Time Series Analysis of the French Industrial Production Index for Electricity Production

Linear time series project

Ilian Hadna Vincent Gimenes

2023 - 2024

Résumé

This report presents a comprehensive time series analysis of the Industrial Production Index (IPI) for electricity production from non-renewable sources in France, spanning from January 1990 to February 2024. The analysis includes data cleaning, transformation to stationarity, model selection, and validation using ARMA and ARIMA models.

I Data cleaning and presentation

1 Description

The chosen series represents the Industrial Production Index (IPIC) for the production, transportation, and distribution of electricity, classified under the NAF revision 2, at the class level, with code 35.11Y. This class encompasses the production of electricity from non-renewable sources.

This series has been created using the correction of seasonal variations and working day effects (CVS-CJO) is a standard statistical procedure applied to monthly or quarterly raw data series. Its primary aim is to remove cyclic components such as seasonality, the influence of working days, and leap years from the data. This correction facilitates a clearer understanding of the fundamental evolution of the series, including trends and cyclical fluctuations.

The dataset spans from January 1990 to February 2024, providing comprehensive insights into the production activity of electricity from non-renewable sources over this period.

The source can be found here : `https://www.insee.fr/fr/statistiques/serie/010768228#Graphique`

2 Transformation to Stationarity

To achieve a stationary series, a first-order differentiation was applied, effectively calculating the difference between consecutive observations. This approach is targeted at removing deterministic trends or seasonality, thereby stabilizing the variance of the series.

The rationale behind differentiation is to eliminate linear trends or periodic fluctuations, thereby rendering the series stationary. This is a critical step in preparing the data for further time series analysis using ARMA models.

After differentiating the series, we employed the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests to confirm stationarity :

- The ADF test provided a p-value of 0.01, leading to the rejection of the null hypothesis of non-stationarity, which supports the effectiveness of the differentiation.

- The KPSS test resulted in a p-value of 0.1, suggesting the series is stationary at conventional significance levels, though there is some ambiguity at the 10% level.

3 Graphical representation

The plot of the original time series shown in Figure 1 depicts a gradual increase in production levels until about 2008, followed by a phase of fluctuation without a clear trend. This initial increase and subsequent plateau indicate the presence of trends and potential cyclical behaviors, which necessitate the application of differentiation to achieve stationarity.

After applying a first-order differentiation to the series, the resulting transformed series (see Figure 2)shows no apparent trend or seasonality, indicating a stationary process suitable for further analysis with time series models. The fluctuations around a zero mean confirm the efficacy of the transformation method used.

II ARMA and ARIMA modelisation

4 The ARMA model

In this part, we aim to identify the best-fitting ARMA model for our time series data. This involves detailed ACF and PACF analysis followed by the application of information criteria and residual tests to ensure the model's adequacy.

ACF and PACF Analysis

— **ACF Plot** : A quick descent below the significance threshold after the first lag, with a notable spike at lags 11 or 12, suggests significant MA components at these lags.

— **PACF Plot** : A similar pattern is observed in the PACF, with a cutoff after the 5-6th lag and spikes around the 11-12th lags, indicating potential AR components.

Combining ACF and PACF Insights

Based on the ACF and PACF plots, an ARMA(1,1) model is hypothesized to be the most appropriate. This model includes :

— One AR term ($p = 1$) : Justified by the sharp cut-off in the PACF.

— One MA term ($q = 1$) : Suggested by the rapid decay in the ACF.

Rethinking Model Selection

Unlike typical tutorials where ARMA models are selected solely based on AIC and BIC minimization, our approach is twofold :

1. Initially, potential models were evaluated based on AIC and BIC values.
2. However, given the failure of some models, notably ARMA(1,1), to pass autocorrelation tests of residuals, a reevaluation was necessary.

Model Selection Process

— First, models were explored that showed no significant autocorrelation in residuals, which pointed us to consider ARMA(3,3) and ARMA(5,1).

— Among these, the model minimizing the information criteria (AIC/BIC) was selected, resulting in the choice of ARMA(3,3) for its robustness in validation tests.

Statistical Validation and Model Fitting

— **Model Testing** : The ARMA(3,3) model was rigorously tested against various metrics, showing lower AIC and BIC values while ensuring that residuals do not exhibit any autocorrelation across all considered lags.

— **Residual Analysis** : The residuals of the ARMA(3,3) model displayed no significant patterns (see Figure 4) , indicating an effective capture of the time series dynamics.

Model Validity Check

The validity of the selected ARMA(3,3) model was further assessed by analyzing the residuals see Figure 5 of the fitted model :

— **Residual Plots** : Visual inspection of the residuals plot did not reveal any obvious patterns or trends, suggesting that the residuals are random, which is a good indicator of model fit.

— **Ljung-Box Test** : The Ljung-Box test for autocorrelation in residuals was performed, where a high p-value (greater than 0.05) indicates that there is no significant autocorrelation for all lags up to 24. The test results for the ARMA(3,3) model residuals showed p-values well above 0.05, confirming that the residuals are white noise.

Conclusion

The ARMA(3,3) model, selected through a rigorous process of ACF and PACF analysis, followed by validation against autocorrelation and information criteria, proves to be a statistically robust model that captures the underlying processes of the time series accurately. It is deemed suitable for forecasting and further analyses.

5 The ARIMA model

Model Configuration and Rationale

The ARIMA(3,1,3) model was chosen based on the initial analysis which indicated the need for first-order differencing to achieve stationarity. The configuration integrates :

— **p (AR part) : 3**, indicated by the sharp cutoff after the first lag in the Partial Autocorrelation Function (PACF), suggesting a significant autoregressive component.

— **d (Differencing order) : 1**, necessary to remove the non-stationarity observed in the initial series data.

— **q (MA part) : 3**, suggested by the rapid decay observed in the Autocorrelation Function (ACF), indicating an essential moving average term.

Parameters were optimally estimated using the maximum likelihood estimation method, considering both the autoregressive and moving average components to model the underlying data structure effectively.

Residual Analysis

The residual plots of the ARIMA Model explains the significance of the absence of patterns in the residuals, which indicates a good fit of the model. Such randomness is critical, indicating that the model has successfully captured the underlying data patterns without any leftover structure in the errors (see Figure 5).

Here, the residuals of the model are plotted against the theoretical quantiles of a normal distribution. The close alignment of the points along the 45-degree line indicates that the residuals approximately follow a normal distribution, affirming the assumption of normality required for many statistical tests used in validating the model (see Figure 6).

Sum-up

The ARIMA(3,1,3) model's effectiveness in forecasting and analyzing economic time series is demonstrated through rigorous statistical analysis and parameter estimation. Its simplicity, combined with robust statistical validation, ensures that it is well-suited for practical applications in economic and financial analyses.

For such a model ARIMA(3,1,3), we obtain the following equation :

$$\Delta X_t = Y_t = -0.84Y_{t-1} - 0.43Y_{t-2} + 0.3851Y_{t-3} + e_t + 0.37e_{t-1} - 0.06e_{t-2} - 0.78e_{t-3} \tag{1}$$

where $\Delta X_t = X_t - X_{t-1}$ represents the first difference of the series

III Prevision

6 Confidence region on the values (X_{T+1}, X_{T+2})

Given the zero expectation of future residuals, optimal forecasts at time T in the context of the ARMA(3,3) model are computed as follows :

$$\begin{aligned}
\hat{X}_{T+1|T} &= \phi_1 X_T + \phi_2 X_{T-1} + \phi_3 X_{T-2} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + \theta_3 \varepsilon_{T-2}, \\
\hat{X}_{T+2|T} &= \phi_1 \hat{X}_{T+1|T} + \phi_2 X_T + \phi_3 X_{T-1} + \theta_2 \varepsilon_T + \theta_3 \varepsilon_{T-1}.
\end{aligned}$$

Then, the prediction errors are given by :

$$\hat{X} = \begin{bmatrix} \hat{X}_{T+1|T} \\ \hat{X}_{T+2|T} \end{bmatrix}, \quad X = \begin{bmatrix} X_{T+1} \\ X_{T+2} \end{bmatrix},$$
$$X - \hat{X} = \begin{bmatrix} X_{T+1} - \hat{X}_{T+1|T} \\ X_{T+2} - \hat{X}_{T+2|T} \end{bmatrix}.$$

The variance of these prediction errors can be structured as :

$$\begin{aligned}
&V(X_{T+1} - \hat{X}_{T+1|T}) = \sigma^2, \\
&V(X_{T+2} - \hat{X}_{T+2|T}) = \sigma^2 \left(1 + (\theta_1 + \phi_1)^2\right).
\end{aligned}$$

Thus, the prediction errors follow a normal distribution with zero mean and covariance matrix Σ :

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \theta_1 + \phi_1 \\ \theta_1 + \phi_1 & 1 + (\theta_1 + \phi_1)^2 \end{bmatrix}.$$

Given the determinant of Σ is $\sigma^4(1 + (\theta_1 + \phi_1)^2 - (\theta_1 + \phi_1)^2) = \sigma^4$, the covariance matrix is invertible, assuming $\sigma^2 > 0$.

By the properties of multivariate normal distributions and the chi-squared distribution, we have :

$$(\mathbf{X} - \hat{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X} - \hat{\mathbf{X}}) \sim \chi^2(2),$$

where $\chi^2(2)$ denotes a chi-squared distribution with 2 degrees of freedom. Therefore, the confidence region of level α is given by :

$$\left\{ \mathbf{X} \in \mathbb{R}^2 \mid (\mathbf{X} - \hat{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X} - \hat{\mathbf{X}}) \leq q_{1-\alpha,\chi^2(2)} \right\},$$

where $q_{1-\alpha,\chi^2(2)}$ is the $(1-\alpha)$ quantile of the $\chi^2(2)$ distribution.

7 Necessary hypotheses to obtain the confidence region

To obtain the results presented previously, certain assumptions were assumed true. Here they are :

— The model is perfectly known.

— The coefficients obtained in part 2 are the true coefficients of our model.

— The white noise follows a normal distribution $t \sim N(0, \sigma^2)$.

— Finally, $\sigma^2 > 0$.

It is the fact that t is a linear innovation, due to the canonical form of the ARMA model (roots outside the unit circle, and no common roots), which allows us to directly determine the form of $\hat{X}_{T+1|T}$ and $\hat{X}_{T+2|T}$. We also assume that the residual follows a normal distribution, which was tested in part 2. Finally, we assume that the variance of the residual is known.

In fact, if the variance of the residual is unknown, it must be estimated, and the confidence interval will then depend on the distribution of a Student's t-distribution, with wider tails and therefore a less precise interval. This is provided that the true values of the ARMA model parameters are known. Indeed, if the true ARMA parameter values are unknown, $\hat{\Sigma}$ is then doubly uncertain (due to the uncertainty associated with estimating the variance of the residual and the uncertainty of estimating the coefficients of the model), which corresponds to the question :

8 Graphical representation of the confidence region

The forecast graph graph shows the estimated values along with their confidence intervals, highlighting the uncertainty in predictions and emphasizing the practical applicability of the model for forecasting in uncertain economic environments. see Figure 7).

9 Open question

The prediction of X_{T+1} can be enhanced using Y_{T+1} if Y_t Granger-causes X_t. Granger causality implies that knowing the past and current values of Y_t helps reduce the mean squared error (MSE) of predicting X_{T+1}, compared to predictions made without the knowledge of Y_t.

Granger causality is based on the principle that if a time series Y_t provides statistically significant information that helps in predicting X_t, beyond the information already contained in the past values of X_t alone, then Y_t is said to

Granger-cause X_t. This concept is particularly applicable when Y_{T+1} is available before X_{T+1}, allowing for an improvement in prediction accuracy under the right circumstances.

To determine whether Y_t Granger-causes X_t, the following statistical tests can be performed :

1. Construct a vector autoregressive (VAR) model including both Y_t and X_t as variables.
2. Estimate the model and conduct hypothesis tests on the coefficients of lagged values of Y_t predicting X_t. The null hypothesis for each coefficient is that it equals zero, implying no predictive power from Y_t to X_t.
3. Use an F-test to assess the joint significance of the coefficients on lagged values of Y_t. A rejection of the null hypothesis suggests that Y_t Granger-causes X_t.

To implement it in practice, we can :
- Fit a VAR model using historical data up to time T.
- Forecast X_{T+1} using the model that includes past values of both X_t and Y_t.
- Compare the forecast accuracy, in terms of MSE, against a model that forecasts X_{T+1} using only past values of X_t.

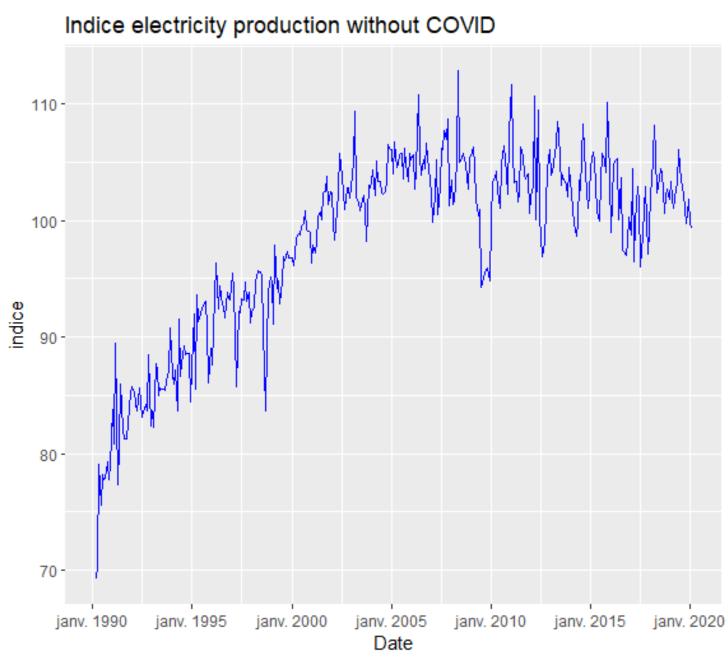

FIGURE 1 – Time series of the evolution of electricity production from 1990 to january 2020

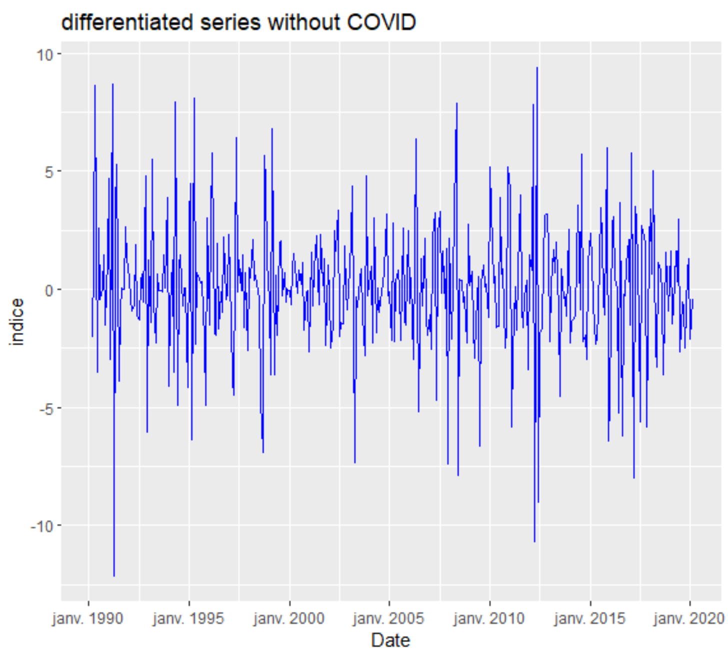

FIGURE 2 – Differentiated of the evolution of electricity production from 1990 to january 2020

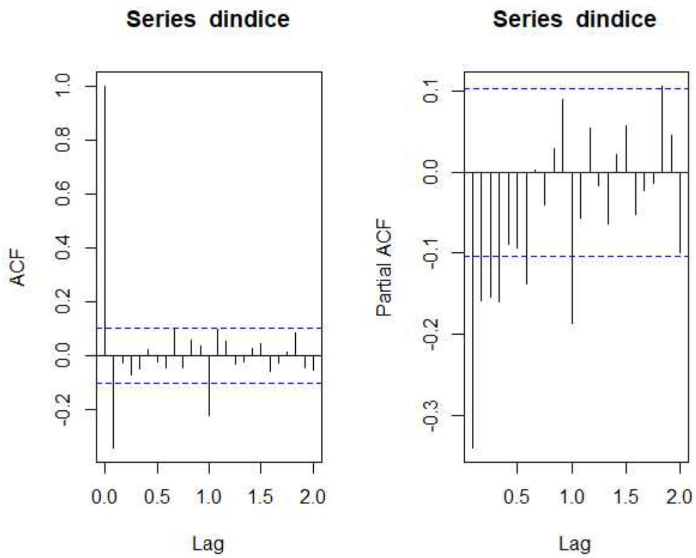

FIGURE 3 – Plot of empirical autocovariances (first plot) and partial empirical autocovariances (second graph) of the series $(Y_t)_{t \in T})$

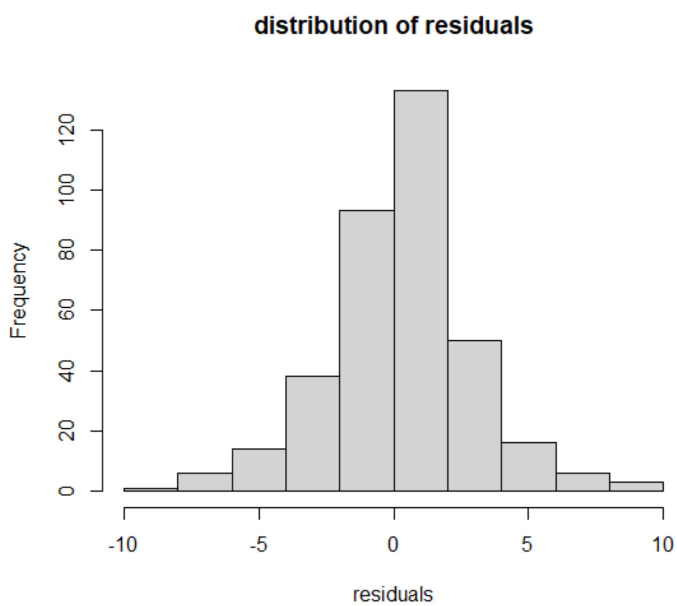

FIGURE 4 – Distribution of residuals of the ARMA(3,3) model.

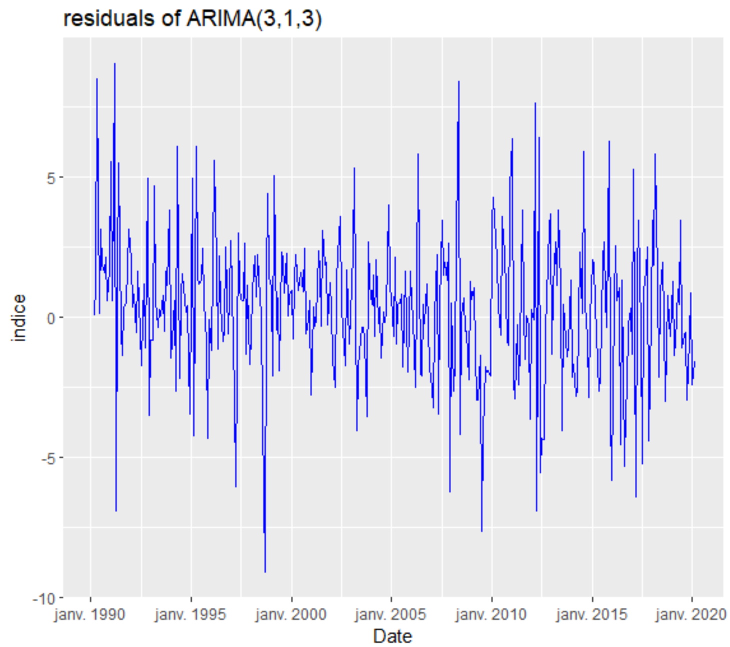

FIGURE 5 – Residuals Plot

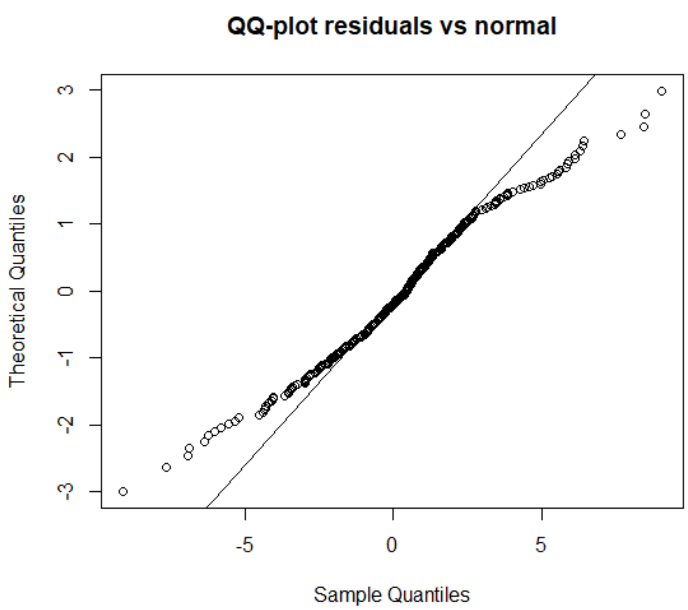

FIGURE 6 – QQ-Plot of Residuals vs Normal

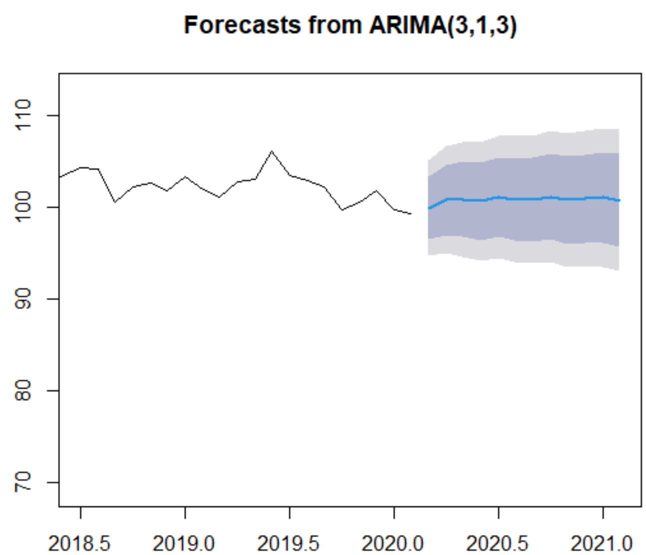

Figure 7 – Forecast Plot from ARIMA(3,1,3)