*stats*

MDPI

*Article*

# Deviations from Normality in Autocorrelation Functions and Their Implications for MA(q) Modeling

**Manuela Royer-Carenzi [1,*,†] and Hossein Hassani [2,†]**

[1] I2M, UMR 7373 Aix-Marseille University, CNRS, 13331 Marseille, France
[2] International Institute for Applied Systems Analysis (IIASA), Schloßpl. 1, 2361 Laxenburg, Austria; hassani.stat@gmail.com
* Correspondence: manuela.royer-carenzi@univ-amu.fr
† These authors contributed equally to this work.

**Abstract:** The identification of the orders of time series models plays a crucial role in their accurate specification and forecasting. The Autocorrelation Function (ACF) is commonly used to identify the order $q$ of Moving Average (MA($q$)) models, as it theoretically vanishes for lags beyond $q$. This property is widely used in model selection, assuming the sample ACF follows an asymptotic normal distribution for robustness. However, our examination of the sum of the sample ACF reveals inconsistencies with these theoretical properties, highlighting a deviation from normality in the sample ACF for MA($q$) processes. As a natural extension of the ACF, the Extended Autocorrelation Function (EACF) provides additional insights by facilitating the simultaneous identification of both autoregressive and moving average components. Using simulations, we evaluate the performance of $q$-order identification in MA($q$) models, which is based on the properties of ACF. Similarly, for ARMA($p, q$) models, we assess the $(p, q)$-order identification relying on EACF. Our findings indicate that both methods are effective for sufficiently long time series but may incorrectly favor an ARMA($p, q-1$) model when the $a_q$ coefficient approaches zero. Additionally, if the cumulative sums of ACF (SACF) behave consistently and the Ljung–Box test validates the proposed model, it can serve as a strong candidate. The proposed models should then be compared based on their predictive performance. We illustrate our methodology with an application to wind speed data and sea surface temperature anomalies, providing practical insights into the relevance of our findings.

**Keywords:** time series analysis; autocorrelation function (ACF); white noise; moving average; normality tests



Academic Editor: Wei Zhu

Received: 7 January 2025
Revised: 16 February 2025
Accepted: 18 February 2025
Published: 20 February 2025

**Citation:** Royer-Carenzi, M.; Hassani, H. Deviations from Normality in Autocorrelation Functions and Their Implications for MA(q) Modeling. *Stats* **2025**, *8*, 19. https://doi.org/10.3390/stats8010019



## 1. Introduction

Time series data, defined by their ordered and sequential nature, are prevalent in a wide range of disciplines, including economics [1–5], finance [6], medicine [7–9], climate science [10,11], and signal processing [12]. Effectively analyzing such data requires an understanding of their intrinsic temporal relationships, which often pose significant challenges. The Autocorrelation Function (ACF) plays a pivotal role in this context [13–24], serving as a fundamental statistical tool to measure the degree of dependency between observations separated by different time lags.

For time series consisting of uncorrelated observations, such as White Noise (WN), the ACF values are zero at all non-zero lags. In contrast, for processes such as moving average models of order $q$, denoted as MA($q$), the ACF becomes zero for lags greater than $q$. This distinctive property makes the ACF a valuable tool for identifying such underlying

models. However, the significance of ACF extends far beyond the distinction of WN or MA($q$) processes. It can also reveal nonstationary structural components in the data, such as trends [25–27] or periodic patterns [28]. Furthermore, the ACF is instrumental in identifying long-memory behaviors as explored in studies like [29–33].

Importantly, the ACF is crucial in model validation, as it permits to assess the adequacy of the residuals from fitted models. By examining the residuals' ACF, researchers can evaluate whether the chosen model adequately captures the series' underlying dynamics. The empirical distribution of the ACF therefore forms the basis for statistical inference in time series analysis. Practical modeling decisions often rely on the theoretical expectations of ACF behavior under different processes [16,23].

In this paper, we focus on MA($q$) processes, whose ACF, calculated beyond lag $q$, is theoretically expected to be asymptotically independent and converge to the same normal distribution. This property is commonly leveraged in practice for identifying an MA($q$) model. However, regardless of the observed time series, the sum of all the ACFs of the sample is constant, equal to $-\frac{1}{2}$ [22]. In this work, we investigate the sum of the sample ACFs for an MA($q$) process to demonstrate that this property contradicts the asymptotic normality of the ACF.

Numerous studies have emphasized the practical importance of Hassani's $-\frac{1}{2}$-Theorem and its integration into time series analysis and model development (e.g., [34–37]). The implications of this remarkable consistency are profound, particularly for building and analyzing time series models [38–40]. For recent investigations into the significance of the sample ACF in the context of Hassani's theorem, see [41–43]. Specifically, in our previous work [44], we analyzed the impact of Hassani's $-\frac{1}{2}$-Theorem on validating ARIMA($p, d, q$) processes using the Box–Pierce and Ljung–Box tests. Our results indicated that the successive ACFs of the residuals cannot be regarded as Gaussian realizations. As a consequence, both the Box–Pierce and Ljung–Box tests exhibit limitations: the former tends to be overly conservative, while the latter appears excessively liberal. To address these issues, we propose considering successive cumulative sums of ACF (SACF) as an alternative method to detect misspecified ARIMA models.

In this paper, we investigate the theoretical and empirical properties of the ACF and its cumulative sums for MA($q$) processes, not for their residuals. The primary goal is to examine the disparities between the theoretical expectations and the empirical behavior of these statistical measures. Our results reveal significant deviations from normality in the sample ACF beyond certain lags, even in MA($q$) processes of various lengths and under different underlying WN assumptions, whether Gaussian or non-Gaussian. These findings challenge traditional assumptions regarding ACF normality in time series modeling, particularly for MA($q$) identification.

Furthermore, we investigate the Extended Autocorrelation Function (EACF) and its utility in identifying the orders $p$ and $q$ of Autoregressive Moving Average (ARMA($p, q$)) models. As an extension of the ACF, the EACF provides additional information by facilitating the simultaneous identification of both autoregressive and moving average components. In its recursive construction, the ACFs of the regressive part are computed, and are expected to behave as for an MA($q$) process. Since the order $q$ of an MA($q$) process is theoretically identifiable based on the normality of the sample ACF beyond lag $q$, this study calls for a re-evaluation of the existing methodologies. We use simulated data to find out whether the procedures for identifying the $q$-order of an MA($q$) or the $(p, q)$-orders of an ARMA($p, q$) model are reliable.

The primary objective of this paper is to examine the extent to which the ACF deviates from normality and assess its consequences for model identification. By conducting extensive simulations and applying our methodology to real-world data, we aim to refine

existing order selection procedures and propose improvements to conventional statistical approaches. This study not only contributes to the theoretical understanding of ACF properties but also provides practical insights for time series practitioners.

The paper is organized as follows. Section 2 introduces the standard methods used for modeling time series, with a particular focus on the definition and properties of the Autocorrelation Function, both theoretical and empirical. Within this section, we describe the asymptotic behavior of ACF estimators in the context of MA($q$) processes and derive a practical rule for identifying the lag $q$ in MA($q$) models. We also introduce EACF estimators and derive a practical rule for identifying the lags $p$ and $q$ in ARMA($p, q$) models. We also explore the asymptotic behavior of the cumulative sums of ACF (SACF) in MA($q$) processes. Additionally, Hassani's $-\frac{1}{2}$-Theorem is revisited, and its contradictions with several results in Section 2 are highlighted, raising questions about the normality of the sample ACF itself. To assess the practical implications of these theoretical findings, we conduct extensive numerical simulations in Section 3. These experiments aim to quantify deviations from normality in sample ACF estimators and evaluate their impact on order selection methods. In particular, we test whether traditional ACF-based model identification techniques remain reliable when the underlying assumptions are not met. The results highlight significant deviations from normality, primarily due to the pronounced non-normality of SACFs. In this section, we also evaluate the performance of the $q$-order identification method on simulated MA(2)-series, showing that it is highly influenced by the value of the $a_2$ coefficient. When this parameter approaches zero, the method tends to suggest an MA(1) model instead. Section 4 extends these experiments to ARMA($p, q$) models, investigating the accuracy of the Extended Autocorrelation Function (EACF) method. We simulate ARMA(1,2) and ARMA(2,2) processes to assess for the reliability of the $(p, q)$-orders identification, based on EACFs. Section 5 applies the methodology to a real-world dataset of wind speed measurements in New York, illustrating the practical implications of the findings. Finally, Section 6 highlights the potential impact of our results on model selection strategies, particularly in cases where deviations from normality may influence order estimation.

## 2. Theoretical Background

### *2.1. Autocorrelation Functions (ACFs)*

For any square-integrable stationary process $(Z_t)_t$, we can consider its theoretical Autocorrelation Function (theoretical ACF)

$$\rho(h) \;=\; \mathrm{cor}(Z_{t+h},\, Z_t)\,, \quad h \in \mathbb{Z}. \tag{1}$$

Note that by definition, we have $\rho(0) \;=\; 1$. The most important example of stationary time series is White Noise (WN), denoted by $(\mathcal{E}_t)_t$ and defined as independent and identically distributed variables, with $\mathbb{E}(\mathcal{E}_t) \;=\; 0$ and $\mathbb{E}(\mathcal{E}_t^2) < \infty$. Then, its theoretical ACF is null for any lag $h \neq 0$. For more complex models such as MA($q$) processes, the theoretical ACF vanishes beyond lag $q$ as stated in the following Proposition:

**Proposition 1.**

*Let $\mathcal{E}_t$ be WN with $\mathbb{E}(\mathcal{E}_t^2) = \sigma^2$. We consider the MA(q) process*

$$Z_t \;=\; \sum_{k=0}^{q} a_k\, \mathcal{E}_{t-k}, \quad \textit{where } a_0 = 1\,.$$

$$\textit{Then,} \quad \rho(h) \;=\; \begin{cases} \sigma^2 \, \frac{\sum_{k=0}^{q-|h|} a_k \, a_{k+|h|}}{\sum_{k=0}^{q} a_k^2} & \textit{if } |h| \le q \\ 0 & \textit{if } |h| > q \,. \end{cases}$$

For a given realization $(z_1, \cdots, z_n)$, and a fixed value of $h = 1, \cdots, n-1$, let us define the sample ACF

$$\hat{\rho}(h) \;=\; \frac{\sum_{j=1}^{n-h} (z_{j+h} - \overline{z}) \, (z_j - \overline{z})}{\sum_{j=1}^{n} (z_j - \overline{z})^2} \,, \quad h = 1, \cdots, n-1 \,. \tag{2}$$

Once again, by definition, we have $\hat{\rho}(0) \;=\; 1$. Note that $\hat{\rho}(h)$ can be computed for any time series, whereas the theoretical ACF $\rho(h)$ are only defined for stationary series. We also consider the associated estimator $\hat{\Xi}(h)$, called the ACF estimator, defined in the same way as in Equation (2), by replacing the observed values $z_j$ by the random variables $Z_j$. ACF estimators satisfy the following fundamental property.

**Theorem 1.**
*Let $\mathcal{E}_t$ be WN with $\mathbb{E}(\mathcal{E}_t^4) < \infty$. We consider the associated MA(q) process*

$$Z_t \;=\; \sum_{k=0}^{q} a_k \, \mathcal{E}_{t-k}, \quad \textit{where } a_0 = 1 \,.$$

*Let us denote by $\mathcal{L}$ the convergence in distribution, by ${}^t v$ the transpose of vector $v$, and by $\underline{0}_H$ the null H-vector. Then,*

$$\sqrt{n} \, \left( {}^t(\hat{\Xi}(1), \hat{\Xi}(2), \cdots, \hat{\Xi}(H)) - \underline{m}_H \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}_H(\, \underline{0}_H, \, V_H) \,,$$

*where $\underline{m}_H \;=\; {}^t(m_1, \cdots, m_H)$ satisfies*

$$\begin{aligned} m_r \;&=\; \rho(r) \,, \quad \textit{for } 1 \le r \le \min(q, H) \\ &=\; 0 \,, \quad \textit{for } q+1 \le r \le H \textit{ when } H > q \,; \end{aligned}$$

*and where $V_H$ is a symmetric $H \times H$-matrix, with terms $v_{i,j}$ given by Bartlett's formula [45]:*

$$\begin{aligned} v_{i,j} \;=\; &\sum_{k=1}^{\infty} (\rho(k+i) \;+\; \rho(k-i) \;-\; 2\,\rho(i)\,\rho(k)) \times \\ &\sum_{k=1}^{\infty} (\rho(k+j) \;+\; \rho(k-j) \;-\; 2\,\rho(j)\,\rho(k)). \end{aligned} \tag{3}$$

Theorem 1 is a particular case of Theorems 7.2.1. or 7.2.2. in [16]. And we can explicitly define several terms of the covariance matrix.

**Proposition 2.**

*Let us introduce*

$$v_0 \;=\; 1 \;+\; 2 \sum_{h=0}^{q} \rho(h)^2 \,, \tag{4}$$

$$\begin{aligned} v_r \;&=\; 2 \sum_{h=0}^{q-r} \rho(h) \, \rho(h+r) \,, \quad \textit{for } \; 1 \le r \le q \\ &=\; 0 \,, \quad \textit{for } \; r > q \,. \end{aligned} \tag{5}$$

*Then*

$$\begin{cases} v_{i,i+r} &= \nu_r \quad \textit{if } 1 \le i \le \min(q,H)\,,\ q-i+1 \le r \le H-i \\ v_{i,i+r} &= \nu_r \quad \textit{if } q+1 \le i \le H\,,\ 0 \le r \le H-i\,, \end{cases}$$

*and when* $H > q$,

$$v_{q,q} \;=\; \nu_0 \left(1 + 2\,\rho(q)^2\right) \;-\; 8\,\rho(q)^2\,.$$

In other words, the first terms $(v_{i,j})_{1\le i\le q,\ 1\le j\le q}$ are computed from Equation (3), which does not simplify further. However, for indices where either $i > q$ or $j > q$, the matrix takes on a specific form with only $q+2$ distinct values. The diagonal elements are given by $\nu_0$, the first off-diagonal elements, positioned at $|i-j| = 1$, which take the value $\nu_1$. Similarly, the second off-diagonal elements, for which $|i-j| = 2$, are equal to $\nu_2$, and so forth. More generally, the elements at $|i-j| = k$ are equal to $\nu_k$ for $k \le q$. Beyond the $q$-th off-diagonal, all elements are zero, reflecting the truncation of the theoretical ACF beyond lag $q$. As an example, the covariance matrix associated to the H-vector of the ACF estimators of an MA(2) process is made explicit in Appendix A.1.

Finally, Theorem 1 states that the vector of the ACF estimators is asymptotically multivariate Gaussian. Let us denote by $\left(\boldsymbol{\mathcal{A}_H}(\underline{\mu}, \boldsymbol{\Sigma})\right)$ the property that ${}^t(\hat{\Xi}(1), \hat{\Xi}(2), \cdots, \hat{\Xi}(H))$ is asymptotically multivariate Gaussian with the expectation that the $H$-vector $\underline{\mu}$, and with covariance matrix $\frac{\Sigma}{n}$, is symmetric and definite-positive.

### *2.2. Identification of Order q in an MA(q) Model*

From Theorem 1 and Proposition 2, we have the following:

- When $h \ge q+1$,

$$\sqrt{n}\,\hat{\Xi}(h) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\ \nu_0). \tag{6}$$

As a consequence, if $(Z_t)_t$ is an MA($q$) process, then $\forall\ h \ge q+1$, and for a large $n$, we should observe that $\hat{\rho}(h) \in \mathcal{J} := \left[-1.96\,\sqrt{\frac{\nu_0}{n}}\,;\,1.96\,\sqrt{\frac{\nu_0}{n}}\right]$.

- When $h = q$,

$$\sqrt{n}\,\left(\hat{\Xi}(q) - \rho(q)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0,\ v_{q,q}\right). \tag{7}$$

As a consequence, if $(Z_t)_t$ is an MA($q$) process, then for a large $n$, we should observe that $\hat{\rho}(q) \in \left[\rho(q) - 1.96\,\sqrt{\frac{v_{q,q}}{n}}\,;\,\rho(q) + 1.96\,\sqrt{\frac{v_{q,q}}{n}}\right]$. We recall that one of the essential characterizations of the order $q$ of an MA($q$) process is that its ACF $\rho(q)$ is non-zero.

In practice, the thresholds provided by Equations (6) and (7) are unknown, but they can be estimated by replacing $\rho(h)$ by $\hat{\rho}(h)$.

**Lemma 1.**

*Let us set*

$$\hat{\lambda}(h) \;=\; 1 \;+\; 2\,\sum_{k=1}^{h} \hat{\rho}(k)^2\,. \tag{8}$$

*Then, we have*

$$\hat{\lambda}(h) \xrightarrow[n\to+\infty]{} \nu_0\,, \quad \textit{for } h \ge q\,. \tag{9}$$

**Proof.** By convergence of $\hat{\rho}(h)$ towards $\rho(h)$ for any $h \le q$, it is obvious that $\hat{\lambda}(q)$ approaches $\nu_0$ when n is large. But this is also true for $\hat{\lambda}(h)$ whatever $h \ge q$, since $\hat{\rho}(k)$ tends towards 0 when $k > q$. □

**Proposition 3.**

*We suggest to identify the order q of an MA(q) process with the following procedure:*

*(i) For any h such that* $1 \leq h \leq \frac{n}{4}$,

*check if* $\hat{\rho}(h)$ *lies in the interval* $\hat{\mathcal{J}}_h := \left[-1.96\sqrt{\frac{\hat{\lambda}(h)}{n}}\,;\,1.96\sqrt{\frac{\hat{\lambda}(h)}{n}}\right]$.

*(ii) A candidate* $\hat{q}$ *for order q is the last lag h with* $|\hat{\rho}(h)| > 1.96\sqrt{\frac{\hat{\lambda}(h)}{n}}$.

*We provide a graphical R-function of this procedure on our website. It is called* `acfMA()`.

**Proof.**

- *If n is sufficiently large, when* $h > q$, *we have*

  $|\hat{\rho}(h)| < 1.96\sqrt{\frac{v_0}{n}} \simeq 1.96\sqrt{\frac{\hat{\lambda}(h)}{n}}$, *so that* $\hat{\rho}(h) \in \hat{\mathcal{J}}_h$.

- *When* $h = q$, *from Proposition 1, we have* $\rho(q) = \sigma^2 \frac{a_q}{\sum_{k=0}^{q} a_k^2} \neq 0$.

  *Then, for sufficiently large n, we have*

  *either* $\rho(q) > 1.96\sqrt{\frac{v_{q,q}}{n}} + 1.96\sqrt{\frac{\hat{\lambda}(q)}{n}}$, *if* $a_q > 0$,

  *or* $\rho(q) < -1.96\sqrt{\frac{v_{q,q}}{n}} - 1.96\sqrt{\frac{\hat{\lambda}(q)}{n}}$, *if* $a_q < 0$,

  *implying in both cases that* $\hat{\rho}(q) \notin \hat{\mathcal{J}}_q$.

□

The use of adjusted significance thresholds, instead of the traditional $[-1.96/n,\ 1.96/n]$, accounts for the cumulative variance of residual autocorrelations in MA($q$) models. This correction reduces the false detections of MA orders by incorporating the actual error structure. As a result, it improves the robustness of order selection, particularly for small sample sizes.

### *2.3. Identification of Orders* $(p, q)$ *in an ARMA(p, q) Model*

An Autoregressive Moving Average model of order $(p, q)$, denoted ARMA$(p, q)$, is defined as

$$Z_t = b_1 Z_{t-1} + b_2 X_{t-2} + \cdots + b_p Z_{t-p} + \mathcal{E}_t - a_1 \mathcal{E}_{t-1} - \cdots - a_q \mathcal{E}_{t-q}, \tag{10}$$

where $b_1, \ldots, b_p$ are the autoregressive (AR) coefficients, $a_1, \ldots, a_q$ are the moving average (MA) coefficients, and $\mathcal{E}_t$ is a White Noise process with mean zero and variance $\sigma^2$. Identifying the correct orders $(p, q)$ of an ARMA$(p, q)$ model is critical for effective modeling and forecasting.

The standard Autocorrelation Function (ACF) is a valuable tool for identifying the order of an MA($q$) process, as it exhibits a well-defined cutoff property: the ACF of an MA($q$) model vanishes for lags greater than $q$. However, for Autoregressive Moving Average (ARMA$(p, q)$) processes, the ACF alone is insufficient for order identification. Indeed, the ACF of the ARMA$(p, q)$ process does not exhibit a clear-cut truncation. Instead, it follows an exponential or damped sinusoidal decay pattern, making it difficult to directly infer the presence and order of the moving average or autoregressive components.

One practical approach is to use the Extended Autocorrelation Function (EACF), which extends the traditional Autocorrelation Function (ACF). By iteratively computing residual autocorrelations after removing autoregressive effects up to a given lag $p$, the EACF allows for the simultaneous identification of both $p$ and $q$ in an ARMA$(p,q)$ model.

#### 2.3.1. Principle of the EACF

**Proposition 4.** *The EACF method is introduced in [46] recursively as follows:*

1. *First, for each candidate AR order k, the series is regressed linearly on its k lagged values:*

$$Z_t = \beta_1^{(k,0)} Z_{t-1} + \beta_2^{(k,0)} Z_{t-2} + \cdots + \beta_k^{(k,0)} Z_{t-k} + \epsilon_t^{(k,0)}.$$

*Here, the coefficients $\beta_1^{(k,0)}, \ldots, \beta_k^{(k,0)}$ are estimated by ordinary least squares, and the residuals $\widehat{\epsilon}_t^{(k,0)}$ are derived. Next, we iterate the regressions, including the successive residuals:*

$$Z_t = \sum_{l=1}^{k} \beta_l^{(k,j)} Z_{t-l} + \sum_{i=1}^{j} \alpha_i^{(k,i)} \widehat{\epsilon}_{t-i}^{(k,j-i)} + \epsilon_t^{(k,j)}.$$

2. *Let us compute the extended residuals $W_t^{(k,j)} = Z_t - \sum_{l=1}^{k} \beta_l^{(k,j)} Z_{t-l}$. The Extended Autocorrelation Function (EACF) is defined as*

$$r_j^{(k)} = \rho_{W_t^{(k,j)}}(j).$$

*If $Z_t$ is an ARMA$(p,q)$ process, then $W_t^{(p,q+h)}$ is an MA$(q)$ process, $\forall h \geq 0$. Consequently, $r_q^{(p)} \neq 0$ and $r_{q+h}^{(p)} = 0$, $\forall h \geq 1$.*

3. *We suggest to test the significance of the EACF $r_j^{(k)}$ referring to the thresholds*

$$\pm 1.96 \sqrt{\frac{\widehat{\lambda}_{W_t^{(k,j)}}(j)}{n}},$$

*where $\widehat{\lambda}_{W_t^{(k,j)}}$, defined in Equation (8), uses the sample ACFs of the process $(W_t^{(k,j)})_t$.*

4. *The EACF matrix is constructed, with value $r_j^{(i-1)}$ in the cell of coordinates $(i,j)$. Each cell is marked as "o" if the autocorrelation is within the thresholds or "×" if it exceeds the thresholds.*

The steps in the algorithm for calculating the EACF are given in Appendix B.

#### 2.3.2. Interpretation of the EACF Matrix

In the EACF matrix, rows correspond to $p$ (AR orders) and columns to $q$ (MA orders), where $p$ and $q$ indices start at 0. Let us denote by $\tau(p,q)$ the EACF, related to orders $p$ and $q$, computed at row $i = p+1$ and column $j = q+1$. Then, $\tau(p,q) = r_{q+1}^{(p)}$. The identification of the underlying $(p,q)$ orders is based on locating the upper-left corner of the largest triangle of "o"s in the EACF matrix. This corner corresponds to the simplest ARMA$(p,q)$ model that adequately explains the time series structure. This approach ensures both parsimony and accuracy in model selection. Table 1 illustrates the expected EACF for an ARMA(1,1) process.

**Table 1.** Expected EACF for an ARMA(1,1) process.

| AR \ MA | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 0 | x | x | x | x | x | x |
| 1 | x | o | o | o | o | o |
| 2 | x | x | o | o | o | o |
| 3 | x | x | x | o | o | o |
| 4 | x | x | x | x | o | o |
| 5 | x | x | x | x | x | o |

2.3.3. Remark on the Adjusted Thresholds in Step 3)

In the computation of the Extended Autocorrelation Function (EACF) suggested in Proposition 4, the introduction of the correction factor $\widehat{\lambda}_{W_t^{(k,j)}}(j)$ plays a crucial role in refining the selection of ARMA orders. Indeed, it allows for a more accurate estimation of significance thresholds by accounting for dependencies within the time series. This adjustment ensures that the statistical evaluation of autocorrelation coefficients properly reflects the structure of the data rather than assuming independence.

*2.4. Normality of the Sum of Sample Autocorrelation Functions (SACF)*

Let us define the partial sum of the sample ACF values (SACF),

$$S_{ACF}^{sample}(H) \;=\; \sum_{h=1}^{H} \hat{\rho}(h)\,, \tag{11}$$

and in the same way, we will call the SACF estimator and denote by $S_{ACF}^{estim.}(H)$, the sum of the associated ACF estimators. At any lag H, the associated SACF is a linear transformation of the random vector ${}^t(\hat{\Xi}(1), \hat{\Xi}(2), \cdots, \hat{\Xi}(H))$. It is well known that any linear transformation of a multivariate Gaussian vector remains Gaussian. More precisely, we have the following proposition.

**Proposition 5.**
*Let $\underline{Y} \;=\; {}^t(Y_1, \cdots, Y_r)$ be a multivariate Gaussian vector, with distribution $\mathcal{N}_r(\underline{\mu}, \Sigma)$, where $\underline{\mu}$ is a r-vector and $\Sigma$ is an $r \times r$-matrix, symmetric, and definite positive.*
*Then, for any matrix $A$ in $\mathbb{R}^{p\times r}$, we have*

$$A\,{}^t(Y_1, \cdots, Y_r) \;\sim\; \mathcal{N}_p(A\,\underline{\mu},\; A\,\Sigma\,{}^tA)\,.$$

Note that $\Sigma$ being symmetric and definite positive implies that $A\,\Sigma\,{}^tA$ is also symmetric and definite positive. Since the vector of the H first SACF estimators is a linear transformation of the random vector ${}^t(\Xi(1), \Xi(2), \cdots, \Xi(H))$ with a matrix $A_H$ being a squared and unitary lower-triangular matrix, then we obtain the following Theorem.

**Theorem 2.**
*Let $\mathcal{E}_t$ be WN with $\mathbb{E}(\mathcal{E}_t^4) < \infty$. We consider the associated MA(q) process*

$$Z_t \;=\; \sum_{k=0}^{q} a_k\,\mathcal{E}_{t-k}, \quad \textit{where } a_0 = 1\,.$$

$$\sqrt{n}\,\left({}^t(S_{ACF}^{estim.}(1), S_{ACF}^{estim.}(2), \cdots, S_{ACF}^{estim.}(H)) - \underline{\mu}_H\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}_H(\underline{0}_H,\, W_H)\,,$$

*where* $\underline{\mu}_H = {}^t(\mu_1, \cdots, \mu_H)$ *satisfies*

$$\begin{aligned} \mu_r &= \sum_{h=1}^{r} \rho(h)\,, \quad \textit{for } 1 \le r \le q \\ &= \sum_{h=1}^{q} \rho(h)\,, \quad \textit{for } q \le r \le H\,; \end{aligned}$$

*and where the terms of* $W_H$ *are* $w_{i,j} = \sum_{k=1}^{i} \sum_{l=1}^{j} v_{k,l}$.

**Proof.** Note that we take $A_H$ as the unitary lower-triangular $H \times H$-matrix, satisfying $det(A_H) = 1$. Then, from Cholesky's decomposition, $W_H = A_H\, {}^tA_H$ is a positive definite matrix. So, we can apply Proposition 5. □

Let us denote by $\left(\boldsymbol{\mathcal{S}_H(\underline{\mu}, \Sigma)}\right)$ the property that ${}^t(S_{ACF}^{estim.}(1), S_{ACF}^{estim.}(2), \cdots, S_{ACF}^{estim.}(H))$ is asymptotically multivariate Gaussian with expectation the $H$-vector $\underline{\mu}$, and with covariance matrix $\frac{\Sigma}{n}$, symmetric and definite positive. From Theorem 2, if $n$ is large, $\left(\boldsymbol{\mathcal{S}_H(\underline{\mu}_H, W_H)}\right)$ is satisfied.

### 2.5. Contradiction with Normality Property

By the definition of parameters $\nu_r$ in Proposition 2, for $k \ge q$, we have

$$w_{k,k} = \sum_{i,j=1}^{k} v_{i,j} = \sum_{i,j=1}^{q} v_{i,j} + (k-q)\left(\nu_0 + 2\sum_{r=1}^{q} \nu_r\right).$$

If $\left(\boldsymbol{\mathcal{S}_{n-1}(\underline{\mu}_{n-1}, W_{n-1})}\right)$ were true (with $q \le n-1$), we would obtain for $n$ large,

$$\begin{aligned} S_{ACF}^{estim.}(n-1) &\sim \mathcal{AN}\left(\sum_{h=1}^{q} \rho(h),\ \frac{\sum_{i,j=1}^{q} v_{i,j}}{n} + \frac{n-q-1}{n}\left(\nu_0 + 2\sum_{r=1}^{q} \nu_r\right)\right) \\ &\sim \mathcal{AN}\left(\sum_{h=1}^{q} \rho(h),\ \nu_0 + 2\sum_{r=1}^{q} \nu_r\right). \end{aligned} \tag{12}$$

where the symbol $\sim$ means *"follows the distribution"* and $\mathcal{AN}$ means *"asymptotically Gaussian"*. But [22] proved that if $n \ge 2$,

$$S_{ACF}^{sample}(n-1) = -\frac{1}{2}\,, \tag{13}$$

for any stationary time series, so, in particular, for MA($q$) processes. Actually, note that this result holds true for any time series with $n \ge 2$, even for the nonstationary ones. Since a Gaussian variable $\mathcal{N}(\mu, \sigma^2)$ can be equal to the constant value $-\frac{1}{2}$, only if $\mu = -\frac{1}{2}$ and in the degenerate case $\sigma^2 = 0$, Equations (12) and (13) imply that

$$\begin{cases} \sum_{h=1}^{q} \rho(h) = -\frac{1}{2} \\ \nu_0 + 2\sum_{r=1}^{q} \nu_r = 0\,. \end{cases}$$

But

$$\begin{aligned} \nu_0 + 2\sum_{r=1}^{q}\nu_r &= 1 + 2\sum_{h=1}^{q}\rho(h)^2 + 4\sum_{k<l}\rho(k)\,\rho(l) \\ &= 1 + 2\left(\sum_{h=1}^{q}\rho(h)\right)^2 + 4\sum_{h=1}^{n-1}\rho(h) \\ &= 1 + 2\times\left(-\frac{1}{2}\right)^2 - 2 = -\frac{1}{2} \neq 0\,. \end{aligned}$$

Consequently, Equations (12) and (13) are contradictory. As a consequence, the Property $\left(\boldsymbol{\mathcal{S}_H}(\underline{\boldsymbol{\mu}}_{\boldsymbol{H}}, \boldsymbol{W_H})\right)$ does not hold for $H = n-1$, and then, neither does $\left(\boldsymbol{\mathcal{A}_{n-1}}(\underline{\boldsymbol{m}}_{\boldsymbol{n-1}}, \boldsymbol{V_{n-1}})\right)$.

However, let us suppose that $\left(\boldsymbol{\mathcal{A}_{n-2}}(\underline{\boldsymbol{m}}_{\boldsymbol{n-2}}, \boldsymbol{V_{n-2}})\right)$ is true. In particular, ${}^t(\hat{\Xi}(1), \hat{\Xi}(2), \cdots, \hat{\Xi}(n-2))$ is supposed to be asymptotically Gaussian multivariate. Since from Equation (13) we have

$$\hat{\Xi}(n-1) = -\frac{1}{2} - \sum_{h=1}^{n-2}\hat{\Xi}(h)\,. \tag{14}$$

Then, by the definition of Gaussian vectors, ${}^t(\hat{\Xi}(1), \hat{\Xi}(2), \cdots, \hat{\Xi}(n-1))$ would also have an asymptotic multivariate distribution:

$${}^t(\hat{\Xi}(1), \hat{\Xi}(2), \cdots, \hat{\Xi}(n-2), \hat{\Xi}(n-1)) \sim \mathcal{AN}_{n-1}\left(\underline{m}'_{n-1}; \frac{V'_{n-1}}{n}\right),$$

with $\underline{m}'_{n-1} = \underline{m}_{n-1} - {}^t\left(0, \cdots, 0, -\frac{1}{2}\right)$ and

$$V'_{n-1} = \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n-2} & -\sum_{j=1}^{n-2} v_{1,j} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n-2} & -\sum_{j=1}^{n-2} v_{2,j} \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ v_{n-2,1} & v_{n-2,2} & \cdots & v_{n-2,n-2} & -\sum_{j=1}^{n-2} v_{n-2,j} \\ -\sum_{i=1}^{n-2} v_{i,1} & -\sum_{i=1}^{n-2} v_{i,2} & \cdots & -\sum_{i=1}^{n-2} v_{i,n-2} & -\sum_{i=1}^{n-2}\sum_{j=1}^{n-2} v_{i,j} \end{pmatrix}.$$

But this matrix is not definite positive since $(1, \cdots, 1)\, V'_{n-1}\, {}^t(1, \cdots, 1) = 0$. Thus, $\left(\boldsymbol{\mathcal{A}_{n-2}}(\underline{\boldsymbol{m}}_{\boldsymbol{n-2}}, \boldsymbol{V_{n-2}})\right)$ is not true either. Actually, Theorem 1, being an asymptotical result, needs $H$ to remain lower than $n-1$ so that $\frac{H}{n}$ converges to 0. For instance, [47] recommends to take a sufficiently long time series ($n \geq 40$) and to consider only $H \leq \sqrt{n}$. So $\left(\boldsymbol{\mathcal{A}_H}(\underline{\boldsymbol{m}}_{\boldsymbol{H}}, \boldsymbol{I_H})\right)$ should be true until $H \leq \sqrt{n}$. And so Theorem 2 should also hold for $H \leq \sqrt{n}$.

## 3. Numerical Results for Simulated MA(q) Processes

To investigate the normality of the sample ACF and of the SACF, we simulate $N_S = 5000$ MA(2) processes with the equation

$$Z_t = \mathcal{E}_t + \frac{1}{2}\mathcal{E}_{t-1} + a_2\,\mathcal{E}_{t-2}\,,$$

where $(\mathcal{E}_t)_t$ is either a Gaussian or an Exponential WN, with length $n = 500$. Note that it suffices to consider standard White Noise processes. Indeed, from the definitions of theoretical and sample ACFs, given in Equations (1) and (2), dividing the underlying process $(\mathcal{E}_t)_t$ by its standard deviation does not alter the ACF values. In each simulation, sample ACF and SACF values are computed for lags $h = 1, \ldots, n-1$. To test for normality, we employ the Shapiro–Wilk test [48] and verify the results using the Lilliefors (composite Kolmogorov–Smirnov) normality test [49], which yields consistent outcomes. Unlike Shapiro–Wilk and Lilliefors tests, which assess sample normality regardless of the expectation and variance values, Kolmogorov–Smirnov's test [50,51] allows direct comparison to the theoretical distribution $\mathcal{N}(\mu, \sigma^2)$, where $\mu$ and $\sigma^2$ can be computed from the parameters used for simulations. This approach provides a comprehensive evaluation of normality in the simulated sample ACF and SACF values.

We successively consider two distinct values for coefficient $a_2$: first $a_2 = \frac{1}{2}$ and next $a_2 = \frac{1}{10}$. The main difference between these MA(2) processes is the proximity of their associated $\rho(2)$ value with 0, and so their ability to be confounded with an MA(1) process. In Appendix A.2, we detail the relationship between the coefficient $a_q$ of an MA($q$) process and the ACF value $\rho(q)$.

### 3.1. Normality of $\hat{\Xi}(h)$ at a Fixed Lag h

For MA(2) processes, the $\hat{\Xi}(h)$ variables are not supposed to be independent, unless they involve indexes $i$ and $j$ such that $j - i > q = 2$. Furthermore all the $\hat{\Xi}(h)$ variables are not identically distributed. Consequently, we cannot test the normality of the set $\hat{\rho}(1), \cdots, \hat{\rho}(H)$, for a fixed simulation. We only test the normality of the $N_S$ values $\hat{\rho}(h)$, at a fixed lag h. At a given lag $h$, the Kolmogorov–Smirnov test evaluates the fit of $\hat{\Xi}(h)$ to the Gaussian distribution $\mathcal{N}\left(\rho(h), \frac{v_{h,h}}{n}\right)$, while the Shapiro–Wilk test directly assesses its normality. Figure 1 displays the $p$-values from the Kolmogorov–Smirnov test applied to $N_S = 200$ and $N_S = 5000$ simulations of MA(2) processes with $a_2 = \frac{1}{2}$ and length $n = 500$. For comparison, Figure 2 presents the $p$-values from the Shapiro–Wilk test. Similar analyses are conducted for MA(2) processes with $a_2 = \frac{1}{10}$, with very similar results, included in the Supplementary Materials.

The upper panels of Figure 1 indicate that $\hat{\Xi}(h)$ fits well with the Gaussian distribution $\mathcal{N}\left(\rho(h), \frac{w_{h,h}}{n}\right)$ as expected from Theorem 1, even for lags $h$ much greater than $\sqrt{n}$. However, the bottom panels demonstrate that as the number of simulations increases, $\hat{\Xi}(h)$ increasingly deviates from this specific normal distribution. This behavior reflects the asymptotic nature of Theorem 1, which ensures that $\hat{\Xi}(h)$ approximates the distribution $\mathcal{N}\left(\rho(h), \frac{v_{h,h}}{n}\right)$ for large $n$. Its slight departure from this distribution is more easily detectable with a great number of simulations.

Nevertheless, focusing on the Gaussian behavior of the ACF estimators $\hat{\Xi}(h)$, Figure 2 highlights a robust normality pattern, especially when the underlying WN is Gaussian. In this case, Shapiro–Wilk tests almost never reject the normality hypothesis, regardless of the lag $h < n - 2$, even with an extensive number of simulations ($N_S = 5000$). Conversely, for an underlying Exponential WN, the normality property rapidly deteriorates as the number of simulations increases.

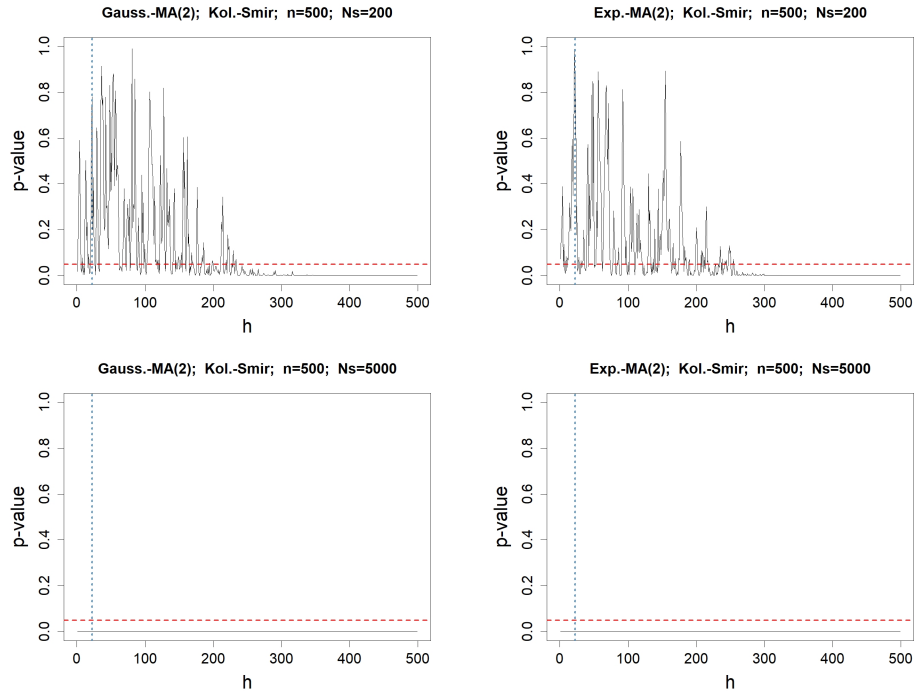


**Figure 1.** $p$-values when testing for the adequacy of the $N_S$ values of $\hat{\rho}(h)$ with $\mathcal{N}\left(\rho(h), \frac{v_{h,h}}{n}\right)$, for any fixed lag $h$ varying from 1 to $n-1$. The involved normality test is Kolmogorov–Smirnov's. The left column concerns MA(2) driven by a Gaussian WN, whereas the right one deals with the Exponential WN process. The length of the simulated WN process is $n = 500$. In the top figures, the number of simulated MA(2) processes is $N_S = 200$, whereas it is $N_S = 5000$ in the bottom. The red-dotted horizontal line represents 5%, while the blue-dotted vertical line represents $h = \sqrt{n}$.

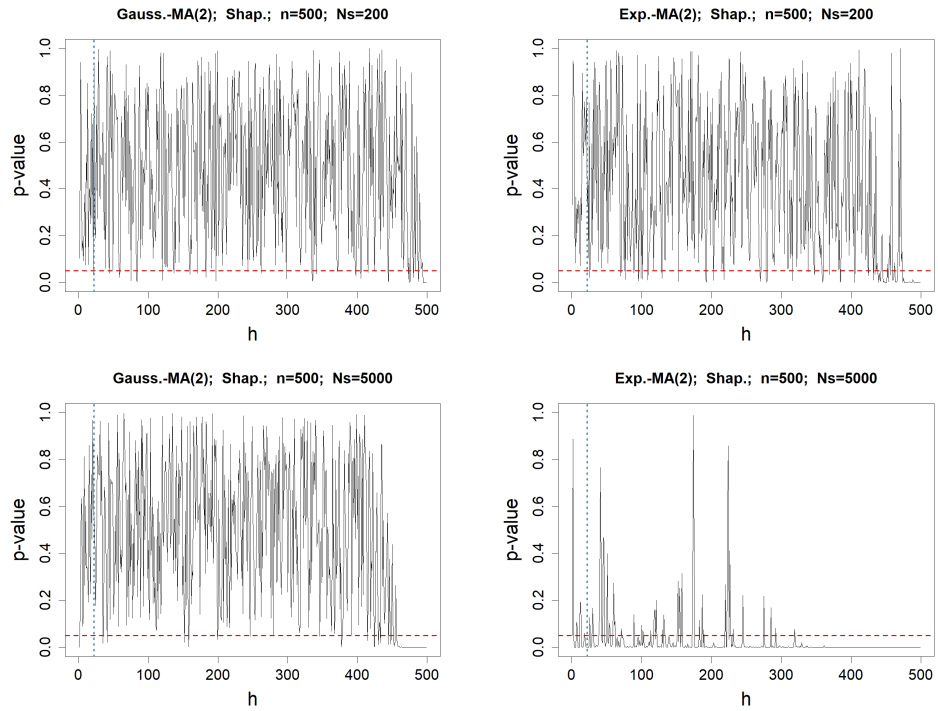


**Figure 2.** $p$-values when testing for the normality of the $N_S$ values of $\hat{\rho}(h)$, for any fixed lag $h$ varying from 1 to $n-1$. The involved normality test is Shapiro–Wilk's. The left column concerns MA(2) driven by a Gaussian WN, whereas the right one deals with the Exponential WN process. The length of the simulated WN process is $n = 500$. In the top figures, the number of simulated MA(2) processes is $N_S = 200$, whereas it is $N_S = 5000$ in the bottom. The red-dotted horizontal line represents 5%, while the blue-dotted vertical line represents $h = \sqrt{n}$.

From Figures 1 and 2, it can be inferred that the lack of adequacy of $\hat{\Xi}(h)$ with the distribution $\mathcal{N}\left(\rho(h), \frac{v_{h,h}}{n}\right)$ is partly due to its asymptotic nature. This explains its more pronounced deviation under Exponential WN. However, the issue also stems from a misspecification of the expectation and variance, as the normality behavior is preserved for Gaussian WN. In fact, $\hat{\Xi}(h)$ may be asymptotically Gaussian, but with either $\mu \neq \rho(h)$ or $\sigma^2 \neq \frac{v_{h,h}}{n}$. Consequently, we find that the condition $(\boldsymbol{\mathcal{A}_H(\underline{m}, V)})$ does not hold.

### *3.2. Normality of SACF*

To assess the normality of SACFs at a given lag $H$, we analyze the $N_S$ computed values of $S_{ACF}^{sample}(H)$. Figure 3 shows the $p$-values from the Shapiro–Wilk test applied to $N_S = 200$ and $N_S = 5000$ simulations of MA(2) processes with $a_2 = \frac{1}{2}$ and length $n = 500$. Similar tests are also conducted on MA(2) processes with $a_2 = \frac{1}{10}$, yielding comparable results, as documented in the Supplementary Materials. In Figure 3, the sum of sample ACF values $S_{ACF}^{sample}(H)$ deviates significantly from normality for nearly all lags $H$, except possibly for the first few lags when the underlying WN is Gaussian or when the number of simulations is relatively low. Moreover, Kolmogorov–Smirnov tests confirm that $S_{ACF}^{sample}(H)$ does not align with the theoretical distribution $\mathcal{N}\left(\sum_{h=1}^{H}\rho(h), \frac{w_{H,H}}{n}\right)$ at any lag $H$, regardless of the nature of the underlying WN (Gaussian or Exponential) or the number of simulations. This departure from $\mathcal{N}\left(\sum_{h=1}^{H}\rho(h), \frac{w_{H,H}}{n}\right)$ can be attributed to earlier findings that $\hat{\Xi}(h)$ does not fully converge to a Gaussian distribution with $\mu = \rho(h)$ and $\sigma^2 = \frac{v_{h,h}}{n}$. Thus, Figure 3 clearly indicates that none of the variables $S_{ACF}^{estim.}(H)$ may follow a Gaussian distribution. This implies that $\left(\boldsymbol{\mathcal{S}_H(\underline{\mu}, \Sigma)}\right)$ cannot hold true, irrespective of $\mu$ and $\Sigma$. Consequently, the Gaussianity of the vector ${}^t(\hat{\Xi}(1), \cdots, \hat{\Xi}(H))$ is also called into question. Indeed, if $\left(\boldsymbol{\mathcal{A}_H(\underline{\mu}, \Sigma)}\right)$ were true, $\left(\boldsymbol{\mathcal{S}_H(\underline{\mu}, \Sigma)}\right)$ would also be.

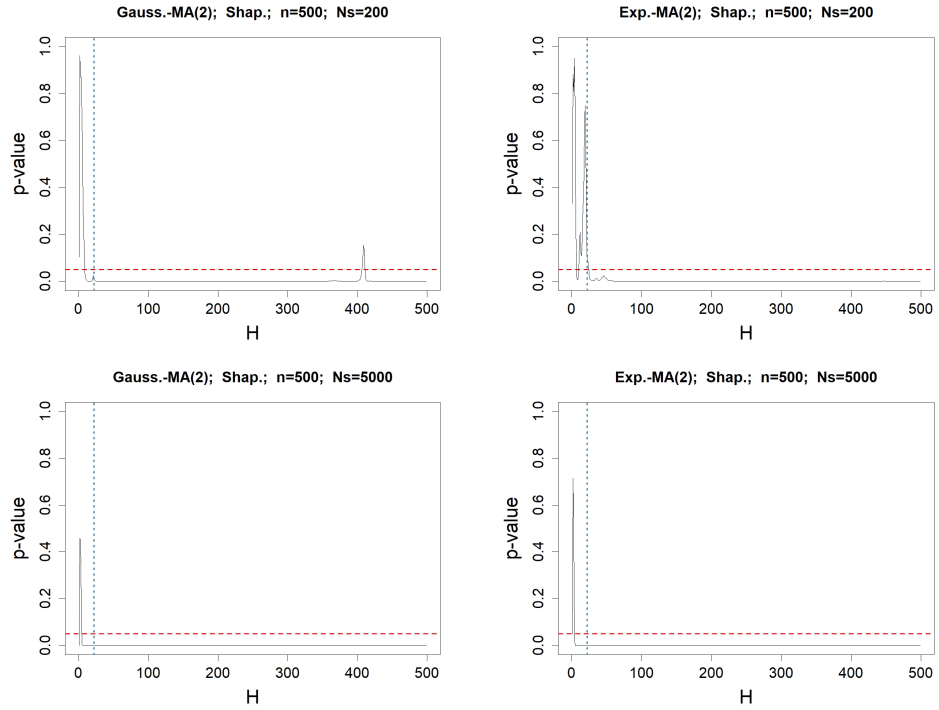


**Figure 3.** $p$-values when testing for the normality of the $N_S$ values of $S_{ACF}^{sample}(H)$, for any fixed lag $H$ varying from 1 to $n-1$. The involved normality test is Shapiro–Wilk's. The left column concerns MA(2) driven by a Gaussian WN, whereas the right one deals with the Exponential WN process. The length of the simulated WN process is $n = 500$. In the top figures, the number of simulated MA(2) processes is $N_S = 200$, whereas it is $N_S = 5000$ in the bottom. The red-dotted horizontal line represents 5%, while the blue-dotted vertical line represents $H = \sqrt{n}$.

Consequently, from Figures 2 and 3, we conclude that at a fixed lag $h$, $\hat{\Xi}(h)$ is roughly Gaussian, with $\mu \simeq \rho(h)$ or $\sigma^2 \simeq \frac{v_{h,h}}{n}$. But the vector ${}^t(\hat{\Xi}(1), \cdots, \hat{\Xi}(H))$ is not a Gaussian vector. This brings into question the relevance of using ACF to identify the order $q$ of an MA($q$) process.

### *3.3. Impact on q-Order Identification*

The order-identification procedure described in Proposition 3 is very simple, but is it really reliable? Indeed, it is based on the fact that the ACF estimator $\hat{\Xi}(h)$ asymptotically follows a normal distribution $\mathcal{N}\left(\rho(h), \frac{v_{h,h}}{n}\right)$, for any lag $h \geq q$, whereas we show in Section 3.1 that this property is not guaranteed, especially for MA($q$) processes with an exponential underlying WN.

In our procedure, the thresholds defining $\hat{\mathcal{J}}_h$ are adapted to MA($q$) processes, in opposition to the classical ACF graphics that use the thresholds $\pm 1.96\sqrt{\frac{1}{n}}$. Thus, for any lag $h > q$, every sample ACF value $\hat{\rho}(h)$ should belong to $\hat{\mathcal{J}}_h$, except in $\alpha = 5\%$ of the simulations led under the null hypothesis (generation of an MA($q$) process). On the contrary, $\hat{\rho}(q)$ should be larger (in absolute value) than the associated MA-threshold at that given lag, when $n$ is sufficiently large.

Here, for any simulated MA(2) process, we compute the associated MA-thresholds, from $h = 1$ to $\frac{n}{4}$. And we run the following steps:

**Step 1.** We check if $\hat{\rho}(3) \in \hat{\mathcal{J}}_3$ and compute the percentage of true diagnosis.
**Step 2.** We check if $\hat{\rho}(2) \notin \hat{\mathcal{J}}_2$ and we compute the percentage of true diagnosis.
**Step 3.** Check if $\hat{\rho}(2) \notin \hat{\mathcal{J}}_2$ and $\hat{\rho}(3) \in \hat{\mathcal{J}}_3$. And we compute the percentage of true diagnosis (both properties must be satisfied).
**Step 4.** We compute the largest lag $h$ such that $\hat{\rho}(h) \notin \hat{\mathcal{J}}_h$. We count each last-lag occurrence.

Table 2 gives the percentage of true diagnosis for **Steps 1** to **3**, when the underlying WN is Gaussian, among $N_S = 5000$ simulations. Simulations with an Exponential WN give very similar percentages. As expected in **Step 1**, we obtain that the $\hat{\rho}(3)$ behaves conveniently, except in 5% of the simulations, which conforms with type-I error. As expected in **Step 2**, we obtain that the frequency of convenient $\hat{\rho}(2)$ behavior improves when the time series length $n$ increases. Moreover, the performance highly depends on the $a_2$ value. As expected in **Step 3**, the performance of the paired properties is smaller than that of the singles.

**Table 2.** Percentage of true diagnosis among $N_S = 5000$ simulations of an MA(2) process, when checking **Steps 1** to **3**. The associated WN is Gaussian and the coefficient $a_2$ is either $a_2 = \frac{1}{2}$ (left column) or $a_2 = \frac{1}{10}$ (right column).

| | $a_2 = \frac{1}{2}$ | | $a_2 = \frac{1}{10}$ | |
|---|---|---|---|---|
| **n** | **100** | **500** | **100** | **500** |
| **Step 1** | 94.3% | 94.7% | 96.7% | 95.7% |
| **Step 2** | 72.1% | 100% | 4.3% | 27.6% |
| **Step 3** | 71.2% | 94.7% | 3.8% | 26.3% |

In Figure 4, we study if the behaviors of $\hat{\rho}(2)$ and $\hat{\rho}(3)$ are compatible with what is expected for simulations of MA(2) processes. Indeed, we plot the performance of **Step 3** versus coefficient $a_2$. It appears that when the coefficient $a_2$ used in the simulations is rather small, **Step 3** has very poor reliability, and it is nearly optimal as soon as $a_2 > 0.25$ when $n = 500$.

Finally, we illustrate the ability of our procedure to identify the underlying order $q = 2$ used to simulate the MA(2) processes by studying **Step 4**. In Figure 5, we give the

frequency of each last lag when $n = 100$, and in Figure A2 in Appendix C, we consider $n = 500$. It appears that, whatever the distribution of the underlying White Noise, the graphical diagnosis is relevant since in most of the simulations, the identified model is an MA(2), as expected, when $a_2 = \frac{1}{2}$. On the contrary, when $a_2 = \frac{1}{10}$, the reliability of the order-identification procedure seems disastrous. Indeed, the procedure rather suggests an MA(1) process. Table 3 shows the number of times where the procedure suggests either the correct MA(2) model or the MA(1) model, when $a_2 = \frac{1}{2}$ and $a_2 = \frac{1}{10}$.

**Table 3.** Number of simulations, among $N_S = 5000$ simulations of an MA(2) process, where the procedure suggests either $\hat{q} = 1$ or $\hat{q} = 2$. The percentage of valid models, validated with the Ljung–Box test, is added in brackets. When no MA(1) model has been suggested by the procedure, a "-" symbol has been placed between the brackets. The associated WN is Gaussian and coefficient $a_2$ is either $a_2 = \frac{1}{2}$ (left column) or $a_2 = \frac{1}{10}$ (right column).

| | $a_2 = \frac{1}{2}$ | | $a_2 = \frac{1}{10}$ | |
|---|---|---|---|---|
| **n** | **100** | **500** | **100** | **500** |
| **nb of $\hat{q} = 1$ (% valid.)** | 836 (28.0%) | 0 (–) | 3243 (92.4%) | 339 (95.2%) |
| **nb of $\hat{q} = 2$ (% valid.)** | 2967 (96.4%) | 709 (96.9%) | 176 (91.5%) | 150 (96.7%) |

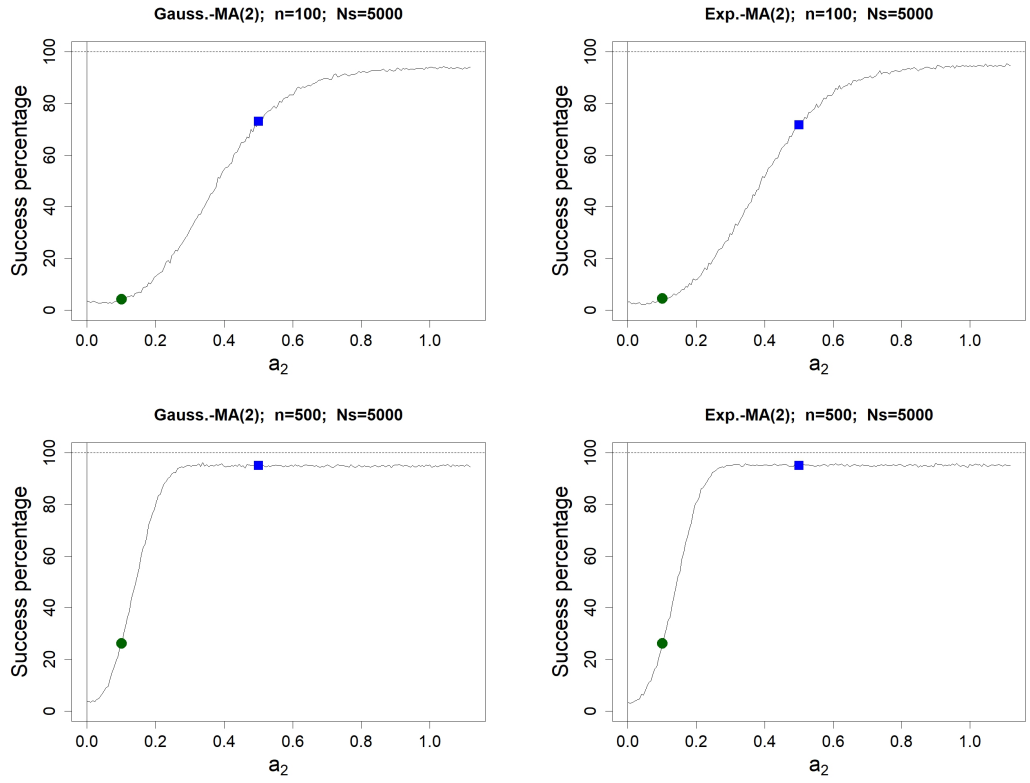


**Figure 4.** Percentage of simulations with $\rho(2)$ out of interval $\hat{\mathcal{J}}_2$ and with $\rho(3)$ inside interval $\hat{\mathcal{J}}_3$ as described in Proposition 3, among $N_S = 5000$ simulations of MA(2) processes. Coefficient $a_2$ varies from 0 to $\alpha$ as defined in Appendix A.2. The left column concerns simulated MA(2) processes with an underlying Gaussian WN, whereas the right one deals with Exponential WN. In the top figures, we simulate MA(2) processes with length $n = 100$, and with $n = 500$ in the bottom. The green circle points for $a_2 = \frac{1}{10}$, whereas the blue square points for $a_2 = \frac{1}{2}$.

But perhaps the procedure has the right behavior in the end, meaning that an MA(1) model is better suited to most of these series with $a_2 = \frac{1}{10}$ than an MA(2) model. For each simulation, we estimate an MA(1) model and an MA(2) model. To assess the quality of these models, we calculate the information criteria (the Akaike information criterion AIC [52], its small-sample equivalent AICc [53,54], and Bayesian Information Criterion BIC [55]) and prediction criteria (RMSE and MAPE, where the training set consists of observations 1 to $n-1$ and the test set corresponds to the final observation $n$). Finally, we use a two-tailed Student's paired $t$-test of equality of means to test the equality of each criterion between model MA(1) and model MA(2). Tables A1 and A2 in Appendix D contain the $p$-values of these tests. When the $p$-value is less than the nominal risk $\alpha = 5\%$, we indicate which model is significantly better (with respect to the criterion under consideration); otherwise, we indicate a symbol of equality because neither model is better or worse than the other. Table A1 shows the results when we consider simulations for which our procedure suggests an MA(1) model, while Table A2 considers simulations for which our procedure suggests an MA(2) model. Table A2 shows that when an MA(2) model has indeed been suggested by our procedure, the MA(1) model is not preferred for any of the criteria. On the contrary, the MA(2) model is significantly preferable, for all information criteria, and for the RMSE prediction criterion when $a_2 = \frac{1}{2}$. For its part, Table A2 shows that when an MA(1) model has been suggested by our procedure, while we have simulated an MA(2) process, the MA(1) model shows significantly better quality for several criteria. Specifically, when $a_2 = \frac{1}{10}$, and the series is short ($n = 100$), all information criteria and the MAPE prediction criterion are significantly better for the suggested MA(1) model, and the RMSEs are equivalent. When the series is longer ($n = 500$), the MA(1) model remains preferable for the BIC and MAPE criteria, but the underlying MA(2) model is more compatible with the data in terms of the AIC and AICc criteria. When $a_2 = \frac{1}{2}$, the MA(1) model is rarely suggested on MA(2) simulations (see Table 3), unless the series is short. In these rare cases, the MA(1) model still has a significant advantage in terms of the MAPE prediction criteria. In conclusion, the MA(2) model estimated on an MA(2) simulation generally offers significantly better quality than an MA(1) model, unless $a_2$ is close to zero and the MA(1) model is suggested by our procedure on this simulation.

We recall that when constructing an ARMA($p, q$) model, the Ljung–Box test must be applied to the residuals of the estimated model [13,14,56], and calculated mainly for $h$ ranging from $p+q+1$ to $H = n/4$ as suggested by [16]. Note that other upper bounds for $h$ are suggested in other papers, such as $H = \min\left(20, \frac{n}{4}\right)$ [57], $H = ln(n)$ [3], and more explicit bounds obtained form simulation procedures [58]. Furthermore, tests at fixed lag $h$ are not independent and provide $p$-values $p_h$, which do not all lead to the same conclusion. Here, we automatically consider that the model is validated by the Ljung–Box tests if $p_{n/4} \geq 0.05$ and if 80% of the $p_h$ values are also $\geq$0.05. In Table 3, we obtain that when $a_2 = \frac{1}{10}$, almost 95% of the models suggested by the procedure are valid, whether it is an MA(1) or an MA(2) model. But when $a_2 = \frac{1}{2}$, the MA(1) model is either not suggested by the procedure (when $n$ is sufficiently large), or it is mainly not validated by Ljung–Box (less than 30%). This means that when the time series length is small, the procedure might err in suggesting an underestimated model. But in this case, the Ljung–Box validation test can detect that such a model is not suitable.

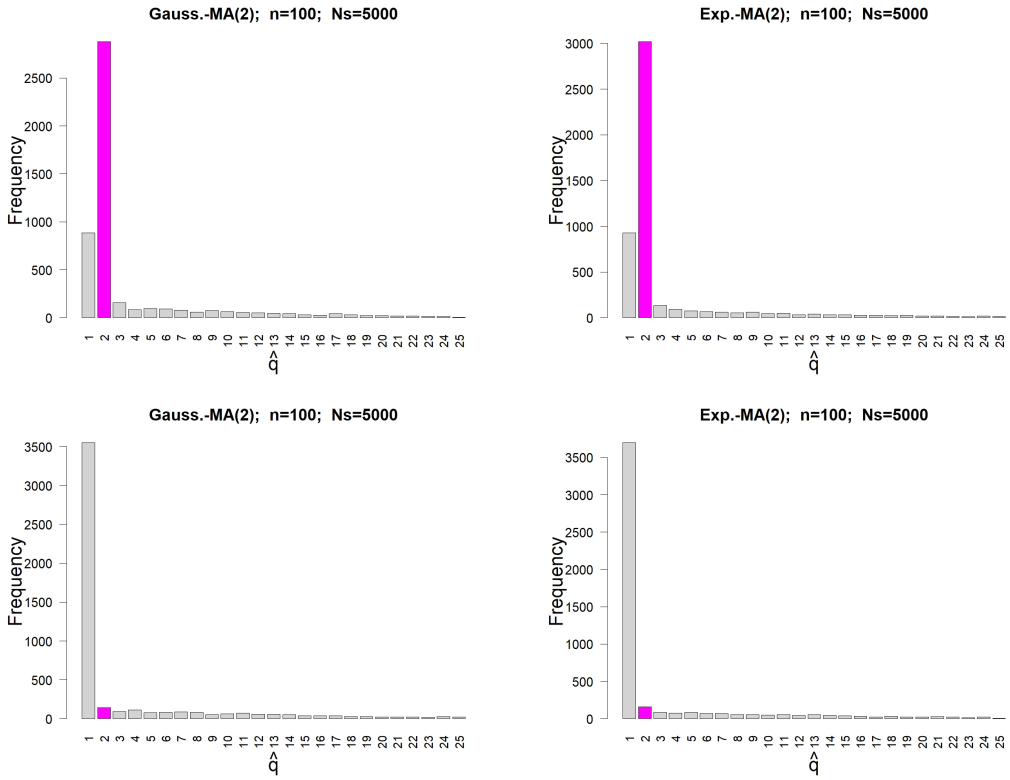


**Figure 5.** Frequency of $\hat{q}$, the last lag $h$ such that $\hat{\rho}(h)$ lies out of interval $\hat{\mathcal{J}}_h$, as described in Proposition 3, for $N_S = 5000$ simulations of MA(2) processes of length $n = 100$. The left column concerns simulated MA(2) processes with an underlying Gaussian WN, whereas the right one deals with Exponential WN. In the top figures, we simulate MA(2) processes with $a_2 = \frac{1}{2}$, whereas we use $a_2 = \frac{1}{10}$ in the bottom. The magenta-colored bar highlights the theoretical convenient order $q = 2$.

## 4. Numerical Results for Simulated ARMA($p, q$) Processes

We simulate the ARMA($p, q$) series to investigate the identification of orders $p$ and $q$ using EACF as presented in Section 2.3. We simulate $N_S = 5000$ ARMA($p$,2) processes ($p = 1, 2$) associated to a Gaussian WN, with length $n = 500$:

$$\text{ARMA(1,2)} \qquad Z_t - \frac{3}{4}Z_{t-1} = \mathcal{E}_t + \frac{1}{2}\mathcal{E}_{t-1} + \frac{1}{2}\mathcal{E}_{t-2}$$

$$\text{ARMA(2,2)} \qquad Z_t - \frac{1}{2}Z_{t-1} + \frac{3}{4}Z_{t-1} = \mathcal{E}_t + \frac{1}{2}\mathcal{E}_{t-1} + \frac{1}{2}\mathcal{E}_{t-2}\,.$$

For any simulated ARMA($p$,2) process, we compute the associated EACF $\tau(k, l)$ for $k = 0, 1, \cdots, 5$ and $l = 0, 1, \cdots 10$ and we run the following steps:

**Step 1.** We check if $\tau(p, 2) =$ "o" and we compute the percentage of true diagnosis.
**Step 2.** We check if $\tau(p, 1) =$ "x" and we compute the percentage of true diagnosis.
**Step 3.** We check if both $\tau(p, 1) =$ "x" and $\tau(p, 2) =$ "o" and we compute the percentage of true diagnosis (both properties must be satisfied).
**Step 4.** We identify the upper-left corner of the first largest triangle (with lowest order $p$) of "o"s in the EACF matrix, computed with $p$ ranging from 0 to 5 and $q$ from 0 to 10.

We recall that the classification as either "o" or "x" is based on the comparison of $\tau(p, 1)$ (respectively, $\tau(p, 2)$) values with their associated thresholds. The thresholds, introduced in Section 2.3, are derived under the assumption of EACF normality. By the recursive construction of the EACF matrix, if the underlying process follows an ARMA($p, q$) model, then $\tau(p, h)$ are Gaussian for $h \geq q - 1$. In this study, we simulate ARMA($p$,2) processes

with $p \in \{1,2\}$. Consequently, $\tau(p,1)$ and $\tau(p,2)$ are expected to follow a Gaussian distribution. Since these two terms of the EACF are pivotal in our methodology, we assess their normality using the Shapiro–Wilk test. Figure 6 demonstrates that for short time series ($n = 100$), normality is not satisfied. However, for sufficiently long series ($n = 500$), the normality assumption holds for $\tau(p,2)$ even under a large number of simulations, whereas $\tau(p,1)$ loses its Gaussian behavior more rapidly.

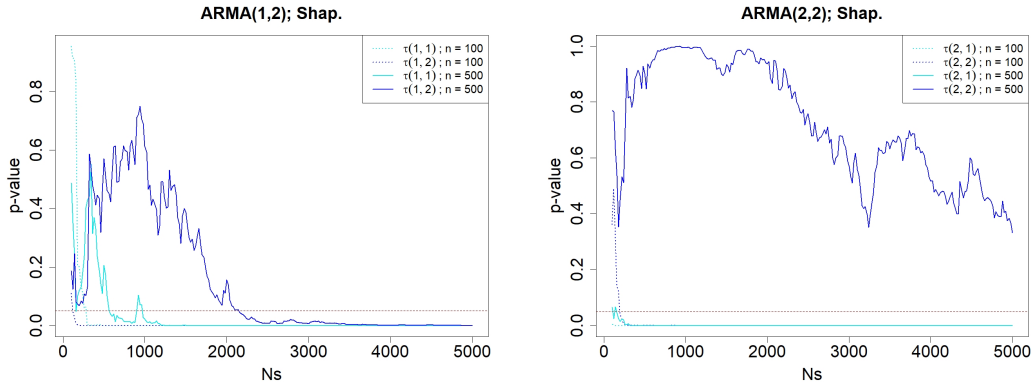


**Figure 6.** $p$-values when testing for the normality of $\tau(p,1)$ and $\tau(p,2)$, for $N_S$ simulated ARMA($p$,2) processes, where $N_s$ varies from 100 to 5000. The involved normality test is Shapiro's. The left graphic concerns ARMA(1,2) simulations, whereas the right one deals with ARMA(2,2). The length of the simulated ARMA$(p,q)$ processes are either $n = 100$ (dotted lines) or $n = 500$ (solid lines). The red-dotted horizontal line represents 5%.

Table 4 gives the percentage of true diagnosis for **Steps 1** to **3**, when simulating $N_S = 5000$ ARMA($p$,2) processes, with a Gaussian underlying WN. The high percentage of true diagnosis in **Step 1**, confirms that $\tau(p,2)$ mainly behaves conveniently, but its error rate seems far enough from the expected 5% type-I error. In **Step 2**, we obtain that the frequency of convenient $\tau(p,1)$ behavior improves when the time series length $n$ increases. As expected in **Step 3**, the performance of the paired properties is smaller than the singles.

**Table 4.** Percentage of true diagnosis among $N_S = 5000$ simulations of a ARMA($p$,2) process, when checking **Steps 1** to **3**. We have $p = 1$ in the left columns and $p = 2$ in the right one. The associated WN is Gaussian.

| | **ARMA(1,2)** | | **ARMA(2,2)** | |
|---|---|---|---|---|
| **n** | **100** | **500** | **100** | **500** |
| **Step 1** | 88.1% | 84.3% | 94.3% | 90.0% |
| **Step 2** | 73.7% | 100% | 56.1% | 100.0% |
| **Step 3** | 64.8% | 84.3% | 53.9% | 90.0% |

Finally, we evaluate whether the discrepancy between the theoretical and empirical properties of the pivotal EACF terms $\tau(p,2)$, and particularly $\tau(p,1)$, affects the accuracy of the procedure described in Section 2.3 for identifying the $p$ and $q$ orders. To this end, we simulate $N_s = 5000$ ARMA$(p,2)$ processes, first with a length of $n = 100$, and then with $n = 500$. For each simulation, we compute the EACF matrix and identify the vertex corresponding to the largest triangle of "o". In Figure 7, we display the percentage of occurrences for all possible orders $(\phi, q)$, where $\phi$ ranges from 0 to 5 and $q$ ranges from 0 to 10. Despite the suboptimal properties of $\tau(p,1)$, the procedure outlined in Section 2.3 performs remarkably well in identifying the $p$ and $q$ orders of an ARMA$(p,q)$ model. Specifically, it predominantly identifies the true orders $p$ and $q$ of the underlying model used for the simulations.

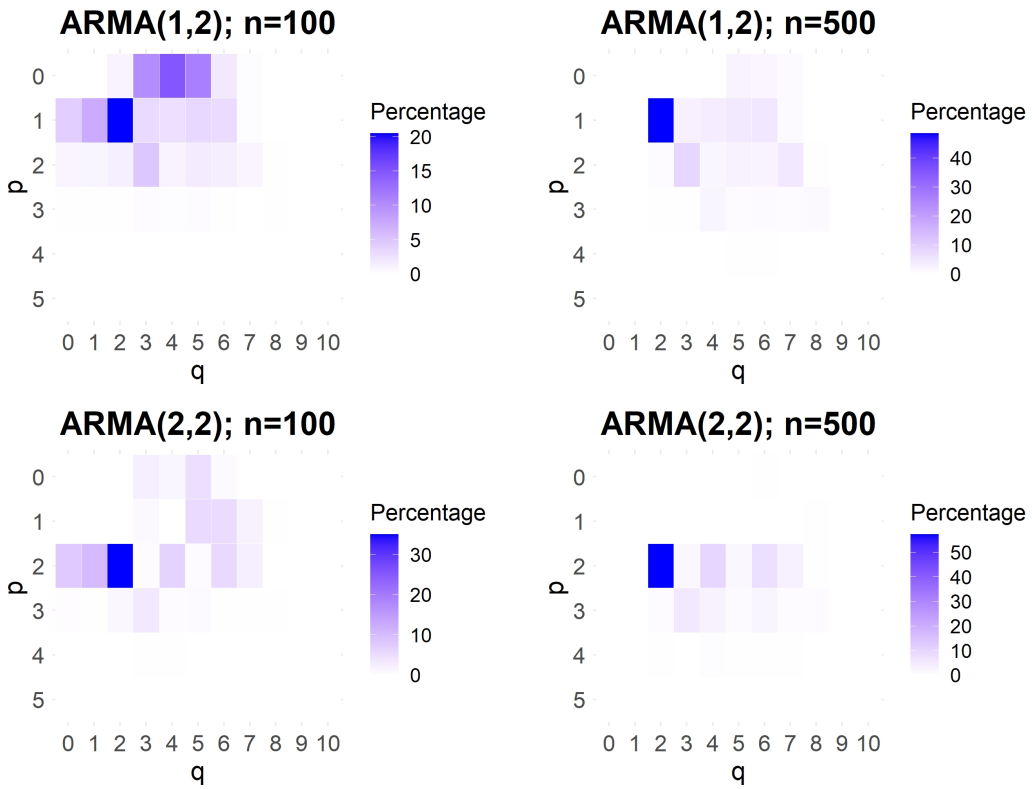


**Figure 7.** Percentage of occurrences of orders $(\phi, q)$, identified as the vertex corresponding to the largest triangle of "o" among $N_s = 5000$ simulations. The darker the box, the higher the occurrence. Each row corresponds to a fixed value of $\phi$, ranging from 0 to 5, while each column corresponds to a fixed value of $q$, ranging from 0 to 10. The top panels represent simulations of ARMA(1,2) models, while the bottom panels correspond to ARMA(2,2) models. The lengths of the simulated ARMA$(p, 2)$ processes are either $n = 100$ (**left panels**) or $n = 500$ (**right panels**).

## 5. Illustration on Meteorological Data

### *5.1. Wind Speed Data*

We consider the wind speed in miles per hour at 7 h and 10 h at *La Guardia Airport*, in New York, every day from 1 May 1973 to 30 September 1973. This time series of length $n = 153$ is available from the `airquality` dataset in R. See Figure 8 (**left**) for its evolution. As illustrated in Figure 8 (**right**), our procedure suggests $\hat{q} = 3$, even if $\hat{\rho}(1)$ lies out the interval $\hat{\mathcal{J}}_1$. Note that, to illustrate the accuracy of our procedure, we do not just construct the MA(3) model but also WN, MA(1), and MA(2) models. In order to verify if WN and the MA$(q)$ candidate models are valid, we check the behavior of their residuals, by plotting their ACF (Figure 9) and also their standardized SACF (Figure 10) as suggested in [44].

In Figure 9, it appears that the ACFs of the residuals are compatible with WN for all the MA$(q)$ models, for $q = 0, 1, 2, 3$. When $q = 0$, plotting the ACFs of the residuals of an MA(0) model reduces to plotting the ACF of the series itself. Observing at most 3 out of 38 ACFs outside the thresholds adapted to WN is not incompatible with the hypothesis that the series is WN, insofar as it can be observed with a probability of 0.296. In Figure 10, we observe very different behaviors from the standardized SACF. If the SACF of the residuals in the MA(3) model is compatible with a good model fit, this is much less clear for MA(1) and MA(2) models. And the SACFs of the simplest model MA(0) show that the WN model cannot fit the data. We recall that when the standardized SACFs calculated on the residuals of an adjusted model fall outside the associated thresholds, the proposed model might be under-calibrated [44]. This is confirmed with Ljung–Box tests, applied from lags $H = q + 1$ to $\frac{n}{4} \simeq 38$. Indeed, Figure 11 shows that MA(1), MA(2) and MA(3) models are valid,

whereas the WN model is totally inaccurate for the wind speed data. Indeed, it still retains some dependency.

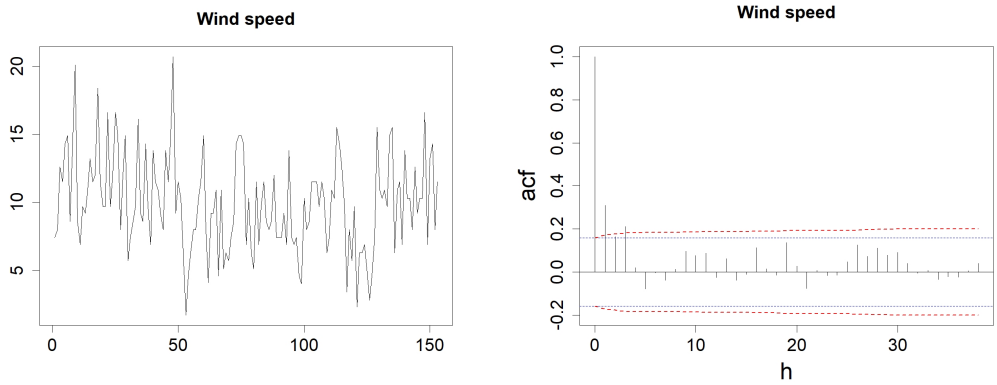


**Figure 8.** Wind speed evolution (**left**) and ACF with $h$ varying from 1 to $\frac{n}{4} \simeq 38$ (**right**). In the left figure, the blue-dotted horizontal lines represent the classical thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$, whereas the red dashed-line represents the corrected thresholds of $\hat{\mathcal{J}}_h$, dedicated to identifying an MA($q$) model, as detailed in Proposition 3.

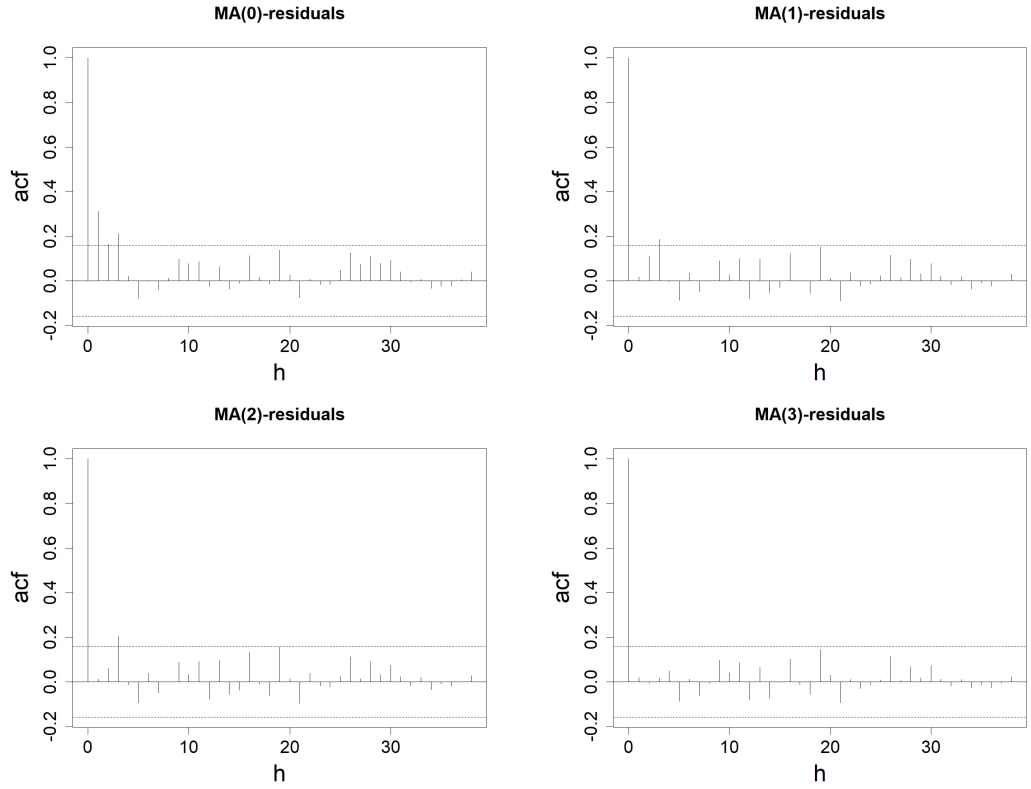


**Figure 9.** Sample ACF of the residuals from the MA($q$) model constructed on wind speed data, with $h$ varying from 1 to $\frac{n}{4}$. Order $q$ is either equal to 0 (**top left**), 1 (**top right**), 2 (**bottom left**), or 3 (**bottom right**). The blue-dotted horizontal lines represent the thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$.

Moreover, note that if we try to build an MA(4) model, it appears to be a valid model as shown by the ACF and SACF computed on its residuals and by Ljung–Box test (see Figure A3 in Appendix E). Indeed, in [44], from simulations, we observe that over-calibrated models produce graphical diagnoses similar to those obtained with the residuals of the underlying model used to produce the series. But the simplification procedure (as described in [16] Equation (8.8.2)), applied on the MA(4) model, leads to an MA(3) model, the model suggested by our procedure in Proposition 3.

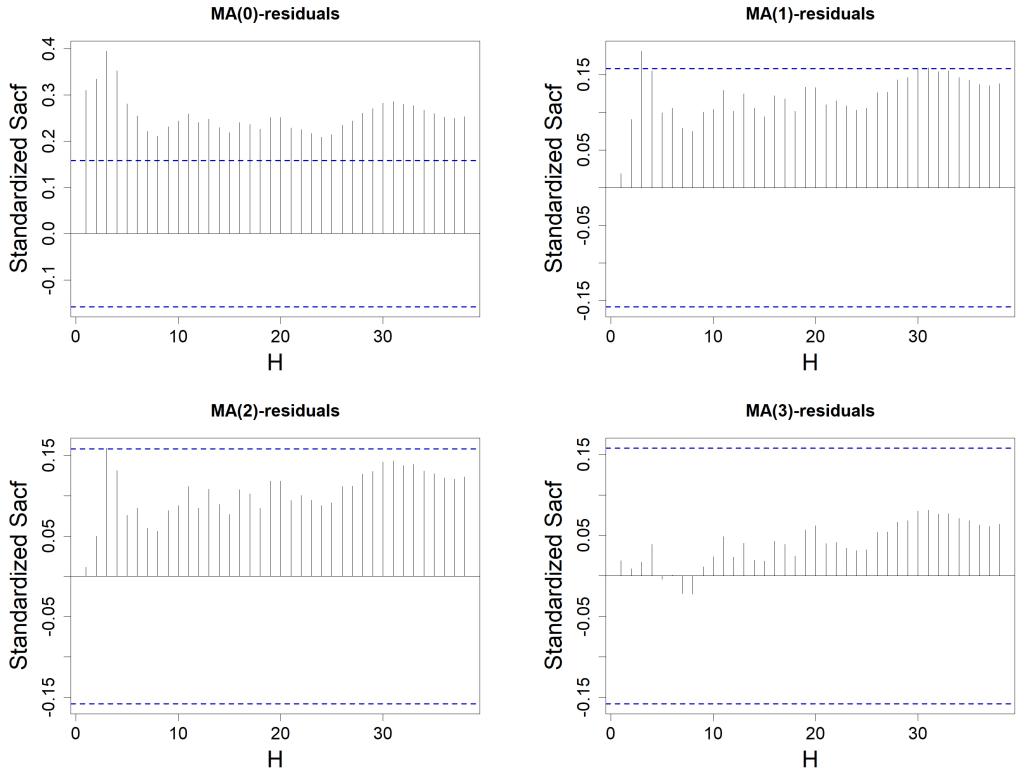


**Figure 10.** Standardized SACF of the residuals from the MA($q$) model constructed on wind speed data, with $H$ varying from 1 to $\frac{n}{4}$. Order $q$ is either equal to 0 (**top left**), 1 (**top right**), 2 (**bottom left**), or 3 (**bottom right**). The blue-dotted horizontal lines represent the thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$.

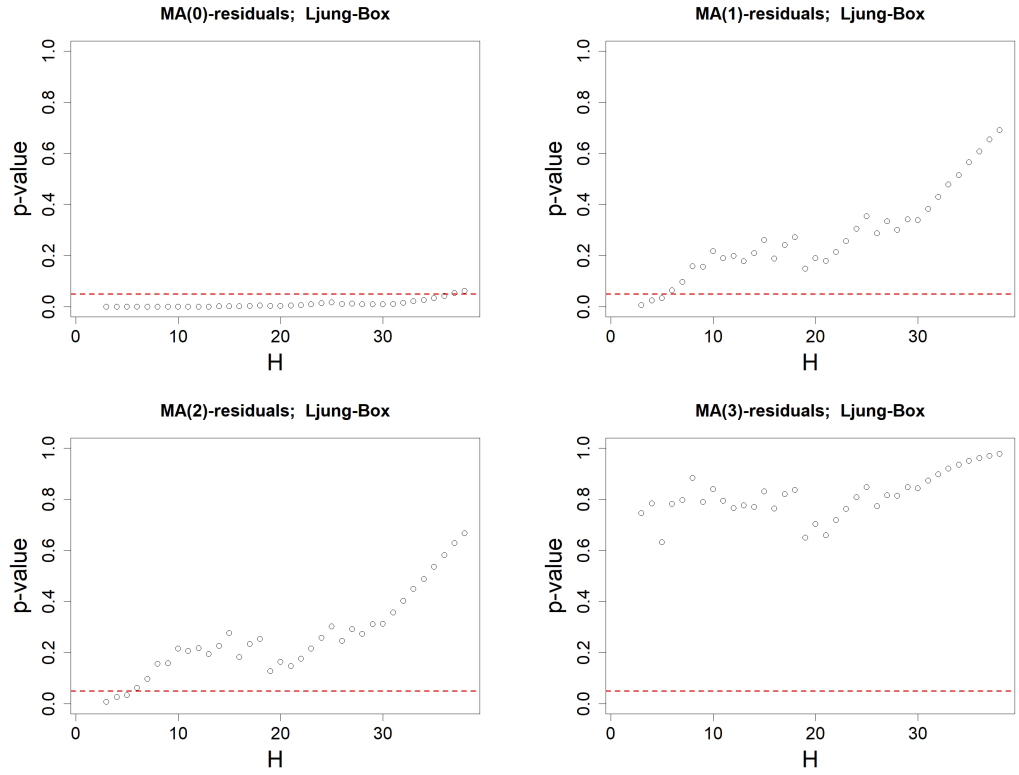


**Figure 11.** $p$-values when using Ljung–Box's test on the residuals associated to an MA($q$) model. Order $q$ is either equal to 0 (**top left**), 1 (**top right**), 2 (**bottom left**) or 3 (**bottom right**). Ljung–Box's test is computed successively for all lags $H = q + 1, \cdots, 38$, with degree of freedom $H - q$. The red-dotted horizontal line represents 5%.

Finally, we compare the reliability of the valid models, in terms of information criteria such as AIC and BIC [52,55] and in terms of prediction criteria (RMSE and MAPE). To compute the predictive criteria, we employ a rolling-window cross-validation approach. Given a time series of length $n$, we use a fixed training window comprising the first 90% of the available observations (from observation 1 to 138). The models are trained on this window and then used to forecast the next observation ($h = 1$). The window is then shifted forward by one time step, discarding the oldest observation and incorporating the next available data point. This process is repeated until all possible predictions are generated. At each iteration, the forecasted value is compared with the actual observation, allowing for the computation of error metrics such as the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE). In Table 5, it appears that the MA(3) model suggested by our procedure is the best model among the MA($q$) models with $1 \leq q \leq 4$, both on information and prediction criteria.

Let us also fit an ARMA($p, q$) model on these data, using the EACF procedure. Here is the EACF matrix:

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10
0 x x x o o o o o o o o
1 x o x o o o o o o o o
2 x x x o x o o o o o o
3 x x o o o o o o o o o
4 x x x o o o o o o o o
5 x o o x o o o o o o o
```

**Table 5.** Comparison of models quality between the suggested MA(3) model, and other models MA(0), i.e., WN, MA(1), MA(2), and MA(4). For each criterion, we have bolded the value corresponding to the best model, i.e., the MA(3) for almost all the criteria.

| | **MA(1)** | **MA(2)** | **MA(3)** | **MA(4)** |
|---|---|---|---|---|
| **RMSE** | 3.43 | 3.42 | **3.19** | 3.26 |
| **MAPE** | 27.66 | 27.70 | **24.80** | 25.98 |
| **AIC** | 810.93 | 812.47 | **807.18** | 807.76 |
| **AICc** | 811.09 | 812.74 | **807.59** | 808.33 |
| **BIC** | **820.02** | 824.59 | 822.33 | 825.94 |

As illustrated in Table 1, the candidate $p$ and $q$ orders correspond to the upper left vertex of the largest observed zero triangle. Unfortunately, empirical EACFs rarely provide a structure as clear as that expected theoretically from an ARMA($p, q$) process. It therefore makes sense to consider several candidate models. Here, the EACF procedure suggests either an ARMA(0,3), an ARMA(2,5), an ARMA(3,2), or an ARMA(4,4) model. Note that no model of the form ARMA(1, $q$) can be suggested from the EACF matrix, notably because $\tau(2, 4) =$"x". It is interesting to see that the EACF-based procedure recommends the MA(3) model as the MA($q$) model, which corresponds to the best MA($q$) model as justified above in Table 5.

In Table 6, we compare the MA(3) model with the ARMA(2,5), ARMA(3,2), ARMA(4,4) models suggested by the EACF matrix and with an ARIMA(3,1,1) model, suggested by an automated procedure (`auto.arima` R-function from package `forecast` [24,59]). Note that for any candidate model, we check its validity as seen in Figures A4–A6 in Appendix E. This step is crucial, especially when using an automated procedure, based in particular on optimizing the BIC criterion since it often leads to invalid models [60]. Regarding the model selection criteria, in Table 6, we obtain that the MA(3) model minimizes the information criteria

AIC, AICc, and BIC, favoring it as the most parsimonious option that balances goodness of fit and model complexity. But for predictive accuracy, the ARMA(4,4) model achieves the lowest RMSE (2.97), indicating the best overall fit to the data in terms of minimizing squared errors. It also has the smallest MAPE error (24.43), which suggests a good relative accuracy. Hence, if the predictive accuracy is the primary concern, ARMA(4,4) may be chosen, and for a trade-off between fit and complexity, the MA(3) model is recommended. Both MA(3) and ARMA(4,4) models are the models suggested by our model selection procedures, based respectively on ACF and EACF and described in Propositions 3 and 4.

**Table 6.** Comparison of models quality between the models ARMA(2,5), ARMA(3,2), ARMA(4,4), suggested by our procedure described in Proposition 4, based on the EACF matrix; an ARIMA(3,1,1) model, suggested by an automated procedure; and the MA(3) model, suggested by the ACF procedure, defined in Proposition 3. For each criterion, we have bolded the value corresponding to the best model.

| | EACF | | | AUTOMATED | ACF |
|---|---|---|---|---|---|
| | **ARMA(2,5)** | **ARMA(3,2)** | **ARMA(4,4)** | **ARIMA(3,1,1)** | **MA(3)** |
| **RMSE** | 3.18 | 3.29 | **2.97** | 3.29 | 3.19 |
| **MAPE** | 25.68 | 26.53 | **24.43** | 26.04 | 24.80 |
| **AIC** | 807.76 | 809.90 | 808.96 | 807.82 | **807.18** |
| **AICc** | 809.02 | 810.68 | 810.51 | 808.23 | **807.59** |
| **BIC** | 835.03 | 831.12 | 839.27 | 822.94 | **822.33** |

### *5.2. Temperature Anomalies*

Since the industrial revolution, the burning of fossil fuels, deforestation, and other anthropogenic activities have significantly increased atmospheric concentrations of carbon dioxide ($CO_2$), methane ($CH_4$), and other greenhouse gases. These emissions enhance the natural greenhouse effect, leading to rising global temperatures, shifts in climate patterns, and increasing frequency of extreme weather events.

To assess the extent and impact of climate change, long-term temperature records are essential. The study of temperature anomalies—deviations from a defined baseline—provides a crucial metric for detecting trends in global and regional climate variations. We analyze historical sea surface temperature (SST) from multiple sources, including ship-based and buoy measurements (ICOADS), autonomous ocean profiling floats (Argo), and satellite-derived sea ice concentration datasets (HadISST2 and NCEP) [61]. In Figure 12 (left), we plot the long-term evolution of SST anomalies (with reference period 1880–1970). The temperature data for the period 1850–1880 are characterized by higher variability and lower reliability due to limited spatial coverage and methodological inconsistencies. During this early period, the number of observation stations is significantly lower, leading to potential biases and regional gaps in the dataset. Furthermore, measurement techniques and data homogenization methods are less developed, contributing to increased uncertainty. For these reasons, this period is excluded from the model to ensure robustness and reliability in the analysis.

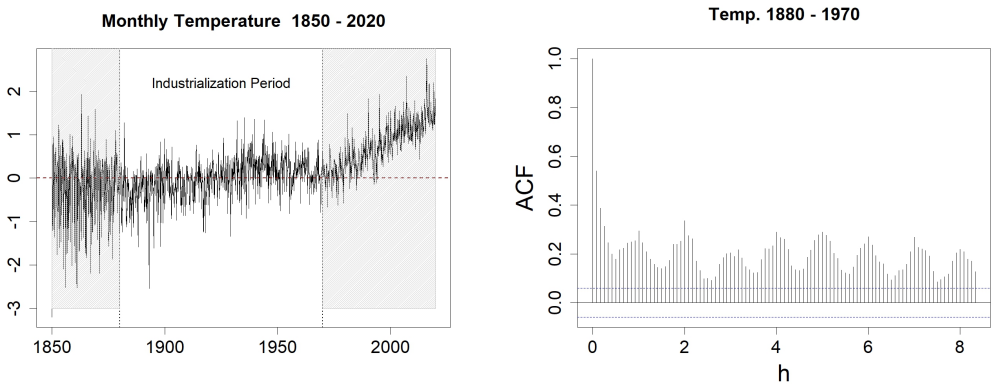


**Figure 12.** Sea Surface Temperature anomalies (**left**) and sample ACF with $h$ varying from 1 to 100 (**right**). In the left figure, the blue-dotted horizontal lines represent the classical thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$.

The ACF of the data given in Figure 12 (right), reveals a clear seasonal structure, indicating periodic dependencies. To confirm the presence of seasonality, we apply the Canova–Hansen (CH) test [62], which suggests the need for seasonal differencing at lag 12. To identify a model for the seasonally differenced series, in Figure 13, we plot its ACF (left), dedicated to identifying an MA($q$) model, and its partial ACF (right), dedicated to identifying rather an AR($p$) model. On the one hand, from the sample ACF with our corrected thresholds, we can suggest an MA(36) model. Note that the non-corrected thresholds would lead to a much greater $q$-order, reflecting spurious correlations. On the other hand, the sample partial ACF cannot identify clearly an AR($p$) model. To stay with a reasonable number of parameters, we consider an AR(36) model. However, we observe that PACFs decrease exponentially as they oscillate, which is characteristic of an MA($q$) process. Note that these oscillations point to lag $h = 13$, and especially to lag multiples of 12, potentially suggesting a SARMA model. Combining both sample ACF and partial ACF, we can suggest a SARIMA(13,0,0)(0,1,3) [12] model for SST anomalies. Finally, we compute EACF on the seasonally differenced series to identify an additional model:

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0 x x x x o o x o o x x  x  x  x  x  o  o  o  o  o  o
1 x o o o o o o o o o o  x  x  o  o  o  o  o  o  o  o
2 x x o o o o o o o o o  x  o  o  o  o  o  o  o  o  o
3 x x x o o o o o o o o  x  x  o  o  o  o  o  o  o  o
4 x x x o o o o o o o o  x  x  x  o  o  o  o  o  o  o
5 x x x o o o o o o o o  x  x  x  x  o  o  o  o  o  o
6 x x x o o o o o o o o  x  x  x  x  x  o  o  o  o  o
7 x x x x x o o o o o o  x  x  x  o  x  o  o  o  o  o
8 x x x o x o o o o o o  x  x  x  o  x  o  o  o  o  o

AR/MA
  21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
0 o  o  x  o  o  o  o  x  o  o  x  x  x  x  x  x  o  o  o  o
1 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
2 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
3 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
4 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
5 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
6 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  x  o  o  o  o
```

```
7 o  o  x  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
8 o  o  o  o  o  o  o  o  o  o  o  o  o  o  o  x  o  o  o  o  o
```

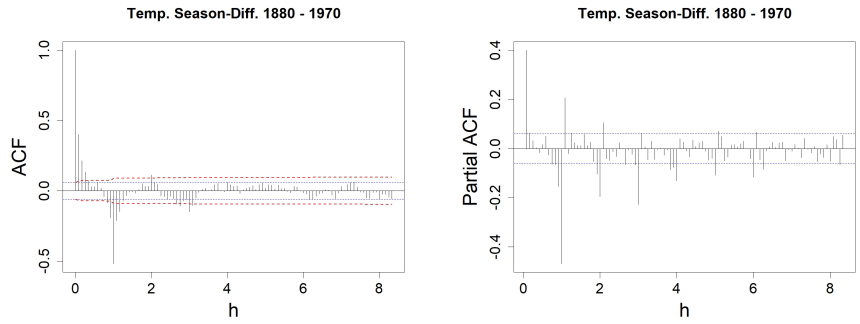


**Figure 13.** Sample ACF (**left**) and partial ACF (**right**) of the seasonally differenced series constructed on SST anomalies, with $h$ varying from 1 to 100. The blue-dotted horizontal lines represent the thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$. The red dashed-line represents the corrected thresholds of $\hat{\mathcal{J}}_h$, dedicated to identifying an MA($q$) model, as detailed in Proposition 3.

The EACF is rather difficult to interpret, as it does not clearly distinguish a large triangle composed of "o". But note that all values of $\tau(p,11), \tau(p,23)$ and $\tau(p,35)$ are significantly non-zero. Moreover, we checked that all EACF $\tau(p,q)$ are null when $1 \leq p \leq 8$ and $41 \leq q \leq 100$. In other words, the ACF of the extended residuals $W_t^{(p,q)}(q)$ are significantly non-zero when $q = 12$, 24, and 36, again suggesting a SARMA model. If we omit the EACF values at these particular orders, we detect a triangle of "o" which suggests an ARMA(1,1) model. All together, this provides SARIMA(1,0,1)(0,1,3)[12] as a candidate model.

We check the validity of all the candidate SARIMA$(p,0,q)$(0,1,$Q$)[12] models previously identified by computing Ljung–Box tests $p$-values from lags $H = max(p, 12Q) + 1$ to 100 (Figure 14). MA(36), SARIMA(13,0,)(0,1,3)[12], and SARIMA(1,0,1)(0,1,3)[12], the models produced by our procedures, are all valid models but AR(36) is not.

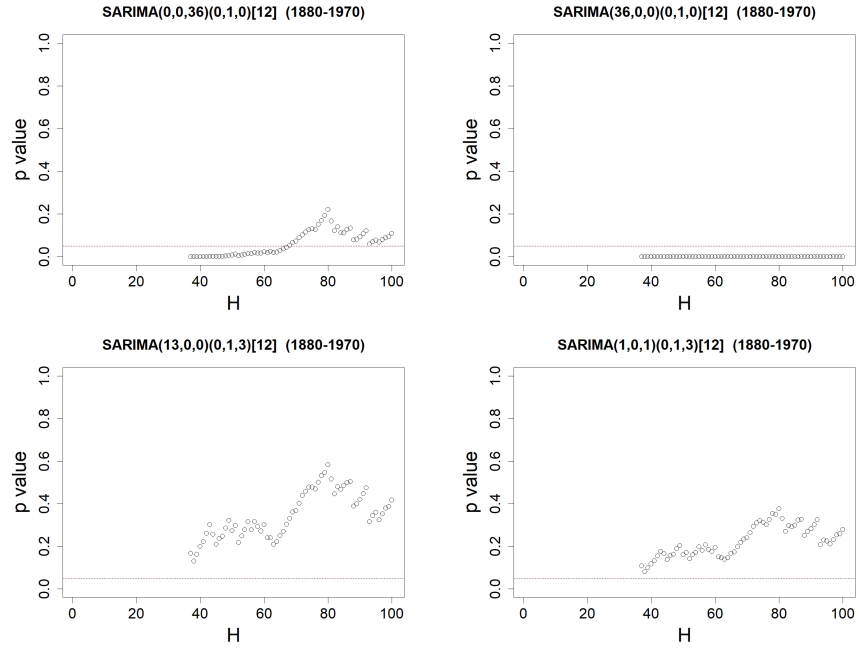


**Figure 14.** $p$-values when using Ljung–Box's test on the residuals from SARIMA$(p,0,q)$(0,1,$Q$)[12] models identified on SST anomalies data, with $H$ varying from $H = max(p, 12Q) + 1$ to 100. Orders $(p,q)(Q)$ are either equal to $(0,36)(0)$ (**top left**), $(36,0)(0)$ (**top right**), $(13,0)(3)$ (**bottom left**) or $(1,1)(3)$ (**bottom right**). The red-dotted horizontal line represents 5%.

Finally, in Table 7, we compare the quality of all the models either with information criteria and prediction criteria. All the models suggested by our procedures provide accurate models, the best ones being the SARIMA(13,0,0)(0,1,3)[12], derived from ACF and PACF and the SARIMA(1,0,1)(0,1,3)[12], derived from EACF. In Figure 15 (top), we compute the forecasts with the best models. Criteria and forecasts provided by the SARIMA(0,0,0)(0,1,0)[12] and SARIMA(36,0,0)(0,1,0)[12] models on the seasonally differenced series are also given as a comparison (Table 7 and Figure 15-bottom). It is interesting to see that the models generated by our procedures reveal global warming [63,64], while models SARIMA(0,0,0)(0,1,0)[12] and SARIMA(36,0,0)(0,1,0)[12] completely miss the point.

**Table 7.** SARIMA$(p,0,q)(0,1,Q)[12]$ models identified on SST anomalies data by our procedures, with other models. SARIMA$(p,0,q)(0,1,Q)[12]$ models are denoted by $(p,q)(Q)$. For each criterion, we have bolded the value corresponding to the best models.

| | ACF **(0,36)(0)** | ACF & PACF **(13,0)(3)** | EACF **(1,1)(3)** | *Others* **(0,0)(0)** | **(36,0)(0)** |
|---|---|---|---|---|---|
| **RMSE** | **0.31** | **0.31** | **0.31** | 0.38 | 0.46 |
| **MAPE** | 331.0 | **309.1** | **312.4** | 360.8 | 363.7 |
| **AIC** | 777.7 | **757.9** | **753.3** | 925.9 | 1573.6 |
| **AICc** | 777.4 | **758.5** | **753.4** | 928.6 | 1573.6 |
| **BIC** | 958.7 | **842.5** | **783.2** | 1109.9 | 1576.2 |

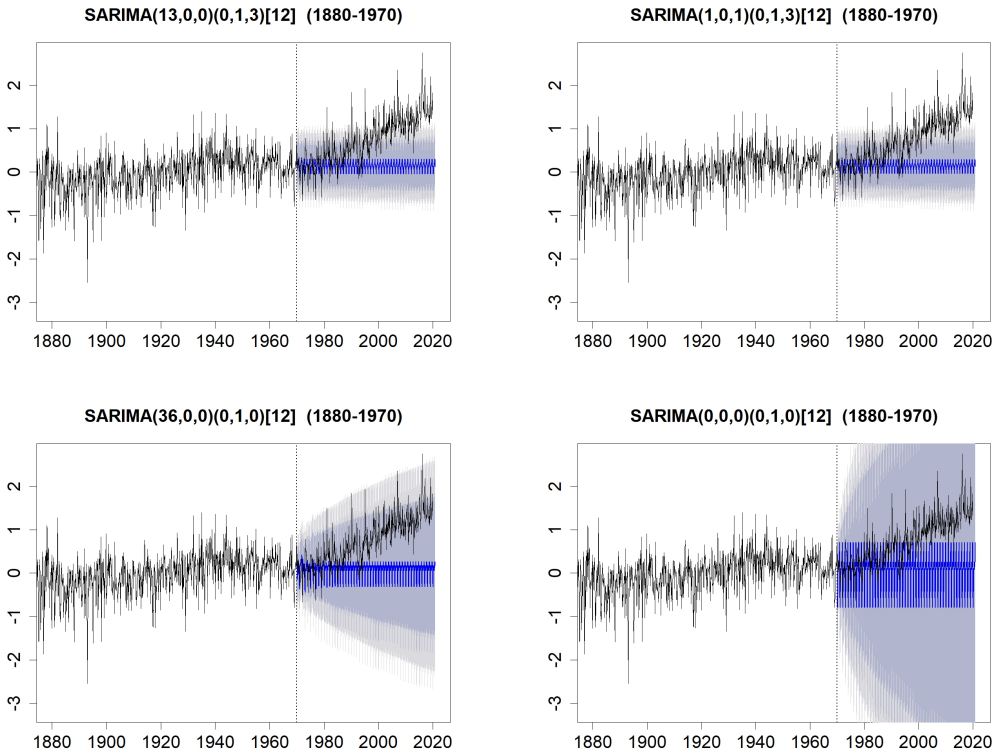


**Figure 15.** Forecasts on the period 1970–2020, with prediction intervals at levels 80% (grey) and 95% (lightgrey), compared with the observed data. Forecasts are computed with the SARIMA(13,0,0)(0,1,3)[12] model (**top-left**), with the SARIMA(1,0,1)(0,1,3)[12] model (**top-right**), with the SARIMA(36,0,0)(0,1,0)[12] model (**bottom-left**) and with the SARIMA(0,0,0)(0,1,0)[12] model (**bottom-right**).

## 6. Discussion

The Autocorrelation Function (ACF) provides a powerful framework for identifying the order $q$ of MA($q$) models. Its theoretical foundation relies on the convergence properties of the sample ACF to its asymptotic Gaussian distribution as derived from Theorem 1.

In this paper, we investigate the normality of sample ACFs, calculated from the realizations of an MA($q$) process. We observe that $\hat{\Xi}(h)$ generally exhibits Gaussian behavior as anticipated by Theorem 1, even for lags $h$ significantly larger than $\sqrt{n}$. However, slight deviations from this distribution are more pronounced for Exponential WN processes, suggesting a potential influence of their asymptotic properties. Additionally, we note a mismatch between $\hat{\Xi}(h)$ and the distribution $\mathcal{N}\left(0, \frac{v_{h,h}}{n}\right)$, which may stem from misspecifications in the expectation and variance. This observation implies that while $\hat{\Xi}(h)$ may converge to a Gaussian distribution, the convergence could involve a mean $\mu \neq 0$ or variance $\sigma^2 \neq \frac{v_{h,h}}{n}$. Moreover, we demonstrate that even when the underlying WN is Gaussian, the sum of sample ACFs, $S_{ACF}^{sample}(H)$, deviates from normality for most lags $H$, except possibly for the smallest lags. This finding indicates that the vector $\left(\hat{\Xi}(1), \cdots, \hat{\Xi}(H)\right)$ is unlikely to follow a Gaussian distribution, highlighting the need for caution when applying Theorem 1. Fortunately, the procedure for $q$-order identification, as outlined in Proposition 3, relies solely on the normality of $\hat{\Xi}(h)$ at fixed lags $h$ and does not depend on the normality of the vector $\left(\hat{\Xi}(1), \cdots, \hat{\Xi}(H)\right)$. However, this procedure assumes that $\hat{\Xi}(h)$ behaves as $\mathcal{N}\left(0, \frac{v_{h,h}}{n}\right)$ for any $h > q$. As a result, the accuracy of the procedure warrants careful consideration.

We simulate time series generated by MA(2) process with several values of parameter $a_2$, and we use the procedure detailed in Proposition 3 to identify an order $\hat{q}$ for a candidate MA($\hat{q}$) model. The procedure does not have perfect efficiency, especially when the length of the time series is not large enough, and may therefore propose a $\hat{q}$-order that is not the underlying q-order. On the one hand, the case where the procedure suggests a more general MA($\hat{q}$) model than the underlying MA($q$) model, i.e., with $\hat{q} > q$, may generally present no major problem. Indeed, all internal data correlations are taken into account. In addition, in the modeling approach, we next try to simplify the model, using likelihood-based criteria, and may approach the underlying order $q$. On the other hand, when the order $\hat{q}$ of the model is underestimated, i.e., when $\hat{q} < q$, it may not take sufficient account of the internal correlations of the data, leading to a poorly performing model. We have shown in [44] that the under-specification of a model generally leads Ljung–Box not to validate the model, when applied to the residuals, and so such models should not be considered.

Moreover, we obtain that the value of the parameter $a_q$ plays a major role in the choice of the order $\hat{q}$. Indeed, if the underlying parameter $a_q$ is far from 0, the procedure does not struggle to identify the order $q$. But if the parameter $a_q$ is close to zero, the procedure will tend to propose the order $q-1$. We also observe that Ljung–Box's test is more likely to validate the MA($q-1$) model if the $a_q$ parameter is close to 0. In this case, an MA($q-1$) model may correspond better to the data obtained, at least in terms of some information or prediction criteria.

Furthermore, we also study the identification of the orders $p$ and $q$ of an ARMA($p, q$) process, using the Extended Autocorrelation Function (EACF). The procedure for identifying $p$ and $q$ orders, detailed in Section 2.3, relies on the Gaussian behavior of $\tau(p, h)$ for $h > q$. Focusing on ARMA($p$,2) simulations, we explore the behavior of pivotal EACF terms $\tau(p, 1)$ and $\tau(p, 2)$ under various conditions. Through extensive simulations of ARMA($p, 2$) processes, we observe that these terms exhibit approximately Gaussian behavior as predicted by Theorem 1. However, deviations from normality are more pronounced in shorter time series ($n = 100$), particularly for $\tau(p, 1)$. For longer time series ($n = 500$), the normality

of $\tau(p,2)$ is generally maintained, whereas $\tau(p,1)$ displays a loss of Gaussian properties more rapidly.

Additionally, the application of our methodology to real-world data such as wind speed data and sea surface temperature anomalies, demonstrates that our procedures suggest valuable candidate models, which have to be explored.

In conclusion, the procedure presented in Proposition 3 (respectively, Section 2.3) is a valuable tool for identifying the $q$-order of the MA($q$) process (resp., the $q$-order of the MA($q$) process). It is particularly effective for longer time series. But any ARMA($\hat{p},\hat{q}$) suggested model should be complemented with model validation techniques, such as the Ljung–Box test. Additionally, the final model selection process should incorporate comparisons with other candidate models, using criteria like RMSE, MAPE, AIC, and BIC to identify the best model for forecasting purposes.

## 7. Conclusions

This study provides new insights into the limitations of standard ACF-based methodologies for time series model identification. By rigorously evaluating the normality of ACF estimators, we demonstrate that widely used statistical assumptions do not always hold. Indeed, while ACF estimates at fixed lags generally follow a Gaussian distribution, the joint behavior of ACF estimates does not conform to a multivariate normal distribution. This is clearly highlighted when computing the Sum of Sample Autocorrelation Functions (SACFs). This challenges the validity of standard statistical tests or model-identification methods relying on these assumptions. Then, we tested whether traditional ACF-based model identification techniques remain reliable when underlying assumptions are not met. In particular, we evaluated the performance of existing methods for identifying MA($q$) and ARMA($p,q$) orders, including threshold-based approaches for MA($q$) models and the Extended Autocorrelation Function (EACF) for ARMA($p,q$) models. Our findings suggest that these methods remain useful for sufficiently large sample sizes but may suffer from systematic biases when the moving average parameter approaches zero. In such cases, our study reinforces the need to complement model selection with residual diagnostics, such as the Ljung–Box test or their SACF examination. Thus, we refined existing order selection procedures and proposed improvements to conventional statistical approaches. This study not only contributes to the theoretical understanding of ACF properties but also provides practical insights for time series practitioners.

**Supplementary Materials:** The following supporting information can be downloaded at: www.i2m.univ-amu.fr/perso/manuela.royer-carenzi/AnnexesR.SacfMA/SacfMA.html (accessed on 5 January 2025).

**Author Contributions:** All the authors were involved in the conceptualization, in developing the method and in writing the original version; they developed the software and wrote the revised version. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the functions have been implemented in the **R** language. The data, codes, and supplementary materials are available on the project website: www.i2m.univ-amu.fr/perso/manuela.royer-carenzi/AnnexesR.SacfMA/SacfMA.html (accessed on 19 February 2025).

**Acknowledgments:** The authors thank the reviewers for their careful reading and their interesting suggestions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Details About MA(2) Processes

*Appendix A.1. Terms $v_{i,j}$*

Given Proposition 2, the covariance matrix associated with the H-vector of the ACF estimators of an MA(2) process has the following form:

$$\begin{pmatrix} v_{1,1} & v_{1,2} & \nu_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ v_{2,1} & v_{2,2} & \nu_1 & \nu_2 & 0 & \cdots & 0 & 0 & 0 \\ \nu_2 & \nu_1 & \nu_0 & \nu_1 & \nu_2 & \cdots & 0 & 0 & 0 \\ 0 & \nu_2 & \nu_1 & \nu_0 & \nu_1 & \cdots & 0 & 0 & 0 \\ \vdots & & & & \vdots & \vdots & & & \\ \vdots & & & & \vdots & \vdots & & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & \nu_0 & \nu_1 & \nu_2 \\ 0 & 0 & 0 & 0 & 0 & \cdots & \nu_1 & \nu_0 & \nu_1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & \nu_2 & \nu_1 & \nu_0 \end{pmatrix}.$$

Terms $v_{i,j}$ are computed thanks to Bartlett's formula, expressed in Equation (3). For more simplicity, we can also use the property given in [16] in Equation (7.2.6):

$$v_{i,j} = \operatorname{cov}\left(\sum_{k=1}^{\infty}(\rho(k+i) + \rho(k-i) - 2\,\rho(i)\,\rho(k))U_k\,; \sum_{l=1}^{\infty}(\rho(l+j) + \rho(l-j) - 2\,\rho(j)\,\rho(l))U_l\right),$$

where $(U_k)_k$ are iid $\mathcal{N}(0,1)$ random variables. Thus,

$$\begin{aligned} v_{1,1} &= 1 + 2\,\rho(2)^2 - 3\,\rho(1)^2 + 2\,\rho(2) + 4\,\rho(1)^4 - 8\,\rho(1)^2\,\rho(2) + 4\,\rho(1)^2\,\rho(2)^2 \\ v_{2,2} &= 1 + 2\,\rho(1)^2 - 3\,\rho(2)^2 + 4\,\rho(1)^2\,\rho(2)^2 + 4\,\rho(2)^4 - 4\,\rho(1)^2\,\rho(2) \\ v_{1,2} &= 2\,\rho(1) - 2\,\rho(1)\,\rho(2) - 4\,\rho(1)\,\rho(2)^2 + 4\,\rho(1)\,\rho(2)^3 + 4\,\rho(1)^3\,\rho(2) - 2\,\rho(1)^3\,. \end{aligned}$$

We use these values when we apply the Kolmogorov–Smirnov test.

*Appendix A.2. Relationship Between $a_q$ and $\rho(q)$*

Let $(Z_t)_t$ be an MA($q$) process defined by

$$Z_t = \sum_{k=0}^{q} a_k\,\mathcal{E}_{t-k}\,,$$

where $a_0 = 1$ and $(Z_t)_t$ is a WN. Then, we know that $\rho(h) = 0$, $\forall\,|h| \geq q$, and

$$\rho(q) = \frac{a_q}{\sum_{k=0}^{q} a_k^2} \neq 0\,.$$

In our simulations, we take $q = 2$, $a_1 = \frac{1}{2}$, and we choose $a_2 \neq 0$. In particular,

$$\rho(2) = \frac{a_2}{\frac{5}{4} + a_2^2}.$$

In Figure A1, we display $\rho(2)$ versus $a_2$. The closer $a_2$ is to 0, the closer $\rho(2)$ is to 0 also. Let us denote $\alpha = \frac{\sqrt{5}}{2} \simeq 1.12$. When $a_2 = -\alpha$, then $\rho(2)$ reaches its minimum value, equal to $-\frac{\sqrt{5}}{5}$. And symmetrically, when $a_2 = \alpha$, then $\rho(2)$ reaches its maximum value, equal to $\frac{\sqrt{5}}{5}$. In our simulations, we choose $a_2 = \frac{1}{10}$, which is rather close to 0 and $a_2 = \frac{1}{2}$, which is further, and leads to a greater value of $\rho(2)$.

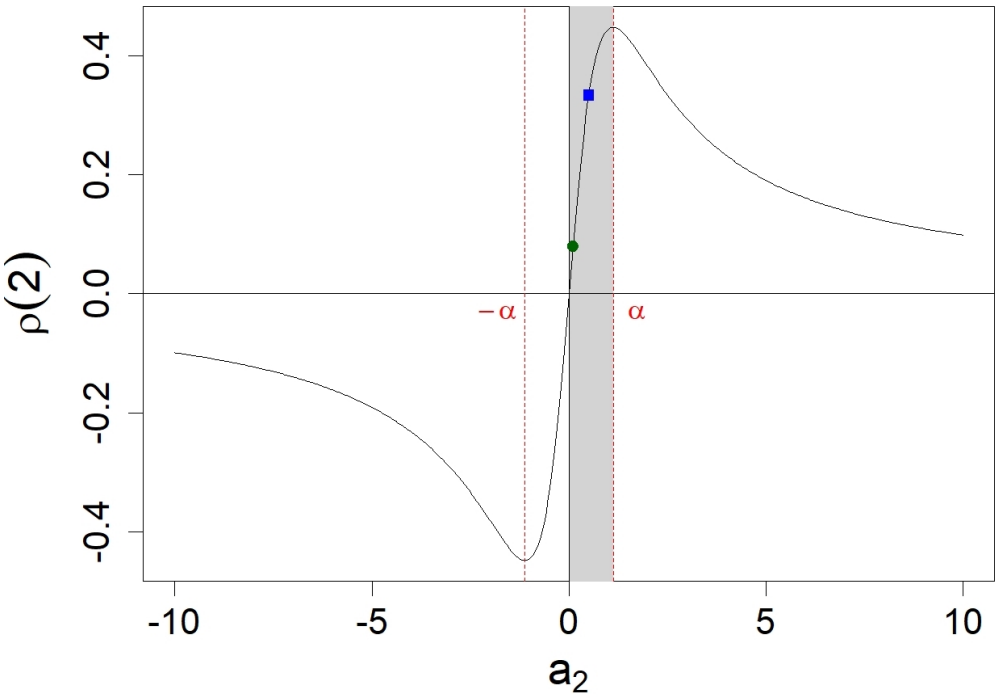


**Figure A1.** $\rho(2)$ versus $a_2$. The values of $a_2$ leading to the minimal and the maximal values of $\rho(2)$, $-\alpha$ and $\alpha$ are written in red. Red vertical lines are plotted at these coordinates. The green circle points for $a_2 = \frac{1}{10}$, whereas the blue square points for $a_2 = \frac{1}{2}$. Moreover, we have shaded the range of the $a_2$ parameters used to study the impact of this coefficient on the ability to identify the order $q$ of simulated MA(2) processes.

## Appendix B. Implementation Details About EACF

EACF function has been implemented in the **R** language, available on the project website: www.i2m.univ-amu.fr/perso/manuela.royer-carenzi/AnnexesR.SacfMA/SacfMA.html (accessed on 19 February 2025).

We provide details about its implementation in Algorithms A1 and A2.

**Algorithm A1** Computation of extended autocorrelation function coefficients.

1: **Input**: Matrix $m$ of residual coefficients, autocovariances $cov1$, maximum autoregressive order $nar$, number of columns $ncol$, lag count $count$, total covariances $ncov$, centered series $z$, lagged matrix $zm$

2: **Output**: Vector of EACF coefficients and correction factor $\lambda_m$

3: Initialize vectors:

$$eacf \leftarrow \mathbf{0} \text{ of size } (nar + 1)$$
$$\lambda_m \leftarrow \mathbf{1} \text{ of size } (nar + 1)$$

4: Compute first coefficient:

$$eacf[1] \leftarrow cov1[ncov + count]$$

5: **for** $i \in [\![1, nar]\!]$ **do**

6: Compute transformed residuals:

$$\text{temp} \leftarrow \begin{bmatrix} z[-(1:i)] & zm[-(1:i), 1:i] \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -m[1:i, i] \end{bmatrix}$$

7: Compute EACF coefficient:

$$eacf[i+1] \leftarrow \text{acf}(\text{temp}, \text{lag.max} = count)[\text{count} + 1]$$

8: Compute correction factor $\lambda_m$:

$$\lambda_m[i+1] \leftarrow 1 + \sum_{j=2}^{\text{count}+1} \text{acf}(\text{temp})[j]^2$$

9: **end for**

10: **Output**: Return $(eacf, \lambda_m)$

**Algorithm A2** Extended Autocorrelation Function for ARMA(p,q) order selection.

**Input**: Time series $z$, maximum orders $ar\_max$ and $ma\_max$
**Output**: Symbol matrix identifying $(p, q)$ orders
**Initialization:**
$ar\_max \leftarrow ar\_max + 1$
$ma\_max \leftarrow ma\_max + 1$
$nar \leftarrow ar\_max - 1$, $nma \leftarrow ma\_max$
$ncov \leftarrow nar + nma + 2$
$nrow \leftarrow nar + nma + 1$
$ncol \leftarrow nrow - 1$
Center the series: $z \leftarrow z - \text{mean}(z)$
Construct lag matrix:
**for** $i \in [\![1, nar]\!]$ **do**
    Store $z$ lagged by $i$ in $zm$
**end for**
Compute autocovariance function:

$$cov1 \leftarrow \text{acf}(z)$$

**AR Model Estimation:**
**for** $i \in [\![1, nar]\!]$ **do**
    Fit an AR model of order $i$:

$$m1[:, i] \leftarrow \text{ar.ols}(z, \text{order} = i)$$

**end for**
**Moving Average (MA) Computation:**
Initialize empty matrices: $eacfm$ and $\lambda_m$
**for** $i \in [\![1, nma]\!]$ **do**
    Perform recursive update on $m1$:

$$m2 \leftarrow \text{reupm}(m1)$$

    Compute EACF coefficients and $\lambda_m$:

$$(eacfm[:, i], \lambda_m[:, i]) \leftarrow \text{ceascf}(m2, cov1, nar, ncol, i, ncov, z, zm)$$

    Update $m1 \leftarrow m2$
**end for**
**Compute Significance Thresholds:**
**for** $i \in [\![1, nma]\!]$ **do**
    Compute symbol matrix based on adjusted threshold:

$$\text{symbol}[:, i] \leftarrow \begin{cases} \text{“}x\text{”}, & \text{if } |eacfm[:, i]| > 1.96 \cdot \sqrt{\lambda_m[:, i]/\text{length}(z)} \\ \text{“}o\text{”}, & \text{otherwise} \end{cases}$$

**end for**
**Output:**
Display the symbol matrix representing $(p, q)$ order selections
Return $(eacfm, ar\_max, ma\_max, \text{symbol matrix})$

## Appendix C. Impact on Order-Identification for Simulations of MA(2) Processes of Length $n = 500$

Figure 5 gives the frequency of each last lag when $n = 100$, and Figure A2 when $n = 500$.

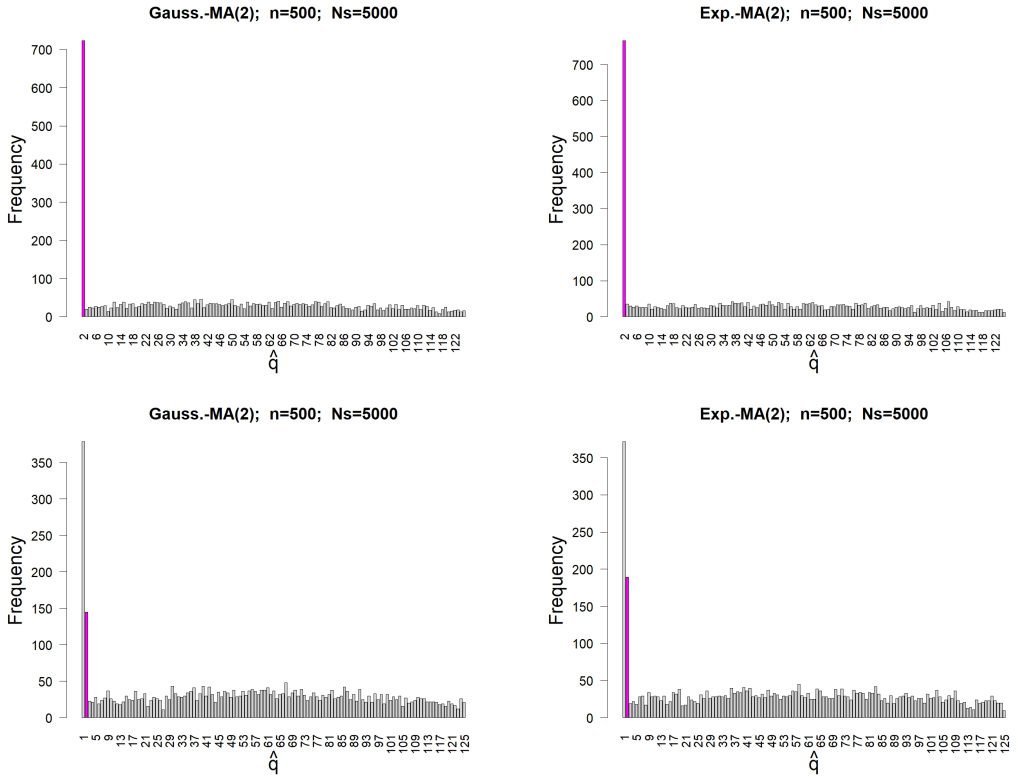


**Figure A2.** Frequency of $\hat{q}$, the last lag $h$, such that $\hat{\rho}(h)$ lies out of interval $\hat{\mathcal{J}}_h$, as described in Proposition 3, for $N_S = 5000$ simulations of MA(2) processes of length $n = 500$. The left column concerns simulated MA(2) processes with an underlying Gaussian WN, whereas the right one deals with Exponential WN. In the top figures, we simulate MA(2) processes with $a_2 = \frac{1}{2}$, whereas we use $a_2 = \frac{1}{10}$ in the bottom. The magenta-colored bar highlights the theoretical convenient order $\hat{q} = 2$.

## Appendix D. Comparison between MA(1) and MA(2) models estimated on MA(2) simulations

**Table A1.** Comparison of models quality when the procedure suggests $\hat{q} = 1$. For any MA(2) simulation where the procedure suggests $\hat{q} = 1$, MA(1) and MA(2) models are estimated. Prediction criteria (RMSE and MAPE) and information criteria (AIC, AICc, and BIC) are compared with the one-tailed Student's paired $t$-test. When the null is rejected, the favored model, either MA(1) or MA(2), is indicated. Otherwise, the "==" symbol is given. Moreover, we add the $p$-value in brackets. When no MA(1) model has been suggested by the procedure, a "-" symbol has been used. The associated WN is Gaussian and coefficient $a_2$ is either $a_2 = \frac{1}{2}$ (left column) or $a_2 = \frac{1}{10}$ (right column).

| | $a_2 = \frac{1}{2}$ | | $a_2 = \frac{1}{10}$ | |
|---|---|---|---|---|
| **Criteria** | **n = 100** | **n = 500** | **n = 100** | **n = 500** |
| **RMSE** | MA(2) $(1.6\times10^{-4})$ | $-(-)$ | == (0.4148) | == (0.5304) |
| **MAPE** | MA(1) $(6.7\times10^{-6})$ | $-(-)$ | MA(1) (0.0247) | MA(1) (0.0437) |
| **AIC** | MA(2) $(<10^{-16})$ | $-(-)$ | MA(1) $(<10^{-16})$ | MA(2) $(<10^{-16})$ |
| **AICc** | MA(2) $(<10^{-16})$ | $-(-)$ | MA(1) $(<10^{-16})$ | MA(2) $(<10^{-16})$ |
| **BIC** | MA(2) $(<10^{-16})$ | $-(-)$ | MA(1) $(<10^{-16})$ | MA(1) $(<10^{-16})$ |

**Table A2.** Comparison of models quality when the procedure suggests $\hat{q} = 2$. For any MA(2) simulation where the procedure suggests $\hat{q} = 1$, MA(1) and MA(2) models are estimated. Prediction criteria (RMSE and MAPE) and information criteria (AIC, AICc and BIC) are compared with the Student's paired $t$-test. When the null is rejected, the favored model, either MA(1) or MA(2), is indicated. Otherwise, the "==" symbol is given. Moreover, we add the $p$-value in brackets. The associated WN is Gaussian, and coefficient $a_2$ is either $a_2 = \frac{1}{2}$ (left column) or $a_2 = \frac{1}{10}$ (right column).

| | $a_2 = \frac{1}{2}$ | | $a_2 = \frac{1}{10}$ | |
|---|---|---|---|---|
| **Criteria** | **n = 100** | **n = 500** | **n = 100** | **n = 500** |
| **RMSE** | MA(2) ($<10^{-16}$) | MA(2) (0.0040) | == (0.6536) | == (0.5257) |
| **MAPE** | == (0.4298) | == (0.0665) | == (0.9965) | == (0.1486) |
| **AIC** | MA(2) ($<10^{-16}$) | MA(2) (0.0040) | MA(2) ($<10^{-16}$) | MA(2) ($<10^{-16}$) |
| **AICc** | MA(2) ($<10^{-16}$) | MA(2) ($<10^{-16}$) | MA(2) ($<10^{-16}$) | MA(2) ($<10^{-16}$) |
| **BIC** | MA(2) ($<10^{-16}$) | MA(2) ($<10^{-16}$) | MA(2) ($3\times10^{-5}$) | MA(2) ($<10^{-16}$) |

## Appendix E. Diagnosis for Several Models Fitted on Wind Speed Data

### *Appendix E.1. MA(4) Model*

Figure A3 displays the graphical diagnosis associated to an MA(4) model fitted on wind speed data: ACF (top left) and standardized SACF (top right) computed on the residuals, and Ljung–Box tests $p$-values (bottom).

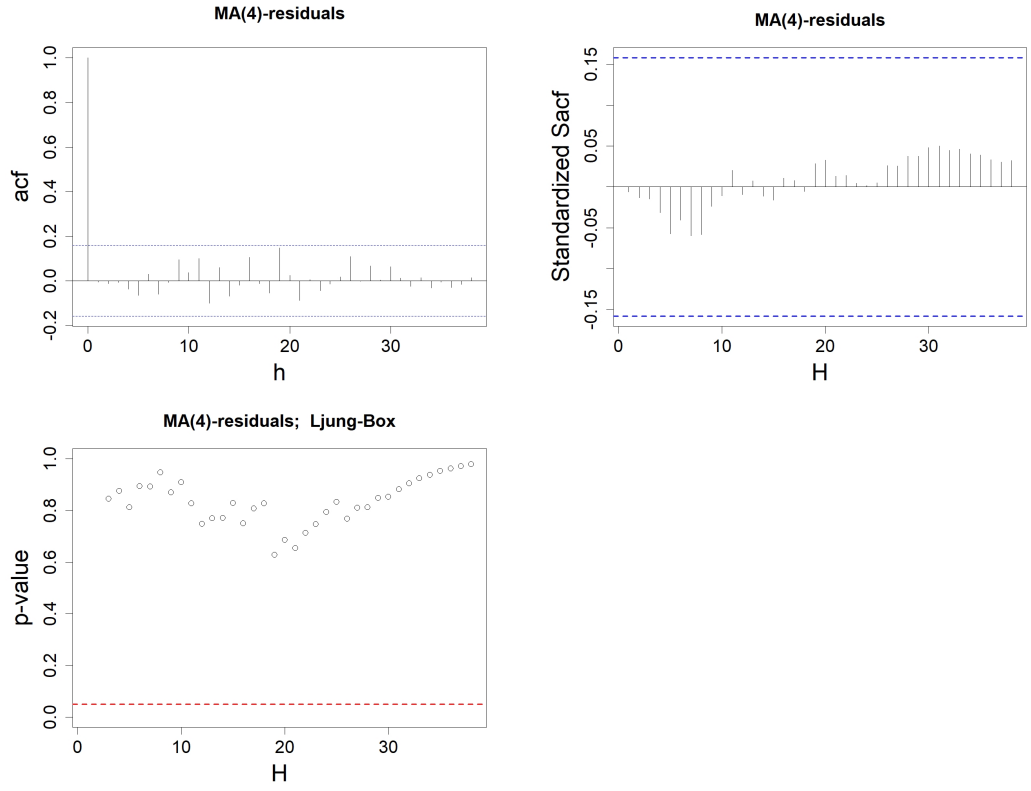


**Figure A3.** Graphical diagnosis for an MA(4) model fitted on wind speed data. ACF (**top left**) and Standardized SACF (**top right**) of the residuals from the MA($q$) model constructed on wind speed data, with $h$ varying from 1 to $\frac{n}{4}$. The blue-dotted horizontal lines represent the thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$. $p$-values when using Ljung–Box's test on the residuals associated to an MA(4) model (**bottom**). Ljung–Box's test is computed successively for all lags $H = q+1, \cdots, 38$, with degree of freedom $H - q$. The red-dotted horizontal line represents 5%.

*Appendix E.2. ARMA(p, q) Models Suggested by the EACF Procedure*

We check that the ARMA($p, q$) models suggested by our EACF procedure and the ARIMA(3,1,1) suggested by the automated procedure `auto.arima` are valid. Figure A4 displays the ACF of the residuals, Figure A5 their standardized SACF, and Figure A6 the Ljung–Box test $p$-values.

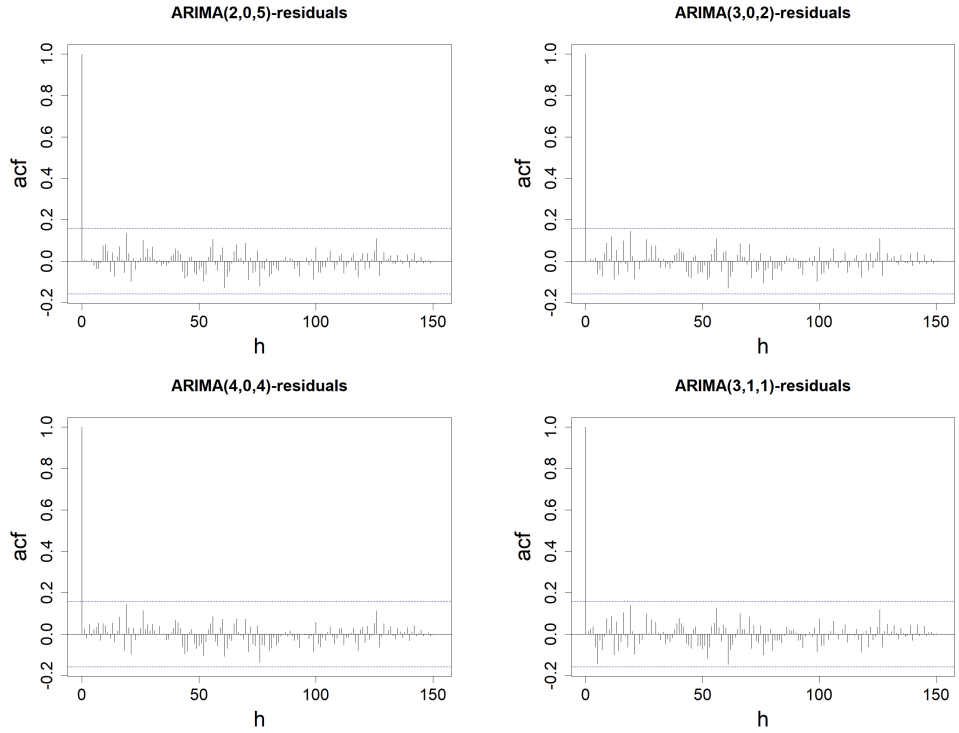


**Figure A4.** ACF computed on the residuals of the ARMA(2,5) (**top left**), ARMA(3,2) (**top right**), ARMA(4,4) (**bottom left**) and ARIMA(3,1,1) (**bottom right**), constructed on wind speed data, with $h$ varying from 1 to $\frac{n}{4}$. The blue-dotted horizontal lines represent the thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$.

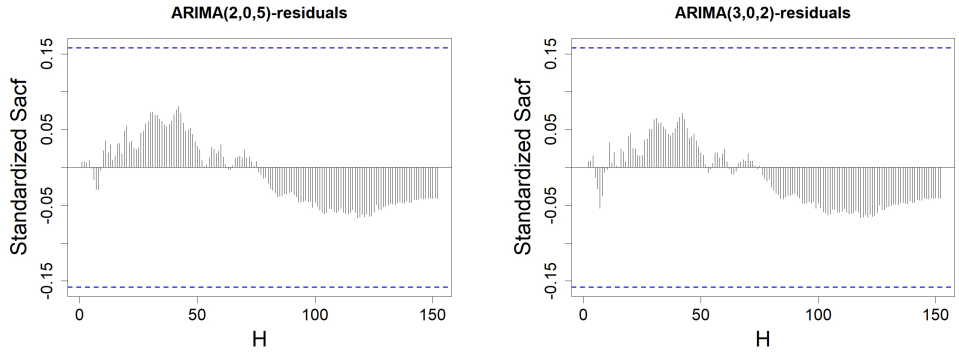


**Figure A5.** *Cont.*

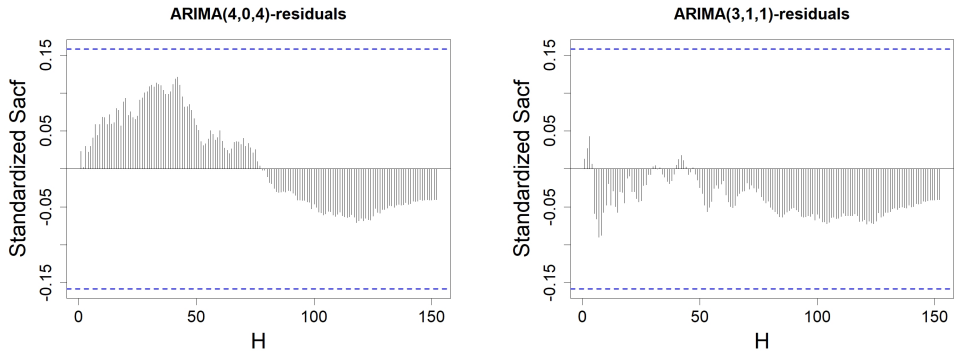


**Figure A5.** Standardized SACF computed on the residuals of the ARMA(2,5) (**top left**), ARMA(3,2) (**top right**), ARMA(4,4) (**bottom left**) and ARIMA(3,1,1) (**bottom right**), constructed on wind speed data, with $h$ varying from 1 to $\frac{n}{4}$. The blue-dotted horizontal lines represent the thresholds $-1.96/\sqrt{n}$ and $1.96/\sqrt{n}$.

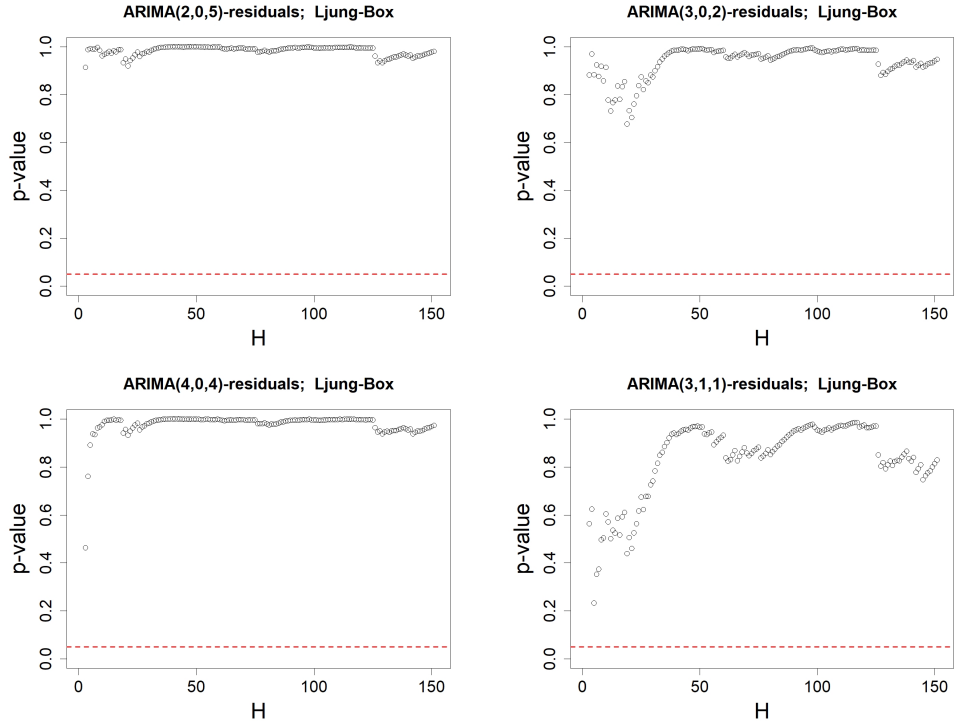


**Figure A6.** $p$-values when using Ljung–Box's test on the residuals of the ARMA(2,5) (**top left**), ARMA(3,2) (**top right**), ARMA(4,4) (**bottom left**) and ARIMA(3,1,1) (**bottom right**), constructed on wind speed data, for all lags $H = q + 1, \cdots, 38$, with degree of freedom $H - q$. The red-dotted horizontal line represents 5%.

## References

1. Nelson, C.R.; Plosser, C.I. Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. *J. Monet. Econ.* **1982**, *10*, 139–162. [CrossRef]
2. Pena, D.; Rodriguez, J. Forecasting Traffic Flow by Using Time Series Models. *Transp. Rev.* **2001**, *21*, 293–317.
3. Tsay, R. *Analysis of Financial Time Series*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 2010.
4. Ogunlana, S.O.; Oyebisi, T.O. Modelling and Forecasting Nigerian Electricity Demand Using Univariate Box-Jenkins Approach. *J. Energy Technol. Policy* **2013**, *3*, 84–91.
5. Elsaraiti, M.; Musbah, H.; Merabet, A.; Little, T. Time Series Analysis of Electricity Consumption Forecasting Using ARIMA Model. In Proceedings of the 2021 IEEE Green Technologies Conference (GreenTech), Denver, CO, USA, 7–9 April 2021; pp. 259–262.
6. Teyssière, G.; Kirman, A. Microeconomic models for long memory in the volatility of financial time series. *Physica A* **2002**, *370*, 26–31.
7. Glass, G.V.; Willson, V.L.; Gottman, J.M. Design and Analysis of Time-Series Experiments. *Annu. Rev. Psychol.* **1975**, *26*, 609–653.

8. Arunachalam, V.; Jaafar, A. Forecasting Dengue Incidence in Penang, Malaysia: A Comparison of ARIMA and GARCH Models. *Am. J. Trop. Med. Hyg.* **2011**, *85*, 827–833.
9. Luis, C.O.; Francisco, G.S.; Jose, M.S. Forecasting of Emergency Department Admissions. *Healthc. Manag. Sci.* **2012**, *15*, 215–224.
10. Zheng, X.; Basher, R.E. Structural Time Series Models and Trend Detection in Global and Regional Temperature Series. *J. Clim.* **1999**, *12*, 2347–2358. [CrossRef]
11. Campbell, J.Y.; Perron, P. An Empirical Investigation of the Relations between Climate Change and Agricultural Yield: A Time Series Analysis of Maize Yield in Nigeria. *J. Agric. Environ. Sci.* **2004**, *5*, 217–230.
12. Sanei, S.; Hassani, H. *Singular Spectrum Analysis of Biomedical Signals*; CRC Press: Boca Raton, FL, USA, 2015. [CrossRef]
13. Box, G.; Pierce, D. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *J. Am. Statist. Assoc.* **1970**, *65*, 1509–1526. [CrossRef]
14. Ljung, G.; Box, G. On a Measure of a Lack of Fit in Time Series Models. *Biometrika* **1978**, *65*, 297–303. [CrossRef]
15. Priestley, M.B. Spectral Analysis and Time Series. *J. Time Ser. Anal.* **1981**, *2*, 85–106. [CrossRef]
16. Brockwell, P.; Davis, R. *Time Series: Theory and Methods*, 2nd ed.; Springer: New York, NY, USA, 1991.
17. Hamilton, J.D. Time Series Analysis. *Econom. Rev.* **1994**, *13*, 147–192.
18. Chatfield, C. *The Analysis of Time Series: An Introduction*; CRC Press: Boca Raton, FL, USA, 2003.
19. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
20. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications: With R Examples*; Springer: Berlin/Heidelberg, Germany, 2006.
21. Wei, W.W.S. *Time Series Analysis Univariate and Multivariate Methods*, 2nd ed.; Addison Wesley: New York, NY, USA, 2006.
22. Hassani, H. Sum of the sample of autocorrelation function. *Random Oper. Stoch. Eqs.* **2009**, *17*, 125–130.
23. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: Cham, Switzerland, 2016.
24. Hyndman, R.J.; Athanasopoulos, G. Forecasting: Principles and Practice. *Int. J. Forecast.* **2018**, *34*, 587–590. Available online: https://otexts.com/fpp2/ (accessed on 20 January 2025).
25. Kendall, M.G. *Time-Series*; Oxford University Press: Oxford, UK, 1976.
26. Bisaglia, L.; Gerolimetto, M. Testing for Time Series Linearity Using the Autocorrelation Function. *Stat. Methods Appl.* **2009**, *18*, 23–50.
27. Boutahar, M.; Royer-Carenzi, M. Identifying trends nature in time series using autocorrelation functions and stationarity tests. *Int. J. Econ. Econom.* **2024**, *14*, 1–22.
28. McLeod, A.I.; Zhang, Y. Partial Autocorrelation Parameterization for Seasonal ARIMA Models. *Int. J. Forecast.* **2006**, *22*, 661–673.
29. Granger, C.W.J.; Joyeux, R. An Introduction to Long-Memory Time Series Models and Fractional Differencing. *J. Time Ser. Anal.* **1980**, *1*, 15–29. [CrossRef]
30. Hosking, J. Asymptotic distribution of the sample mean, autocovariances, autocorrelations of long-memory time series. *J. Econom.* **1996**, *73*, 261–284. [CrossRef]
31. Boutahar, M.; Marimoutou, V.N.L. Estimation Methods of the Long Memory Parameter: Monte Carlo Analysis and Application. *J. Appl. Stat.* **2007**, *34*, 261–301. [CrossRef]
32. Hassani, H.; Leonenko, N.; Paterson, K. The sample autocorrelation function and the detection of long-memory processes. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 6367–6379. [CrossRef]
33. Hassani, H.; Yarmohammadi, M.; Mashald, L. Uncovering hidden insights with long-memory-proscess detection: An in-depth overview. *Risks* **2023**, *11*, 113. [CrossRef]
34. Phojanamongkolkij, N.; Kato, S.; Wielicki, B.A.; Taylor, P.C.; Mlynczak, M.G. A Comparison of Climate Signal Trend Detection Uncertainty Analysis Methods. *J. Clim.* **2014**, *27*, 3363–3376. [CrossRef]
35. Dimitriadis, P.; Koutsoyiannis, D. Climacogram versus Autocovariance and Power Spectrum in Stochastic Modelling for Markovian and Hurst-Kolmogorov Processes. *Stoch. Environ. Res. Risk Assess.* **2015**, *15*, 1649–1669. [CrossRef]
36. Xie, Y.; Liu, S.; Fang, H.; Wang, J. Global Autocorrelation Test Based on the Monte Carlo Method and Impacts of Eliminating Nonstationary Components on the Global Autocorrelation Test. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 1645–1658. [CrossRef]
37. Liu, S.; Xie, Y.; Fang, H.; Du, H.; Xu, P. Trend Test for Hydrological and Climatic Time Series Considering the Interaction of Trend and Autocorrelations. *Water* **2022**, *14*, 3006. [CrossRef]
38. Belmahdi, B.; Louzazni, M.; El Bouardi, A. One month-ahead forecasting of mean daily global solar radiation using time series models. *Optik* **2020**, *219*, 165207. [CrossRef]
39. Gostischa, J.; Massolo, A.; Constantine, R. Multi-species feeding association dynamics driven by a large generalist predator. *Front. Mar. Sci.* **2021**, *8*, 739894.
40. Yang, Y.; Qin, S.; Liao, S. Ultra-chaos of a mobile robot: A higher disorder than normal-chaos. Chaos. *Solitons Fractals* **2023**, *167*, 113037. [CrossRef]
41. Orlando, G.; Bufalo, M. Empirical evidences on the interconnectedness between sampling and asset returns's distributions. *Risks* **2021**, *9*, 88. [CrossRef]

42. Bai, M.; Zhou, Z.; Chen, Y.; Liu, J.; Yu, D. Accurate four-hour-ahead probabilistic forecast of photovoltaic power generation based on multiple meteorological variables-aided intelligent optimization of numeric weather prediction data. *Earth Sci. Inform.* **2023**, *16*, 2741–2766. [CrossRef]
43. Wang, X.; Yang, J.; Yang, F.; Wang, Y.; Liu, F. Multilevel residual prophet network time series model for prediction of irregularities on high-speed railway track. *J. Transp. Eng. Part A Syst.* **2023**, *149*, 04023012. [CrossRef]
44. Hassani, H.; Royer-Carenzi, M.; Mashhad, L.M.; Yarmohammadi, M.; Yeganegi, M.R. Exploring the depths of autocorrelation function: Its departure from normality. *Information* **2024**, *15*, 449. [CrossRef]
45. Bartlett, M.S. On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Suppl. J. R. Stat. Soc.* **1946**, *8*, 27–41. [CrossRef]
46. Tsay, R.; Tiao, G. Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *J. Am. Stat. Assoc.* **1984**, *79*, 84–96 [CrossRef]
47. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 3rd ed.; Prentice Hall: Englewood Clifs, NJ, USA, 1994.
48. Shapiro, S.; Wilk, M. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
49. Dallal, G.E.; Wilkinson, L. An analytic approximation to the distribution of Lilliefors' test for normality. *Am. Stat.* **1986**, *40*, 294–296. [CrossRef]
50. Kolmogorov, A. Sulla determinazione empirica di una legge di distribuzione. *Giorn. Inst. Ital. degli Att.* **1933**, *4*, 83–91.
51. Smirnov, N. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.* **1948**, *19*, 279–281.
52. Akaike, H. Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*; Petrov, B.N., Csaki, F., Eds.; Springer: Budapest, Hungary, 1973; Volume 1557, pp. 30–32.
53. Sugiura, N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat.-Theory Methods* **1978**, *7*, 13–26. [CrossRef]
54. Burnham, K. and Anderson, D. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. In *Second International Symposium on Information Theory*, 2nd ed.; Springer: New York, New York, USA, 2002; Volume 1557, pp. 30–32.
55. Schwartz, G.E. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464.
56. Li, W.K. *Diagnostic Checks in Time Series*; Chapman & Hall: New York, NY, USA, 2004.
57. Anderson, O. The box-jenkins approach to time series analysis. *RAIRO Rech. OpÉrationnelle* **1977**, *11*, 3–29. [CrossRef]
58. Hassani, H.; Yeganegi, M. Selecting optimal lag order in Ljung-Box test. *Physica A* **2020**, *541*, 123700. [CrossRef]
59. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *2*, 1–22.
60. Hassani, H.; Mashhad, L.M.; Royer-Carenzi, M.; Yeganegi, M.R.; Komendantova, N. White Noise and Its Misapplications: Impacts on Time Series Model Adequacy and Forecasting. *Forecasting* **2025**, *7*, 8. [CrossRef]
61. Huang, B.; Thorne, P.W.; Banzon, V.F.; Boyer, T.; Chepurin, G.; Lawrimore, J.H.; Menne, M.J.; Smith, T.M.; Vose, R.S.; Zhang, H.-M. *NOAA Extended Reconstructed Sea Surface Temperature (ERSST)*; NOAA National Centers for Environmental Information: Asheville, NC, USA, 2017.
62. Canova, F.; Hansen, B.E. Are seasonal patterns constant over time? A test for seasonal stability. *J. Bus. Econ. Stat.* **1995**, *13*, 237–252. [CrossRef]
63. Hansen, J.; Ruedy, R.; Sato, M.; Lo, K. Global surface temperature change. *Rev. Geophys.* **2010**, *48*, RG4004. [CrossRef]
64. Lenssen, N.; Schmidt, G.; Hansen, J.; Menne, M.; Persin, A.; Ruedy, R.; Zyss, D. Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.* ***2019****, 124, 6307–6326. [CrossRef]*