TIME SERIES REPORT

TZIEMI NGANSOP

TANGOUO KUETE

Linear Time series

ENSAE de PARIS

Supervisor: Mr. FERMANIAN Jean David

Mai 2025

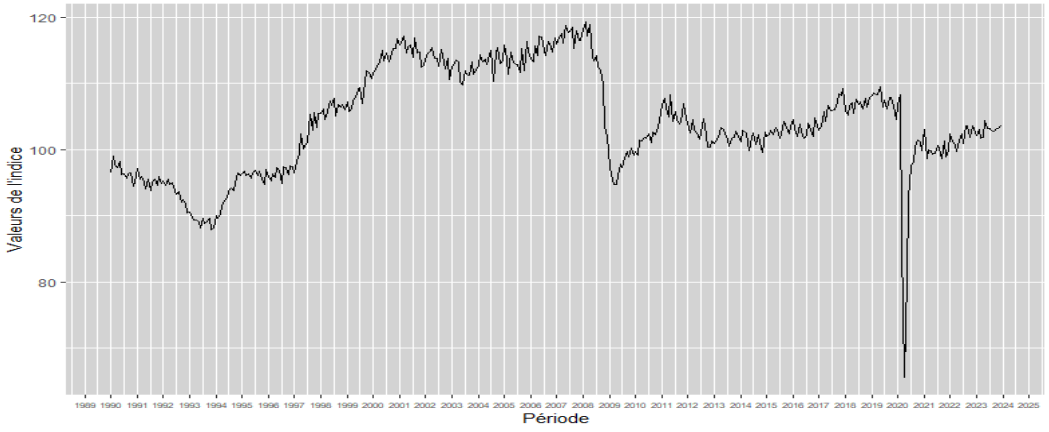
TABLE OF CONTENTS

1 DATA PREPARATION

1.1 Description of the Series

This study aims to model and forecast the Industrial Production Index (IPI) in the manufacturing sector, based on monthly observations spanning from January 1990 to January 2024. The time series used has been seasonally and calendar-adjusted. The data, dated January 2025, comes from the French National Institute of Statistics and Economic Studies (INSEE), specifically from the Annual Production Survey (EAP). We denote this series as $(\text{indice}_t)_{t\in\mathcal{T}}$, where $\mathcal{T}$ represents the set of monthly observation dates.

Figure 1: Description of the time series



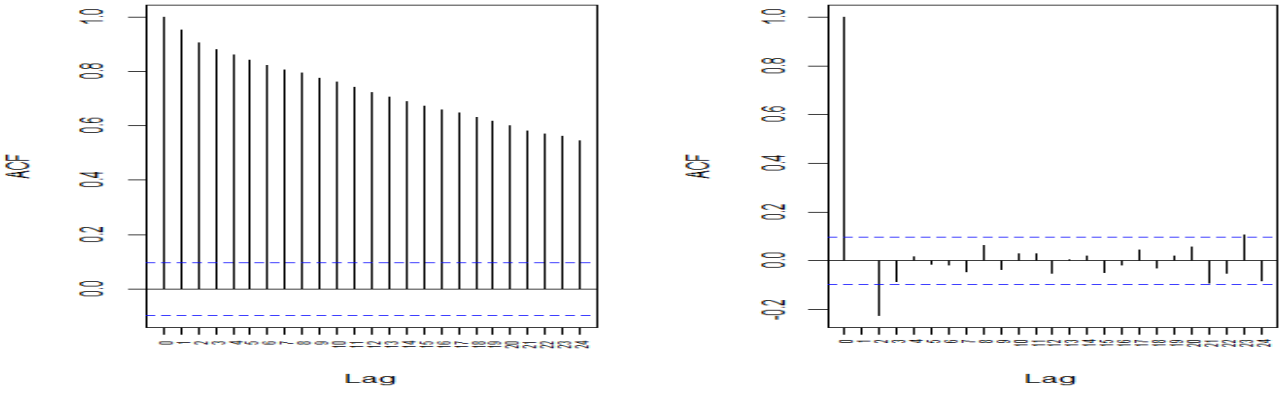
Between 1990 and 2024, we first observe a slight downward trend until 1993, likely linked to the European economic crisis of that time. From 1994 onwards, the index shows sustained growth, indicating a phase of industrial expansion continuing into the early 2000s. Between 2001 and 2008, the index remains relatively stable at a high level, though with noticeable fluctuations. However, starting in 2008, the index drops sharply due to the global financial crisis, before stabilizing at a lower plateau. The 2010–2019 period is marked by stagnation, with some temporary recoveries, but the pre-crisis levels are not regained. In 2020, a historic collapse occurs, likely caused by the COVID-19 pandemic, which triggers a sharp decline in manufacturing activity. Finally, between 2021 and 2024, a partial recovery is observed, although it does not reach the levels seen before 2008. The overall trend is piecewise constant despite short-term fluctuations.

1.2 Stationarization

We observe that the autocorrelations of the series ($\texttt{indice}_t$) are significantly different from zero and decay very slowly, supporting the hypothesis of non-stationarity. The Augmented Dickey-Fuller (ADF) test shows the presence of a deterministic trend, indicating a unit root with a p-value of 0.3471, which is not significant at the 5% level. [Appendix 1]

Since our index is an **economic time series**, we apply a first-order differencing, i.e., $d\_\texttt{indice}_t = \texttt{indice}_t - \texttt{indice}_{t-1}$. In the autocorrelation function of $d\_\texttt{indice}_t$, only lags 1 and 3 are significantly different from zero at the 5% level. An ADF test rejects the presence of a unit root at the 5% level. The KPSS test, along with the Phillips-Perron test, supports the ADF result by accepting the null hypothesis of stationarity. [Appendix 2]

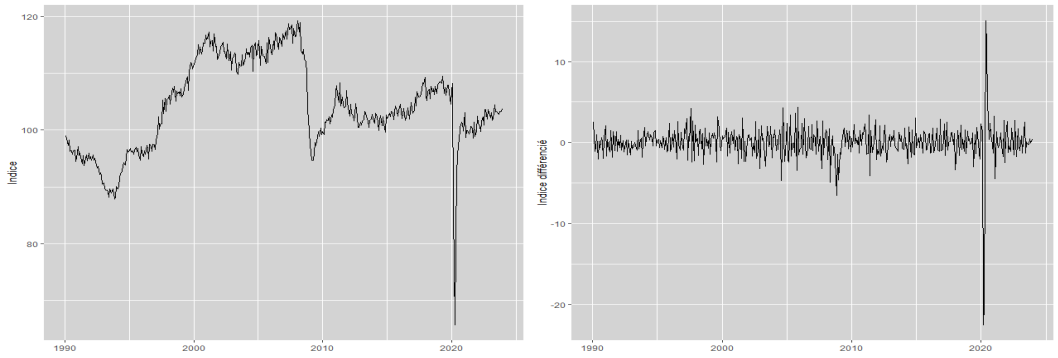
Figure 2: Autocorrelations of the manufacturing production index and its differenced series



2 ARMA AND ARIMA MODELING

In the previous section, we differenced the original series once ($d = 1$) to achieve stationarity. We now seek an $ARMA(p, q)$ model that best fits the differenced series ($d\_\texttt{indice}_t$). Consequently, the corresponding model for the original series ($\texttt{indice}_t$) will be an $ARIMA(p, d, q)$ with $d = 1$.

Figure 3: Plots of the original series $\texttt{indice}_t$ and the differenced series $d\_\texttt{indice}_t$



2.1 ARMA Model

2.1.1 Model Specification

We assume that the series $d\_\texttt{indice}_t$ is modeled by an ARMA(p, q) process, where p is the autoregressive order and q the moving average order. The standard form of the ARMA(p, q) model is given by:

$$d\_\texttt{indice}_t = \phi_1 d\_\texttt{indice}_{t-1} + \cdots + \phi_p d\_\texttt{indice}_{t-p} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q} \tag{2.1}$$

where (ε_t) is a white noise process.

This model can also be written in polynomial form:

$$\phi(B)\, d\_\texttt{indice}_t = \psi(B)\, \varepsilon_t \tag{2.2}$$

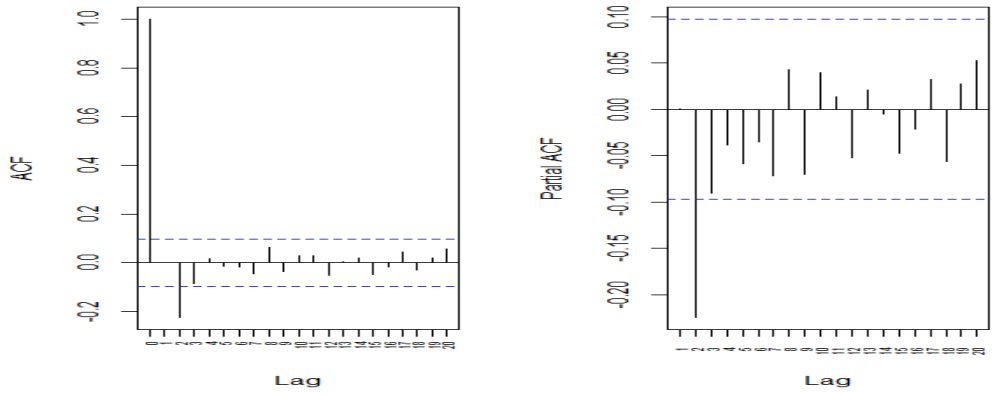
with B being the backshift operator, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, and $\psi(B) = 1 + \psi_1 B + \cdots + \psi_q B^q$.

For the model to be stationary, the roots of $\phi(z)$ must lie outside the unit circle ($|z| > 1$). Similarly, for invertibility, the roots of $\psi(z)$ must satisfy the same condition.

2.1.2 Choice of Orders p and q

The orders p and q are determined through analysis of the autocorrelation (ACF) and partial autocorrelation (PACF) functions. The ACF becomes insignificant after lag 2, suggesting that the moving average order should not exceed 2, i.e., $q_{\max} = 2$. Likewise, the PACF becomes insignificant after lag 2, indicating a maximum autoregressive order of $p_{\max} = 2$.

Figure 4: Autocorrelation and partial autocorrelation functions of $d\_\texttt{indice}_t$



Selection of Valid ARMA Models

Models to test: All combinations such that $p \leq 2$ and $q \leq 2$.

Model validity: A model is valid if: - The residuals are white noise (Ljung-Box test), - The coefficients are significantly different from zero (Student's t-test).

The Ljung-Box test is applied with:

$$H_0 : \rho(1) = \rho(2) = \cdots = \rho(K) = 0 \quad \text{vs} \quad H_1 : \exists k \text{ such that } \rho(k) \neq 0$$

with $K = T/3$, and the test statistic:

$$\text{LB}(K) = \frac{T(T+2)}{T-p-q-1} \sum_{k=1}^{K} \frac{\hat{\rho}_k^2}{T-k} \sim \chi^2(K-p-q-1)$$

Coefficient significance is tested by:

$$t = \frac{\hat{\phi}_i - 0}{\hat{\sigma}_{\hat{\phi}_i}} \sim T(T-p-q)$$

with standard deviation estimated by:

$$\hat{\sigma}_{\hat{\phi}_i} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T-1} \left(d\_\texttt{indice}_t - \bar{d\_\texttt{indice}} \right)^2}$$

Selected models: The following models are valid: MA(2), ARMA(1,2), and ARMA(2,1).

Best ARMA Model Selection

Using AIC and BIC criteria, ARMA(1,2) is preferred by AIC, while BIC favors MA(2). A likelihood ratio test was used to decide between them. [Appendix 3]

Ultimately, the ARMA(1,2) model is selected. Its equation is:

$$d\_\texttt{indice}_t = 0.3593\, d\_\texttt{indice}_{t-1} + \varepsilon_t - 0.3912\, \varepsilon_{t-1} - 0.2267\, \varepsilon_{t-2}$$

The Ljung-Box test confirms that the residuals are white noise [Appendix 4], and the Wald test confirms all coefficients are significant [Appendix 5].

Causality and Invertibility Verification

To ensure parameter stability and convergence, it is crucial to verify that the ARMA(1,2) model is both **causal** and **invertible**.

The root of the AR polynomial is $Z_1 = 1/0.3593 \approx 2.784 > 1$, confirming causality.

The roots of the MA polynomial are $Z_1 = 1.5613$, $Z_2 = -2.8328$, both with modulus greater than 1, ensuring invertibility.

Thus, the ARMA(1,2) process is stable, causal, and invertible. The model can be written as:

$$\varepsilon_t = \psi^{-1}(B) d\_\texttt{indice}_t = \sum_{j=0}^{\infty} b_j d\_\texttt{indice}_{t-j}, \quad \text{with } \sum_{j=0}^{\infty} |b_j| < \infty$$

2.2 ARIMA Model

Since the selected model for the differenced series $d\_\texttt{indice}_t$ is ARMA(1,2) and it satisfies the conditions of causality and invertibility, the corresponding ARIMA(1,1,2) model for the original series $\texttt{indice}_t$ is well defined. The residuals are white noise [Appendix 6], and all coefficients are significant at the 95% level.

The ARIMA(1,1,2) model is:

$$\texttt{indice}_t - \texttt{indice}_{t-1} = 0.3593\, \texttt{indice}_{t-1} + \varepsilon_t - 0.3912\, \varepsilon_{t-1} - 0.2267\, \varepsilon_{t-2}$$

3 FORECASTING

3.1. Equation Satisfied by the Confidence Region at Level α for Future Values ($\texttt{indice}_{T+1}, \texttt{indice}_{T+}$

We concluded in the previous sections that the differenced series ($\texttt{d\_indice}_t$) follows an ARMA(1,2) model without a constant. We will forecast the next two values of the series, i.e., $\texttt{d\_indice}_{T+1}$ and $\texttt{d\_indice}_{T+2}$, which correspond to the March and April 2022 observations of the production index.

The estimated model is given by:

$$\texttt{d\_indice}_t = \phi \texttt{d\_indice}_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

Since the model is causal and invertible, ε_t is the innovation of $\texttt{d\_indice}_t$, hence:

$\mathbb{E}(\varepsilon_{T+h} \mid \mathcal{I}_T) = 0, \quad \forall h > 0, \quad$ where $\mathcal{I}_T$ denotes the information available at time T (i.e., $\texttt{d\_indice}_T, \texttt{d\_indic}$

$$\text{then, } \begin{cases} \texttt{d\_indice}_{T+1} = \phi \texttt{d\_indice}_T + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} \\ \texttt{d\_indice}_{T+2} = \phi \texttt{d\_indice}_{T+1} + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T \\ \qquad\qquad = \phi^2 \texttt{d\_indice}_T + (\phi + \theta_1)\varepsilon_{T+1} + (\phi\theta_1 + \theta_2)\varepsilon_T + \phi\theta_2 \varepsilon_{T-1} + \varepsilon_{T+2} \end{cases} \tag{3.1}$$

From the above, we deduce:

$$\begin{cases} \widehat{\texttt{d\_indice}}_{T+1|T} = \mathbb{E}(\texttt{d\_indice}_{T+1} \mid \texttt{d\_indice}_T, \texttt{d\_indice}_{T-1}, \dots) = \phi \texttt{d\_indice}_T + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} \\ \widehat{\texttt{d\_indice}}_{T+2|T} = \mathbb{E}(\texttt{d\_indice}_{T+2} \mid \texttt{d\_indice}_T, \texttt{d\_indice}_{T-1}, \dots) = \phi^2 \texttt{d\_indice}_T + (\phi\theta_1 + \theta_2)\varepsilon_T + \phi\theta_2 \varepsilon_{T-1} \end{cases} \tag{3.2}$$

By subtracting the previous equations, we obtain:

$$\begin{cases} \texttt{d\_indice}_{T+1} - \widehat{\texttt{d\_indice}}_{T+1|T} = \varepsilon_{T+1} \\ \texttt{d\_indice}_{T+2} - \widehat{\texttt{d\_indice}}_{T+2|T} = (\phi + \theta_1)\varepsilon_{T+1} + \varepsilon_{T+2} \end{cases} \tag{3.3}$$

Now, we know that:

$$\begin{cases} \texttt{d\_indice}_{T+1} = \texttt{indice}_{T+1} - \texttt{indice}_T, & \widehat{\texttt{d\_indice}}_{T+1|T} = \widehat{\texttt{indice}}_{T+1|T} - \widehat{\texttt{indice}}_{T|T} \\ \texttt{d\_indice}_{T+2} = \texttt{indice}_{T+2} - \texttt{indice}_{T+1}, & \widehat{\texttt{d\_indice}}_{T+2|T} = \widehat{\texttt{indice}}_{T+2|T} - \widehat{\texttt{indice}}_{T+1|T} \end{cases} \tag{3.4}$$

Combining (3.3) and (3.4), we finally obtain:

$$\begin{cases} e_{T+1} = \texttt{indice}_{T+1} - \widehat{\texttt{indice}}_{T+1|T} = \varepsilon_{T+1} \\ e_{T+2} = \texttt{indice}_{T+2} - \widehat{\texttt{indice}}_{T+2|T} = (\phi + \theta_1)\varepsilon_{T+1} + \varepsilon_{T+2} \end{cases} \tag{3.5}$$

Thus, the forecast errors on $\texttt{indice}_{T+1}$ and $\texttt{indice}_{T+2}$ are given by:

$$\begin{cases} e_{T+1} = \varepsilon_{T+1} \\ e_{T+2} = (1 + \phi + \theta_1)\varepsilon_{T+1} + \varepsilon_{T+2} \end{cases}$$

Variance-Covariance Matrix of the Forecast Errors

Let $E = \begin{pmatrix} e_{T+1} \\ e_{T+2} \end{pmatrix}$ be the vector of errors. Then:

$$\Sigma = \mathbb{E}[EE'] = \sigma^2 \begin{pmatrix} 1 & 1 + \phi + \theta_1 \\ 1 + \phi + \theta_1 & 1 + (1 + \phi + \theta_1)^2 \end{pmatrix}$$

The matrix Σ is positive definite since $\text{Det}(\Sigma) = \sigma^2 > 0$, which is assumed strictly positive. We can thus perform a joint null hypothesis test of the components of E at level $\alpha \in]0, 1[$. The hypotheses are: $H_0 : E = 0_{\mathbb{R}^2}$ versus $H_1 : E \neq 0_{\mathbb{R}^2}$. The test statistic is $E^T\Sigma^{-1}E$, which follows a $\chi^2(2)$ distribution under H_0. The rejection region for H_0 at level α is: $E^T\Sigma^{-1}E > q^2_{1-\alpha}(2)$, where $q^2_{1-\alpha}(2)$ is the $1 - \alpha$ quantile of the $\chi^2(2)$ distribution.

Based on the previous results, the 95% confidence intervals for the next two values are:

$$\begin{cases} CI_{95\%}(\texttt{indice}_{T+1}) = \left[\widehat{\texttt{indice}}_{T+1|T} - 1.96 \times \hat{\sigma},\ \widehat{\texttt{indice}}_{T+1|T} + 1.96 \times \hat{\sigma}\right] \\ CI_{95\%}(\texttt{indice}_{T+2}) = \left[\widehat{\texttt{indice}}_{T+2|T} - 1.96 \times \hat{\sigma}\sqrt{1 + (\phi + \theta_1)^2},\ \widehat{\texttt{indice}}_{T+2|T} + 1.96 \times \hat{\sigma}\sqrt{1 + (\phi + \theta_1)^2}\right] \end{cases}$$

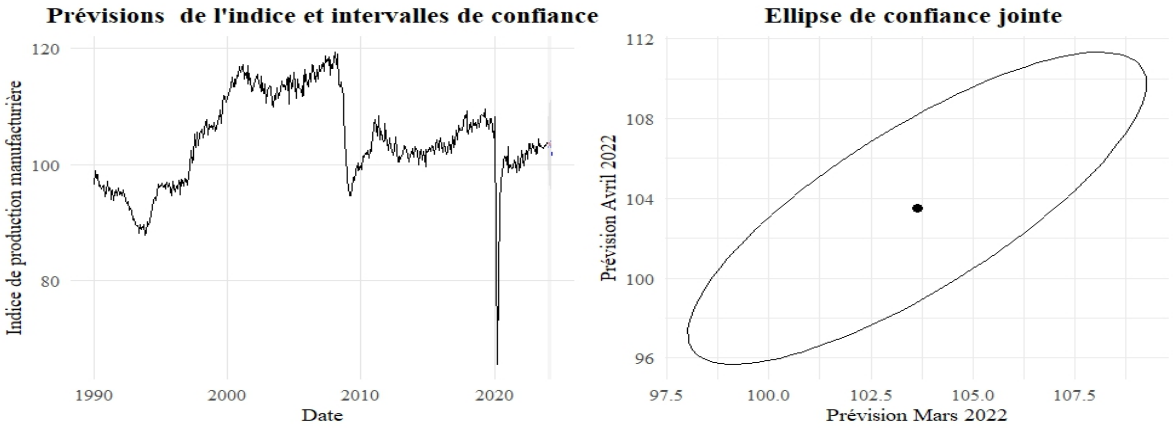
3.2. Assumptions Used to Construct the Confidence Intervals

We made four main assumptions in constructing the previous confidence intervals: All estimators used are consistent; The residuals are independent and identically distributed; The residuals are Gaussian: $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$; The variance-covariance matrix (Σ) is invertible and positive definite ($\sigma^2 > 0$).

3.1 Visualization of the Confidence Region

The predicted values of the manufacturing production index on January 1, 2024, and February 1, 2024, are respectively 103.62 and 103.50, with their associated 95% confidence regions shown below.

Figure 5: Forecast Visualization and Ellipse for $\texttt{indice}_t$



3.2 Early Use of Y_{T+1} to Improve the Forecast of X_{T+1}

Suppose Y_t is a stationary time series observed for $t = 1$ to T, and that Y_{T+1} is available earlier than X_{T+1}. The objective is to determine whether this information can improve the forecast of X_{T+1} and under what conditions.

Necessary Conditions

For Y_{T+1} to effectively improve the forecast of X_{T+1}, the following are required:

- There exists a significant correlation between X_{T+1} and Y_{T+1}: $\text{Corr}(X_{T+1}, Y_{T+1}) \neq 0$.
- The series (X_t, Y_t) can be jointly modeled using a VAR (Vector AutoRegressive) model, capturing their dynamic interactions.
- Y_{T+1} contains non-redundant information compared to the history of X_t alone, i.e., Y_{T+1} is not deterministic given X_T and its past.

Testing Methodology

We propose the following methodology to test the usefulness of Y_{T+1} in forecasting:

1. **Estimate a direct regression model:**

$$X_{T+1} = \beta_0 + \beta_1 Y_{T+1} + \varepsilon$$

   Then test $H_0 : \beta_1 = 0$. If H_0 is rejected, this indicates that Y_{T+1} has a significant effect.

2. **Compare forecasting performance:**

   - Construct two forecasting models: one using only the history of X_t, the other incorporating Y_{T+1}.
   - Compare the prediction errors (e.g., RMSE) of both models.

   A significant improvement in performance (i.e., reduction in RMSE) upon adding Y_{T+1} suggests the usefulness of this information.

3. **Estimate a VAR model:** If (X_t, Y_t) follows a VAR model:

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = A_1 \begin{pmatrix} X_t \\ Y_t \end{pmatrix} + \cdots + A_p \begin{pmatrix} X_{t-p+1} \\ Y_{t-p+1} \end{pmatrix} + \varepsilon_{t+1}$$

   Then the forecast of X_{T+1} can be refined using the early knowledge of Y_{T+1} if the equations are cross-correlated and residuals are correlated.

CONCLUSION

This project allowed us to deeply explore classical time series analysis tools, applied to a concrete economic issue: modeling the industrial production index in the manufacturing sector. Through the rigorous steps of the **Box-Jenkins** methodology—visualization, stationarity, differencing, estimation, validation, and forecasting—we were able to select a parsimonious, robust, and interpretable model: an ARIMA(1,1,2).

Furthermore, we showed how anticipating certain exogenous variables such as Y_{T+1} can enhance the forecasting of variables of interest like X_{T+1}, provided certain conditions are met. This approach illustrates the value of multivariate analysis when series exhibit dynamic interdependencies.

In sum, this work highlights the importance of combining statistical expertise, methodological

rigor, and economic interpretation to build relevant and decision-supportive models. It also opens up promising perspectives in vector modeling, integration of exogenous data, and short-term causal analysis.

ANNEXE

Table 1: Test de dicker fuller augmenté d'ordre 2 sur l'indice

Lag order	Statistique	P-valeur
2	−2.5451	0.3471

Table 2: Résultats des tests de stationnarité sur la série différenciée de l'indice

Test	Lag order	Statistique	P-valeur
ADF (Augmented Dickey-Fuller)	2	−14.524	0.01
KPSS	5	0.054155	> 0.1
PP (Phillips-Perron)	5	−329.23	0.01

Table 3: Critères d'information pour différents modèles ARMA

Critère	MA(2)	ARMA(1,2)	ARMA(2,1)
AIC	1843.8571	1841.650	1842.1874
BIC	1859.8924	1861.694	1862.2315
logLik	−917.9286	−915.825	−916.0937

Table 4: Test de Ljung-Box d'autocorrélation sur le modèle ARMA(1,2)

Lag	1	2	3	4	5	6	7	8	9	10	11	12
P-va	NA	NA	0.8	0.7	0.8	0.9	0.8	0.7	0.7	0.7	0.8	0.8

Table 5: Estimation des coefficients du modèle ARIMA(1,1,2)

Coefficient	Estimation	Statistique t
AR(1)	0.36	$2.25 > \pm 1,96$
MA(1)	−0.39	$-2.52 > \pm 1,96$
MA(2)	−0.23	$-4.33 > \pm 1,96$

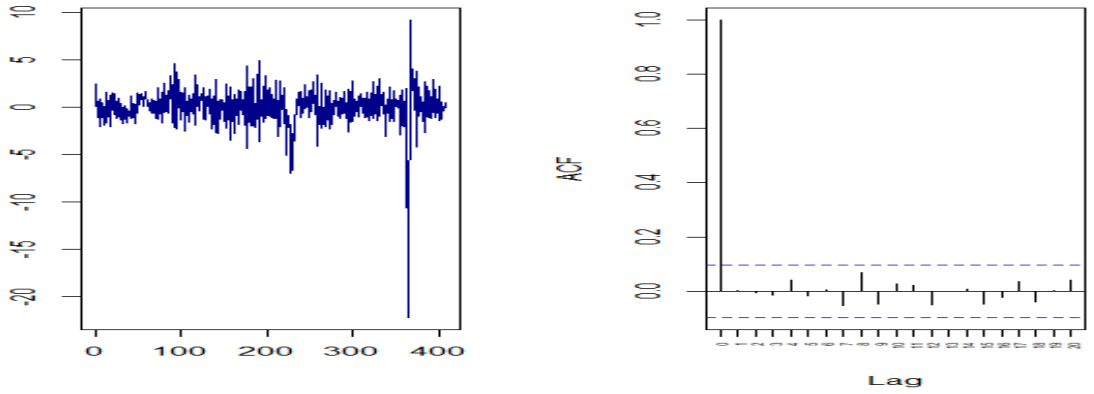


Figure 6: Autocorrélation des résidus