  Shapiro-Wilk normality test
##
## data:  resid_star_lin
## W = 0.9821, p-value = 0.0008379
```

If we study the normality of the residuals, we can see that they still do not follow a normal distribution. This is evident both from the visual inspection of the plots, where deviations from normality are still noticeable, and from the statistical results. Specifically, the p-value of the normality test remains below 0.05, which suggests that the null hypothesis of normality should be rejected. Therefore, we can conclude that our residuals continue to deviate significantly from a normal distribution, which could impact the validity of certain statistical inferences based on the model.

Finally we are going to study the independence

We can observe that the p-values obtained are even higher than those from the model that only ignores calendar effects. This indicates a stronger result in favor of the null hypothesis of independence. Therefore, we can conclude that the residuals from this model can indeed be considered independent. This is a positive outcome, as the assumption of residual independence is crucial for the validity of many statistical models and ensures that there is no autocorrelation affecting the results.

If we now make a prediction of our series using this new series, we obtain the following results and metrics.

```
## [1] 17.95891

## [1] 0.1032013

## [1] 22.54406

## [1] -326.2528

## [1] 16.8943

## [1] 9.03954

## [1] 21.57165

## [1] -621.2785
```

Table 1: Comparison of Forecast Metrics: Original vs Adjusted Model

Metric	Original_Model	Adjusted_Model
MAE	17.9589	16.8943
MAPE	0.1032	9.0395
RMSE	22.5441	21.5717
AIC	-326.2528	-621.2785

We can observe that the results obtained from the adjusted model—one that no longer includes calendar effects or outliers—are significantly better than those of the original model. This improvement is reflected in all key performance metrics, such as lower error values and a better model fit. Therefore, we can confidently conclude that the forecasts will be more accurate and reliable when using the model that has been corrected for these phenomena. Removing calendar effects and anomalous events leads to a clearer understanding of the underlying pattern in the data, which translates into improved forecasting performance.

5 Conclusions and discussion

This study applied the Box-Jenkins methodology to model and forecast the monthly production of passenger cars in Spain, using data provided by the Ministry of Industry, Energy and Tourism. The analysis followed a rigorous process including stationarity assessment, model identification, parameter estimation, residual diagnostics, and out-of-sample forecast evaluation. Several ARIMA models were proposed and evaluated, ultimately selecting the ARIMA(0,1,1)(0,1,1)12 model as the most suitable for capturing both regular and seasonal dynamics in the series.

Residual analysis confirmed the need to address heteroscedasticity, non-normality, and autocorrelation issues, which initially indicated that the model, although structurally sound, could benefit from further refinement. To improve model performance and interpretability, two additional adjustments were made: the treatment of **calendar effects** and the **correction of outliers**.

The inclusion of external regressors to model calendar effects, such as the number of working days and the timing of Easter holidays, led to a modest improvement in forecast accuracy. Although the reduction in error metrics (RMSE, MAE, RMSPE, MAPE) and the average confidence interval width was not substantial,

the adjusted model provided slightly more precise and reliable forecasts. This reinforces the importance of considering domain-specific calendar structures when modeling economic or industrial time series.

The removal of **outliers** had a more pronounced impact. After adjusting the model to exclude anomalies and calendar influences, all key performance indicators improved significantly, suggesting that these elements were obscuring the true signal in the data. The refined model achieved both a better fit and more accurate forecasts, highlighting the value of careful preprocessing and anomaly detection in time series modeling.

In summary, the methodology applied in this project demonstrates that robust time series forecasting requires not only the correct selection of model structure but also careful attention to external factors and irregular events. Even when improvements appear incremental, they can be meaningful in practical contexts—especially in industrial forecasting, where small gains in predictive accuracy can translate into better planning and resource allocation. The final model, corrected for both calendar and outlier effects, offers a more faithful representation of the underlying data-generating process and improves confidence in future projections.

6 References

This report is based on the data analysis and statistical procedures taught in the course, without the use of additional external sources.