

The Table Auto-Regressive Moving-Average Model for (Categorical) Stationary Series: Mathematical Perspectives (Invertibility; Maximum Likelihood Estimation)

Chrysoula Dimitriou-Fakalou

Independent Researcher, Athens, Greece

Email: my.submission@mail.com

How to cite this paper: Dimitriou-Fakalou, C. (2022) The Table Auto-Regressive Moving-Average Model for (Categorical) Stationary Series: Mathematical Perspectives (Invertibility; Maximum Likelihood Estimation). *Open Journal of Statistics*, 12, 385-407. <https://doi.org/10.4236/ojs.2022.123025>

Received: May 23, 2022

Accepted: June 26, 2022

Published: June 29, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Once invertibility for a causal TARMA series is defined and accompanied by conditions on the probability parameters of the model, all focus concentrates on the maximum likelihood estimators. Under the coexistence of essential causality and invertibility, the estimators are shown to be convergent to the real values and asymptotically obedient to the Gaussian distribution: their variance matrix identifies with a classic result. Some real-like examples are simulated and simplification attempts include the derivation of the non-parametric chi-square test extension for stationary TAR series.

Keywords

Asymptotic Normality, Consistency, Scores Efficiency, Strict Stationarity

1. Introduction

The scientific progress, from defining a valid time series model to making it useful in practice, depends on the consolidation of inference results. With regard to the non-linear stationary time series, the general TARMA model ([1]) has recently pledged an unprecedented flexibility: given the mild requirement to categorize the variables' values, its competence to express any serial order conditional or joint probability dependence is unquestionable. So it will be the aim of this paper to bond the theoretical with the most deserving TARMA sample properties. Before that, a scent of other results and popular models in the field is offered.

The DARMA model undoubtedly triggered a significant number of scientists'

interest for the discrete time series analysis: the reader may skim through [2] to get acquainted with the definition and some asymptotic properties. A recent fix to replication issues of the older model, can be discovered in [3]. Having to do with an additive model though, it should be crystal clear that its (and those pieces' following it) simplicity and parsimony contribution to the discrete stationary series modelling, are due to its inarguable impairment to manage any better than the marginal and covariance dependence (which is not the case for the TARMA).

A valuable equivalence has been established in the past between the analysis of Bernoulli variables and the strictly stationary series obeying any law ([4]): this saves the trouble of generalizing the Gaussian ARMA as in [5] for a family of infinitely divisible distributions. Nevertheless, most of the derivations concern the Gaussian stationarity, such as [6] who deals with the alternative (via 0 - 1) estimation of autoregressive parameterizations: those binary data have traditionally benefited from a rich bibliography anyway.

A promotion to the maximum likelihood estimation together with some solid statements can be found in [7], but the dense and short paper uses an approximation to the fixed Markovian or auto-regressive order; the inclusion of a moving-average part only presents itself with the Gaussian ARMA (1, 1) illustration, but it is exactly converting the AR to the ARMA that makes the problem challenging, ever more so when the distribution is not Gaussian. In the best estimators' properties department, it is worth reading [8] to view the well-known ARMA efficiency result.

Notable attempts have taken place for count series as well; besides modelling as usual, [9] reproduced standard discrete-time renewal process inference results. In the other hand there are the INARMA models, for example, the INAR (1) (with random coefficient) asymptotic behaviour has been studied extensively by [10]. A recent review on the subject of count time series is [11]. One can always incorporate a count variable into a categorical one, encouraging the reader to elect the TARMA model as superior for the series stationarity. For bilinear models (and 'any'-valued variables) there has been a plethora of inference results too.

Though a distinction wall is often raised between the time series and the Markov chain schemes, the causal TAR model embodies the homogeneous Markov dependence attached to a unique stationary distribution. Hence, [12] were the original contributors for the sake of the Markov chain inference. Later, [13] restricted the derivations to two states basic chain dependence managing explicit efficiency-related results, when (for two parameters only) the inverse of a square matrix can be computed with ease.

In this paper, the main objective is to perform the TARMA parameters inference, filling the gap of estimation for the infinite Markovian dependence: Section 2 summarizes the TARMA definition together with some raw statements concerning the use of the model parameters for a stationary series. Then Section

3 newly defines the invertible TARMA series and Section 3.1 invents a condition that safeguards invertibility, using a parallel with the existing definition and condition for causality. Section 3.2 continues with the invertibility topic, delivering news on the limit of probabilities based on a finite to the countably infinite past. Then all Section 4 is to establish the two grand properties of the maximum likelihood estimators: the weak convergence to the parameters is in Section 4.1, while Section 4.2 uses numerous lemmas and a theorem to conclude with the asymptotic normality, as this is explicitly stated in the beginning of Section 6; there, the normally distributed estimator vector transforms to a chi-square statistic to test a null hypothesis on the specified TARMA model fitting stationarity. The purely theoretical presentations are quickly reminded (Sections 2, 3), established (Sections 3, 4 and 6) or discussed (Section 6). In addition, the simulation Section 5 examines the performance of TARMA estimators (or how to compute the estimates) for small sample sizes, and returns favorable conclusions as well.

2. Reminders

Previously in [1], a strictly stationary time series $\{X_t, t \in \mathbb{Z}\}$ was considered: the variables can be of any (bounded or unbounded) range, which for the serial stationarity clothing *must* be assigned to $(k+1)$ categories (for fixed $k \in \mathbb{N}$) with conventional (or, natural, if that is the case) category codes, say, 0 and $v_1, \dots, v_k \neq 0$. The X -variables are built on a multivariate sequence of independent in time and identically distributed $(k+1)^{p+q} \times 1$ random vectors: for each $\mathbf{i} = (i_1, \dots, i_p)$, $\mathbf{j} = (j_1, \dots, j_q)$, $i_1, \dots, i_p, j_1, \dots, j_q = 0, v_1, \dots, v_k$, a univariable $I_t^{(i,j)}$ ($I_t^{(0_p, 0_q)} \equiv I_t$) with same range as X is considered at time t , and marginally it has to hold $\{I_t^{(i,j)}\} \sim \text{IID}$ with probabilities

$$\pi_x^{(i,j)} = \mathbb{P}(I_t^{(i,j)} = x) \in (0, 1), x = 0, v_1, \dots, v_k, \quad \sum_{x=0, v_1, \dots, v_k} \pi_x^{(i,j)} = 1$$

$$(\text{and } \pi_x^{(0_p, 0_q)} \equiv \pi_x).$$

It is reminded that a multivariate IID time series is such that the random vectors are independent in time only, or that any two univariables indexed at different times are independent: the variables' dependence at the same time within the same vector will be referred to as *interdependence* with *interindependence* being a special case; hence the $\pi_x^{(i,j)}$ above are marginal probabilities that do not necessarily suffice to determine the joint dependence at the same timing.

In the general setting, $\{I_t^{(i,j)} = x\}$ will simplify to the event that the variable $I_t^{(i,j)}$ belongs into the category with code $x = 0, v_1, \dots, v_k$. The Markov chain terminology of a state space, say $\mathcal{S} := \{0, v_1, \dots, v_k\}$ and $\dots \in \mathcal{S}$ will be avoided deliberately. Then the definition introduced in [1] is reminded here.

For fixed $p, q \in \mathbb{N}_0$ (such that $p+q \in \mathbb{N}$), $\{X_t, t \in \mathbb{Z}\}$ is a Table Auto-Regressive Moving-Average process of order (k, p, q) , and it is written $\{X_t\} \sim \text{TARMA}(k, p, q)$, if (it is causal and invertible and) it holds that

$$X_t := I_t^{((X_{t-1} \cdots X_{t-p})(I_{t-1} \cdots I_{t-q}))}, \text{ for every } t \in \mathbb{Z}. \tag{1}$$

In explanation, X_t is set equal to a $I_t^{(\cdot)}$ (at the original range); nevertheless, which I (out of the $(k+1)^{p+q}$) contributes its value, depends on the previous $X_{t-1}, \dots, X_{t-p}, I_{t-1}, \dots, I_{t-q}$, *i.e.*, which category each one of these $(p+q)$ variable realization falls into.

As an *example* of a TARMA (1, 1), the real-valued variables may be categorized into $k+1=3$ groups, say $(-\infty, -3]$, $(-3, 3)$ and $[3, \infty)$ with codes “-3”, “0” and “3”, respectively; then the (9×1) random vectors $I_t^{(ij)}$, $i, j = -3, 0, 3$ are considered with fixed distribution for every $t \in \mathbb{Z}$, and then over time with $I_t^{(ij)}$ being independent of $I_{t^*}^{(i^*j^*)}$ for $t \neq t^*$. The X_t is defined to be equal to $I_t^{(-3|-3)}$ if both X_{t-1} and $I_{t-1}^{(0|0)}$ are less than or equal to -3 , or equal to $I_t^{(0|-3)}$ if $-3 < X_{t-1} < 3$ and $I_{t-1}^{(0|0)} \leq -3$, or...., or equal to $I_t^{(3|3)}$ if both X_{t-1} and $I_{t-1}^{(0|0)}$ are greater than or equal to 3.

Hopefully, the example above clarifies that it is not necessarily a count series that is under study. This is an ‘occurrence-not occurrence’ analysis of categorical variables and, for the serial evolution, it matters not whether X_t is set equal to the exact value or the relevant category code: it is a probability of occurrence analysis and it works subject to the categorization (together with k of course).

Then (1) can be re-arranged as

$$f_x(X_t) := \sum_{i_1, \dots, i_p, j_1, \dots, j_q=0, v_1, \dots, v_k} f_x(I_t^{(ij)}) \prod_{l=1}^p f_{i_l}(X_{t-l}) \prod_{n=1}^q f_{j_n}(I_{t-n}), \tag{2}$$

for $x = 0, v_1, \dots, v_k$,

where for $x, y = 0, v_1, \dots, v_k$, the index functions

$$f_x(y) = \frac{\prod_{x^*=0, v_1, \dots, v_k, x^* \neq x} (y - x^*)}{\prod_{x^*=0, v_1, \dots, v_k, x^* \neq x} (x - x^*)} = \begin{cases} 1, & \text{if } y = x \\ 0, & \text{if } y = 0, v_1, \dots, v_k, y \neq x \end{cases}$$

have been set.

For (2) to hold for every $t \in \mathbb{Z}$, at the very least the assumption of uniqueness of a stationary distribution solution should be applicable; in fact, $\{X_t, t \in \mathbb{Z}\}$ should not be considered at all, unless it is a (table) causal process (based on $\{I_t\}$) according to the definition in [1]. The causality element will be revisited here together with the new element of invertibility.

Since the definition of this model has been the focus of past work, the reader is encouraged to look for examples and form a clearer picture from there ([1]). The understanding and appreciation of the TARMA definition must have already taken place to proceed with this paper’s inference related goals. For the reader that wishes to derive the auto-covariance function of TARMA series, it is highlighted that the second moments stationarity is a special case of the all moments stationarity: the work of [1] has handed over a methodology to compute *all joint probabilities*, and hence the auto-covariance function is a special case, as it demands the joint probabilities of two variables. Not only are those derivations outside the scope of *this* paper, but also taming the all- rather than just second-

order stationarity is the great TARMA contribution over other models.

From this point on, it will be taken for granted (and without a formal proof) that the TARMA construction may represent any strictly stationary time series, particularly under the convention- if there needs to be one- of tabulating the variables' values. Together with a solution on identification (of k and p, q) issues, the TARMA theory is a contribution to the non-parametric version of statistical science. When $q = 0$ the dependence is p th order Markovian, while for $q \geq 1$ a valid Markovian dependence of infinite order is achieved; when $p = 0$, a q -dependent strictly stationary series is concerned, and the joint dependence limits to infinity for $p \geq 1$. Hence it will be considered ([1]) that, subject to the categorization, the TARMA can be the all-moments stationary analogue to the ARMA model for the covariance stationary series.

Some Basic Requirements

The probabilities π will be taking the role of parameters of the model, for which the estimation results will be established. The conditions in [1] are remembered:

(C1): $\{I_t^{(ij)}, t \in \mathbb{Z}\}$ are jointly interdependent series, i.e., it holds that

$$\mathbb{P}\left(I_t^{(i_1 j_1)} = x_1, \dots, I_t^{(i_n j_n)} = x_n\right) = \prod_{m=1}^n \mathbb{P}\left(I_t^{(i_m j_m)} = x_m\right), t \in \mathbb{Z}$$

for any $n = 2, \dots, (k+1)^{p+q}$, $(i_m, j_m) = (i_{m,1}, \dots, i_{m,p}, j_{m,1}, \dots, j_{m,q})$, $i_{m,1}, \dots, i_{m,p}, j_{m,1}, \dots, j_{m,q} = 0, v_1, \dots, v_k$, $m = 1, \dots, n$, $(i_{m_1}, j_{m_1}) \neq (i_{m_2}, j_{m_2})$, $m_1, m_2 = 1, \dots, n$ ($m_1 \neq m_2$) and $x_1, \dots, x_n = 0, v_1, \dots, v_k$.

Condition (C1) enforces the $I_t^{(\dots)}$ variables (at the same time t) to be independent: otherwise, the form of their interdependence should accompany (1), in order to properly define a TARMA process. Under an interdependence scenario, the computation of all joint TARMA probabilities heavily relies on

$$\pi_{x|y}^{*(ij)} := \mathbb{P}\left(I_t^{(ij)} = x \mid I_t = y\right), x, y = 0, v_1, \dots, v_k, (i, j) \neq \mathbf{0}_{p+q}$$

(as well as π) according to [1]. Hence in this case, π^* are also parameters for an apposite TARMA definition.

In the absence of (C1), an “ h to $h + 1$ ” way to plex the interdependence may work by first considering π_x and then gradually building for $n = 1, \dots, p + q$ and every $l_1, \dots, l_n = 1, \dots, p + q$ ($l_1 < \dots < l_n$), $i_{l_1}, \dots, i_{l_n} = v_1, \dots, v_k$, the conditional probabilities $\mathbb{P}\left(I_t^{(0_{l_1-1} i_{l_1} \dots i_{l_n} 0_{p+q-l_n})} = x_n \mid I_t = x, \dots\right)$: “ \dots ” in the condition

sets all $I_t^{(0_{l_1^*-1} i_{l_1^*} \dots i_{l_n^*} 0_{p+q-l_n^*})}$, $n^* = 1, \dots, n-1$, $l_1^*, \dots, l_n^* = l_1, \dots, l_n$ ($l_1^* < \dots < l_n^*$).

That way one gets a feel of how (C1) can be replaced, especially when other attributes need to be accomplished.

Other conditions in [1] are listed below:

(C2): $\{X_t, t \in \mathbb{Z}\}$ is not a deterministic process, i.e.,

$$\mathbb{P}(X_t = x \mid X_{t-1} = x_1, X_{t-2} = x_2, \dots) \in (0, 1)$$

for all $x, x_n = 0, v_1, \dots, v_k, n \in \mathbb{N}$.

(C3): $\{X_t, t \in \mathbb{Z}\}$ is *not overparameterized*, i.e., in the simplest of cases, it cannot be that $\pi_x^{(i|j)} = \pi_x^{(i'|j')}$, $x = 0, v_1, \dots, v_k$ for $(i, j) \neq (i', j')$.

It would be misleading that the process $\{X_t\}$ is of order (p, q) with $(k + 1)$ categories, when in fact there are fewer parameters that are active (due to identical distributions), so this is preserved by (C3): the reader may reflect on the extreme example that *all* distributions are identical and $\{X_t\}$ is in fact an IID process! Nevertheless, the usual (opposite) notion that the different I -distributions should be as alike as possible, will be fundamental for the process validity according to other conditions (causality/invertibility) as in the next section.

Finally, it is stressed that the assumption of a convoluted multivariate IID time series, which is attached to the definition of the TARMA, is the necessary block for the inference that will be presented in this paper, the same way that the univariate error IID series is for the well-known ARMA.

3. Invertibility

In [1], the process $\{X_t, t \in \mathbb{Z}\}$ was called *causal* (based on $\{I_t, t \in \mathbb{Z}\}$) also implying that it is strictly stationary. After setting

$$\begin{aligned} & d^h f_x \left(I_t^{(0_{h-1}, i_{h-1}, \dots, 0_{2-h-1}, i_{2-h-1}, \dots, i_h, 0_{p+q-h})} \right) \\ & := f_x \left(I_t^{(0_{h-1}, i_{h-1}, \dots, 0_{2-h-1}, i_{2-h-1}, \dots, i_h, 0_{p+q-h})} \right) - \sum_{n=1}^h f_x \left(I_t^{(0_{h-1}, i_{h-1}, \dots, i_{n-1}, 0_{n+1-h}, \dots, i_{n+1}, \dots, i_h, 0_{p+q-h})} \right) \\ & \quad + \dots + (-1)^{h-1} \sum_{n=1}^h f_x \left(I_t^{(0_{n-1}, i_{n-1}, \dots, 0_{p+q-n})} \right) + (-1)^h f_x(I_t) \end{aligned}$$

for $h = 1, \dots, p + q$, $l_1, \dots, l_h = 1, \dots, p + q$ ($l_1 < \dots < l_h$), $i_{l_1}, \dots, i_{l_h} = v_1, \dots, v_k$, it can be derived from (2), that

$$\begin{aligned} f_x(I_t) &= f_x(X_t) - \sum_{h_p=1}^p \sum_{\substack{l_1, \dots, l_{h_p}=1 \\ l_1 < \dots < l_{h_p}}} \sum_{i_{h_p} = v_1, \dots, v_k} d^{h_p} f_x \left(I_t^{(0_{h-1}, i_{h-1}, \dots, i_{h_p}, 0_{p-h_p})} \right) \\ & \cdot \left(\prod_{r=1}^{h_p} f_{i_r}(X_{t-l_r}) \right) - \sum_{h_q=1}^q \sum_{h_p=0}^p \sum_{\substack{l_1, \dots, l_{h_p}=1 \\ l_1 < \dots < l_{h_p}}} \sum_{\substack{n_1, \dots, n_{h_q}=1 \\ n_1 < \dots < n_{h_q}}} \sum_{i_{h_p}, \dots, i_{h_p}, j_{n_1}, \dots, j_{n_{h_q}} = v_1, \dots, v_k} d^{h_p+h_q} \\ & \cdot f_x \left(I_t^{(0_{h-1}, i_{h-1}, \dots, i_{h_p}, 0_{p-h_p})} (0_{n_1-1}, j_{n_1}, \dots, j_{n_{h_q}}, 0_{q-n_{h_q}}) \right) \left(\prod_{r=1}^{h_q} f_{j_{n_r}}(I_{t-n_r}) \right) \left(\prod_{r=1}^{h_p} f_{i_r}(X_{t-l_r}) \right); \end{aligned} \tag{3}$$

similarly was done for causality ([1]) and the representation

$$\begin{aligned} f_x(X_t) &= f_x(I_t) + \sum_{h \in \mathbb{N}} \sum_{n_1, \dots, n_h \in \mathbb{N}} \sum_{\substack{j_{n_1}, \dots, j_{n_h} = v_1, \dots, v_k \\ n_1 < \dots < n_h}} f_x \Psi_{t, (n_1, \dots, n_h)}(j_{n_1}, \dots, j_{n_h}) \\ & \cdot \left(\prod_{r=1}^h f_{j_{n_r}}(I_{t-n_r}) \right). \end{aligned} \tag{4}$$

Here, a new substitution of $\left(\prod_{r=1}^{h_q} f_{j_{n_r}}(I_{t-n_r}) \right)$ and so on, will also lead to a countably infinite representation (of the upcoming form (5)). An adequate defi-

dition for *invertibility* follows.

Definition 3.1: The process $\{X_t, t \in \mathbb{Z}\}$ as defined in (1) will be called *invertible*, in the sense that it can be written

$$f_x(X_t) = f_x(I_t) + \sum_{h \in \mathbb{N}} \sum_{\substack{l_1, \dots, l_h \in \mathbb{N} \\ l_1 < \dots < l_h}} \sum_{i_{l_1}, \dots, i_{l_h} = v_1, \dots, v_k} f_x \Phi_{t, (l_1, \dots, l_h)}(i_{l_1}, \dots, i_{l_h}) \cdot \left(\prod_{r=1}^h f_{i_{l_r}}(X_{t-l_r}) \right) \tag{5}$$

for $x = v_1, \dots, v_k$ and $t \in \mathbb{Z}$; the random variable $f_x \Phi_{t, (l_1, \dots, l_h)}(i_{l_1}, \dots, i_{l_h})$, ($h \geq 1$) is *independent* of $I_{t+n}^{(\dots)}$, $n \in \mathbb{N}$, $I_{t-l}^{(\dots)}$, $l \geq l_h$, $l \in \mathbb{N}$ (it is a function of $I_{t-l}^{(ij)}$, $0 \leq l \leq l_h - 1$) and remains unchanged for $t \in \mathbb{Z}$.

Additionally, it *must* hold that the probabilities

$$\mathbb{P}(X_t = x | X_{t-l} = i_l, l \in \mathbb{N}), i_l = 0, v_1, \dots, v_k, l \in \mathbb{N} \tag{6}$$

can be *uniquely determined* from (5).

Remark 1: For $h, h^*, h_L \in \mathbb{N}$, $h_L \leq h^* \leq h$, $l_1, \dots, l_h \in \mathbb{N}$, ($l_1 < \dots < l_h$), consider $\{l_1^*, \dots, l_{h_L}^* \equiv l_{h^*}^*\} \subseteq \{l_1, \dots, l_h\}$, $l_1^* < \dots < l_{h_L}^*$ and any $i_{l_1}, \dots, i_{l_h} = v_1, \dots, v_k$, and write the events

$$\mathcal{B}_{(t,h), (l_1, \dots, l_h), (i_{l_1}, \dots, i_{l_h})} := \{I_{t-l_1} = i_{l_1}, \dots, I_{t-l_h} = i_{l_h}, I_{t-l} = 0, l \in \mathbb{N}, l \neq l_1, \dots, l_h\},$$

$$\mathcal{C}_{(t,h), (l_1, \dots, l_h), (i_{l_1}, \dots, i_{l_h})} := \{X_{t-l_1} = i_{l_1}, \dots, X_{t-l_h} = i_{l_h}, X_{t-l} = 0, l \in \mathbb{N}, l \neq l_1, \dots, l_h\}$$

and the probability of interest

$$\mathbb{P}\left(f_x \Phi_{t, (l_1^*, \dots, l_{h_L}^*)}(i_{l_1^*}, \dots, i_{l_{h_L}^*}) \leq y | \mathcal{C}_{(t,h), (l_1, \dots, l_h), (i_{l_1}, \dots, i_{l_h})}\right).$$

As opposed to the probability

$$\begin{aligned} &\mathbb{P}\left(f_x \Psi_{t, (l_1^*, \dots, l_{h_L}^*)}(i_{l_1^*}, \dots, i_{l_{h_L}^*}) \leq y | \mathcal{B}_{(t,h), (l_1, \dots, l_h), (i_{l_1}, \dots, i_{l_h})}\right) \\ &\equiv \mathbb{P}\left(f_x \Psi_{t, (l_1^*, \dots, l_{h_L}^*)}(i_{l_1^*}, \dots, i_{l_{h_L}^*}) \leq y | I_{t-l_1} = i_{l_1}, \dots, I_{t-l_{h^*}^*} = i_{l_{h^*}^*}, \right. \\ &\quad \left. I_{t-l} = 0, 1 \leq l < l_{h^*}^*, l \neq l_1, \dots, l_{h^*}^*\right) \end{aligned}$$

from the ‘‘causality’’ topic in [1] when it is used that $\{I_t^{(ij)}\} \sim \text{IID}$ ($f_x \Psi_{t, (l_1, \dots, l_h)}$ are functions of $I_{t-l}^{(ij)}$, $0 \leq l < l_h$), in the case of ‘‘invertibility’’ $\{X_t\}$ is a serially dependent series. Nevertheless, a simplification is in order, *i.e.*, it holds that

$$\begin{aligned} &\mathbb{P}\left(f_x \Phi_{t, (l_1^*, \dots, l_{h_L}^*)}(i_{l_1^*}, \dots, i_{l_{h_L}^*}) \leq y | \mathcal{C}_{(t,h), (l_1, \dots, l_h), (i_{l_1}, \dots, i_{l_h})}\right) \\ &= \mathbb{P}\left(f_x \Phi_{t, (l_1^*, \dots, l_{h_L}^*)}(i_{l_1^*}, \dots, i_{l_{h_L}^*}) \leq y | \Phi_{(t-1,h), (l_1-1, \dots, l_h-1), (i_{l_1}, \dots, i_{l_h})}^* = 0, \dots \right. \\ &\quad \left. \Phi_{(t-1,h-1), (l_2-1, \dots, l_h-1), (i_2, \dots, i_{l_h})}^* = i_{l_1}, \dots, \Phi_{(t-1,h-1), (l_h-1, l_h-1), (i_{l_h})}^* = i_{l_{h-1}}, \dots \right. \\ &\quad \left. \Phi_{(t-l_h+1,1), (1, i_{l_h})}^* = 0, I_{t-l_h} = i_{l_h}\right) \end{aligned} \tag{7}$$

as it can be shown (together with the notation Φ^*).

3.1. Conditions Relating to Causality and Invertibility

To perform the TARMA parameters inference, there will be the need to somehow “squeeze” the random coefficients $f_x \Phi_{t,(i_1, \dots, i_h)}$ to become smaller as $l_h \rightarrow \infty$.

To manage convergence results, first set the remainder

$$\begin{aligned}
 f_x R_{t,>l} &:= \sum_{r=l+1}^{\infty} \sum_{i_r=v_1, \dots, v_k} f_x \Phi_{t,r}(i_r) \cdot f_{i_r}(X_{t-r}) \\
 &+ \sum_{\eta=1}^l \sum_{\nu_2=l+1}^{\infty} \sum_{i_{\eta}, i_{\nu_2}=v_1, \dots, v_k} f_x \Phi_{t,(\eta, \nu_2)}(i_{\eta}, i_{\nu_2}) \cdot f_{i_{\eta}}(X_{t-\eta}) f_{i_{\nu_2}}(X_{t-\nu_2}) \\
 &+ \sum_{\substack{\eta, \nu_2=l+1 \\ \eta < \nu_2}}^{\infty} \sum_{i_{\eta}, i_{\nu_2}=v_1, \dots, v_k} f_x \Phi_{t,(\eta, \nu_2)}(i_{\eta}, i_{\nu_2}) \cdot f_{i_{\eta}}(X_{t-\eta}) f_{i_{\nu_2}}(X_{t-\nu_2}) + \dots
 \end{aligned}$$

from the “invertible” representation (5). Similarly, it has been set by [1] that

$$\begin{aligned}
 f_x r_{t,>l} &:= \sum_{r=l+1}^{\infty} \sum_{j_r=v_1, \dots, v_k} f_x \Psi_{t,r}(j_r) \cdot f_{j_r}(I_{t-r}) \\
 &+ \sum_{\eta=1}^l \sum_{\nu_2=l+1}^{\infty} \sum_{j_{\eta}, j_{\nu_2}=v_1, \dots, v_k} f_x \Psi_{t,(\eta, \nu_2)}(j_{\eta}, j_{\nu_2}) \cdot f_{j_{\eta}}(I_{t-\eta}) f_{j_{\nu_2}}(I_{t-\nu_2}) \\
 &+ \sum_{\substack{\eta, \nu_2=l+1 \\ \eta < \nu_2}}^{\infty} \sum_{j_{\eta}, j_{\nu_2}=v_1, \dots, v_k} f_x \Psi_{t,(\eta, \nu_2)}(j_{\eta}, j_{\nu_2}) \cdot f_{j_{\eta}}(I_{t-\eta}) f_{j_{\nu_2}}(I_{t-\nu_2}) + \dots
 \end{aligned}$$

from the “causal” representation (4).

Consider generic constants $C > 0$ and $\alpha \in (0, 1)$. The condition is presented below:

(C4): The parameters of the TARMA equation are such that it holds that

$$\mathbb{E} \left\{ \left| f_x R_{t,>l} \right| \mid X_{t-n} = x_n, n \in \mathbb{N} \right\} \leq C \cdot \alpha^{l+1}, \text{ for } x = v_1, \dots, v_k, \tag{8}$$

$t \in \mathbb{Z}, l \in \mathbb{N}_0$ and any $x_n = 0, v_1, \dots, v_k, n \in \mathbb{N}$.

(C5): The parameters of the TARMA equation are such that it holds that

$$\mathbb{E} \left| f_x r_{t,>l} \right| \leq C \cdot \alpha^{l+1}, \text{ for } x = v_1, \dots, v_k, t \in \mathbb{Z}, l \in \mathbb{N}_0. \tag{9}$$

To justify (C4), see that it can be written that

$$f_x(X_t) = f_x(I_t) + \sum_{l \in \mathbb{N}_0} (f_x R_{t,>l} - f_x R_{t,>l+1}),$$

so that the conditional probability of interest (6) is bounded by

$$\mathbb{P}(I_t = x \mid X_{t-n} = i_n, n \in \mathbb{N}) + \sum_{l \in \mathbb{N}_0} C^* \cdot \alpha^{l+1},$$

where a converging geometric series is involved and, under causality, it is

$\mathbb{P}(I_t = x \mid X_{t-n} = i_n, n \in \mathbb{N}) \equiv \pi_x$; straight from (8), it holds that

$$\begin{aligned}
 &\mathbb{E} \left\{ \left| f_x R_{t,>l} - f_x R_{t,>l+1} \right| \mid X_{t-n} = i_n, n \in \mathbb{N} \right\} \\
 &\leq \mathbb{E} \left\{ \left| f_x R_{t,>l} \right| \mid X_{t-n} = i_n \right\} + \mathbb{E} \left\{ \left| f_x R_{t,>l+1} \right| \mid X_{t-n} = i_n \right\}, \text{ so that it has been inserted} \\
 &C^* := C \cdot (1 + \alpha).
 \end{aligned}$$

Regarding (C5) and how it secures that all joint probabilities can be safely

bounded, the answer is easy to show.

Since (C4) (as opposed to (C5)) involves conditional expectations, it might be wished for $|f_x R_{t,>l}|$ given “...” to simplify the condition “...” from $\{X_{t-n} = x_n, n \in \mathbb{N}\}$ to $\{X_{t-n} = x_n, n = 1, \dots, l\}$. According to the argument laid out above, it is in fact the absolute value of

$$f_x R_{t,>l} - f_x R_{t,>l+1} \equiv \sum_{i_{l+1}=v_1, \dots, v_k} f_x \Phi_{t,l+1}(i_{l+1}) \cdot f_{i_{l+1}}(X_{t-(l+1)}) \\ + \sum_{\eta=1}^l \sum_{i_{\eta}, i_{\eta+1}=v_1, \dots, v_k} f_x \Phi_{t,(\eta,l+1)}(i_{\eta}, i_{\eta+1}) \cdot f_{i_{\eta}}(X_{t-\eta}) f_{i_{\eta+1}}(X_{t-(l+1)}) + \dots$$

(instead of that of $f_x R_{t,>l}$) that is needed, which is “contained” within the random coefficients $f_x \Phi_{t,(l_1, \dots, l_h)}$, $l_1 < \dots < l_h \leq l+1$ (that are functions of $I_{t-n}^{(\dots)}$, $n = 0, 1, \dots, l$). In the case that $X_{t-n} = 0$, $n \geq l+1$, it is clear from Remark 1 that (under causality) a simplification is possible, though not concerning the variables X in the condition: otherwise, it'd better not be attempted. The interested reader may look at Section 3.2 to verify what happens in the general case *given the requirement in (C4) as it is*.

For causality, due to the difficulties in determining the random coefficients $f_x \Psi_t$ (as functions of $I_{t-l}^{(\dots)}$, $l \in \mathbb{N}_0$), [1] resorted to an alternative (to (4)) representation: then based on the new (rather than (4)) form, salvaged a condition relating to causality. The equivalent representation for invertibility may be demonstrated with a Proposition 1 and Propositions 2 and 3 may lead to an invertibility-related condition. Nevertheless, it might be stressed here that Proposition 2 (and, consequently, Proposition 3) is established using the prerequisite of causality; this is because an “invertible” representation relies on the index variables $f(X_{t-l})$ (not $f(I_{t-l})$) and, under causality, it can be certified that X_{t-l} , $l \in \mathbb{N}$ is independent of $I_t^{(\dots)}$.

Furthermore, remember that the causality consideration is attached to the definition of a TARMA process, as it guarantees probability stationarity, which is what this theory is about: so this is before the results for the inference are sought. In the contrary, the contribution of invertibility will shine, when the weak convergence (consistency) of the maximum likelihood estimators for the relevant TARMA probabilities is established. Nevertheless, it is not wished to undermine the value of invertibility as compared to that of causality: after all, both type coefficients $f_x \Psi_t$ and $f_x \Phi_t$ are built as functions of the $I^{(\dots)}$ variables from present and past, using a similar mechanism. The conditions obtained seem to complete rather than contradict each other too, so it is clear that causality and invertibility need to work together for the TARMA body to stand straight.

The reader needs to acquaint themselves with the following notation: it is set for $x = v_1, \dots, v_k$, that

$$\gamma_x^{(v)} := \begin{cases} \max \left\{ \mathbb{E} \left| d^v f_x \left(I_t^{(0_p | (0_{m-1} j_{m-1} \dots j_{n_v} 0_{q-n_v}))} \right) \right| \right\}, & \text{if } v = 1, \dots, q \\ \max_{i,j} \left\{ \pi_x^{(ij)} \right\}, & \text{if } v = 0 \end{cases},$$

as well as for each $h_p = 1, \dots, p, v = 0, 1, \dots, q$, that

$$\gamma_{p,x}^{(v)}(h_p) := \max \left\{ \mathbb{E} \left| d^{h_p+v} f_x \left(I_t \left(\left(\mathbf{0}_{h_1-1} \ i_{h_1} \dots i_{h_p} \ \mathbf{0}_{p-h_p} \right) \left(\mathbf{0}_{n_1-1} \ j_{n_1} \dots j_{n_v} \ \mathbf{0}_{q-n_v} \right) \right) \right) \right| \right\},$$

where both maximums take place for any $n_1, \dots, n_v = 1, \dots, q$ ($n_1 < \dots < n_v$), $j_{n_1}, \dots, j_{n_v} = v_1, \dots, v_k$; the second also taking place for any $l_1, \dots, l_{h_p} = 1, \dots, p$ ($l_1 < \dots < l_{h_p}$), $i_{l_1}, \dots, i_{l_{h_p}} = v_1, \dots, v_k$. It is defined for $x = v_1, \dots, v_k$

$$\gamma_x := \max_{v=1, \dots, q} \left\{ \gamma_x^{(v)} + \sum_{h_p=1}^p \binom{p}{h_p} \gamma_{p,x}^{(v)}(h_p) \right\} \text{ and}$$

$$\gamma^* := \max_{v=0, \dots, q} \left\{ \left(\sum_{x=v_1, \dots, v_k} \gamma_x^{(v)} \right) + \sum_{h_p=1}^p \binom{p}{h_p} \left(\sum_{x=v_1, \dots, v_k} \gamma_{p,x}^{(v)}(h_p) \right) \right\}.$$

By relaxing the sum of maximums into a maximum of sums, $\sum_{x=v_1, \dots, v_k} \gamma_x^{(0)}$ is replaced by $\max_{i,j} \{1 - \pi_0^{(ij)}\}$. More research is welcome on the subject of the equivalence with (C4), which was demonstrated to be sufficient for invertibility.

Remark 2: (i) Under (C4) and (C5), all joint probabilities and all probabilities $\mathbb{P}(X_t = x | X_{t-n} = x_n, n \in \mathbb{N})$, $x, x_n = 0, v_1, \dots, v_k$ are contained (away from infinity): it can be concluded that *any joint probability is away from zero (hence away from one as well)*, because if any such probability, say $\mathbb{P}(\mathcal{X}_t^{(a)})$, was 0 it would have to be that any conditional probability as above- with $\mathcal{X}_t^{(a)} \subseteq \{X_{t-n} = x_n, n \in \mathbb{N}\}$ for some t^* - would not be bounded and properly defined; then it can be concluded (by division of a joint non zero probability over a joint non infinity probability) that *all conditional probabilities are strictly larger than zero (hence smaller than one as well)*. Consequently, it will be taken that (C4) and (C5) can suffice for (C2).

(ii) For the π that yield a causal and invertible TARMA series $\{X_t\}$, it can be arranged under (8) in (C4) (or, (9) in (C5)) that

$$\left| \frac{\partial}{\partial \pi} \mathbb{E}(f_x R_{t,>l} | X_{t-n} = x_n, n \in \mathbb{N}) \right| \leq C \cdot \alpha^{l+1}, \left(\text{or } \left| \frac{\partial}{\partial \pi} \mathbb{E}(f_x r_{t,>l}) \right| \leq C \cdot \alpha^{l+1} \right) \quad (10)$$

for some $0 < C < \infty$ and $\alpha \in (0, 1)$, and

$$\left| \frac{\partial^2}{\partial \pi \partial \pi} \mathbb{E}(f_x R_{t,>l} | X_{t-n} = x_n, n \in \mathbb{N}) \right| \leq C \cdot \alpha^{l+1} \left(\text{or } \left| \frac{\partial^2}{\partial \pi \partial \pi} \mathbb{E}(f_x r_{t,>l}) \right| \leq C \cdot \alpha^{l+1} \right) \quad (11)$$

under (8) (or (9)): this can be shown.

It is being taken for granted that the $\frac{\partial}{\partial \pi} \mathbb{P}(X_t = x | X_{t-n} = x_n, n \in \mathbb{N}, \pi)$, $\frac{\partial^2}{\partial \pi \partial \pi} \mathbb{P}(X_t = x | X_{t-n} = x_n, n \in \mathbb{N}, \pi)$ derivatives exist.

3.2. Random Coefficient Modelling Based on the Past

For $l, n \in \mathbb{N}_0, n \leq l$ and a set $\{l_1^*, \dots, l_n^*\} \subseteq \{1, \dots, l\}$, it is written (for fixed $i_1^*, \dots, i_n^* = v_1, \dots, v_k$ but this is omitted from the symbol \mathcal{N}), that

$$\mathcal{N}_{t,n \leq l} := \left\{ X_{t-l_1^*} = i_{l_1^*}, \dots, X_{t-l_n^*} = i_{l_n^*}, X_{t-l^*} = 0, l^* = 1, \dots, l, l^* \neq l_1^*, \dots, l_n^* \right\}.$$

Eventually it can be shown that

$$\lim_{l \rightarrow \infty} \mathbb{P} \left(\left| \mathbb{P} \left(X_t = x \mid \mathcal{N}_{t,n \leq l}, X_{t-l-l^*}, l^* \in \mathbb{N} \right) - \mathbb{P} \left(X_t = x \mid \mathcal{N}_{t,n \leq l} \right) \right| \leq C \cdot \alpha^{l+1} \right) = 1 \quad (12)$$

($n, l_1^*, \dots, l_n^*, i_{l_1^*}, \dots, i_{l_n^*}$ remain unchanged as $l \rightarrow \infty$).

Remark 3: From (10) and (11) and in the same manner, it will be considered occasionally that

$$\lim_{l \rightarrow \infty} \mathbb{P} \left(\left| \frac{\partial}{\partial \pi} \mathbb{P} \left(X_t = x \mid \mathcal{N}_{t,n \leq l} \right) - \frac{\partial}{\partial \pi} \mathbb{P} \left(X_t = x \mid \mathcal{N}_{t,n \leq l}, X_{t-l^*}, l^* \geq l+1 \right) \right| \leq C \cdot \alpha^{l+1} \right) = 1 \quad (13)$$

and

$$\lim_{l \rightarrow \infty} \mathbb{P} \left(\left| \frac{\partial^2}{\partial \pi \partial \pi} \mathbb{P} \left(X_t = x \mid \mathcal{N}_{t,n \leq l} \right) - \frac{\partial^2}{\partial \pi \partial \pi} \mathbb{P} \left(X_t = x \mid \mathcal{N}_{t,n \leq l}, X_{t-l^*}, l^* \geq l+1 \right) \right| \leq C \cdot \alpha^{l+1} \right) = 1, \quad (14)$$

respectively. A justification may be offered.

4. Maximum Likelihood Estimation

First, the parameter space, say Θ , must be considered; the fixed $k \in \mathbb{N}$, categories v_1, \dots, v_k and order $p, q \in \mathbb{N}_0$, $p+q \in \mathbb{N}$ are attached to it:

The parameter space Θ is a set that includes all candidate parameter vectors $\pi \in \Theta$ that model a process of interest $\{X_t, t \in \mathbb{Z}\}$ using a TARMA (k, p, q) equation with a predetermined form of interdependence, such that conditions (C3), (C4) and (C5) are satisfied (plus any extra requirements added in this section).

Now, suppose that $\{X_1, \dots, X_T\}$ have been made available from (1), and the target is to estimate the true parameters of the model, say $\pi_0 \in \Theta$, based on these observations. The likelihood expressed as a function of π , takes the form

$$L(\pi) = \prod_{t=1}^T (\tilde{p}_{t,t-1}(\pi)),$$

where

$$\tilde{p}_{t,0}(\pi) := \mathbb{P}(X_t = x; \pi), \text{ if } X_t = x, \text{ where } x = 0, v_1, \dots, v_k,$$

and for the general $i \in \mathbb{N}$, it is written

$$\tilde{p}_{t,i}(\pi) := \mathbb{P}(X_t = x \mid X_{t-1}, \dots, X_{t-i}; \pi), \text{ if } X_t = x, \text{ where } x = 0, v_1, \dots, v_k;$$

$\tilde{p}_{t,i}$ are functions of X_{t-1}, \dots, X_{t-i} as well as X_t and it is considered that these (conditional) probabilities result from $\{X_t, t \in \mathbb{Z}\}$ as if those have been generated by $\pi \in \Theta$. Of course, the natural logarithm of the likelihood may be taken

$$l(\pi) = \sum_{t=1}^T \ln(\tilde{p}_{t,t-1}(\pi))$$

as the maximum likelihood estimators $\hat{\pi}$ then satisfy the equations

$$\frac{\partial}{\partial \pi} l(\pi) \Big|_{\pi=\hat{\pi}} = 0 \text{ or } \sum_{t=1}^T \frac{1}{\tilde{p}_{t,t-1}(\hat{\pi})} \frac{\partial}{\partial \pi} \tilde{p}_{t,t-1}(\pi) \Big|_{\pi=\hat{\pi}} = 0. \quad (15)$$

4.1. Weak Convergence of Estimators

For $\pi \in \Theta$ and any $t \in \mathbb{Z}$, it is written

$$p_t(\pi) := \mathbb{P}(X_t = x | X_{t-i}, i \in \mathbb{N}; \pi), \text{ if } X_t = x, \text{ where } x = 0, v_1, \dots, v_k,$$

which is a function of X_{t-i} , $i \in \mathbb{N}_0$ and the (conditional) distribution law is generated by $\pi \in \Theta$.

According to (12) (thanks to *invertibility*) and for any $\pi \in \Theta$ and $x_1, \dots, x_T = 0, v_1, \dots, v_k$, it holds as $T \rightarrow \infty$ that

$$\left(\frac{\mathbb{P}(X_1 = x_1; \pi)}{\mathbb{P}(X_1 = x_1 | X_{1-n}, n \in \mathbb{N}; \pi)} \frac{\mathbb{P}(X_2 = x_2 | X_1 = x_1; \pi)}{\mathbb{P}(X_2 = x_2 | X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \pi)} \dots \frac{\mathbb{P}(X_T = x_T | X_{T-1} = x_{T-1}, \dots, X_1 = x_1; \pi)}{\mathbb{P}(X_T = x_T | X_{T-1} = x_{T-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \pi)} \right)^{1/T} \xrightarrow{P} 1, \tag{16}$$

because the geometric mean weighs with (rather than just multiplies) the previous values: as the new values get closer to 1, so must the mean itself. The geometric mean from $(T + 1)$ observations is a weighted product mean of the geometric mean of the T observations and the $(T + 1)$ th observation (with a C_A adjusted constant for the ratio of probabilities to one, by allowing the exponential rate for the geometric mean of ratios, this becomes apparent with the bound

$$\left\{ C_A^T \cdot \alpha^{\frac{(T+1)T}{2}} \right\}^{1/T} \rightarrow 0 \text{ as } T \rightarrow \infty.$$

Once the fixated sample series x_1, \dots, x_T has been collected (this has been generated by π_0), remember that the maximum likelihood estimate $\hat{\pi}$ is set that way, such that it is true for any $\pi \in \Theta$, that

$$\begin{aligned} & \prod_{t=1}^T \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1; \hat{\pi}) \\ & \geq \prod_{t=1}^T \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1; \pi) \end{aligned}$$

where the (conditional) probabilities (the distribution law of X_1, \dots, X_T) are calculated as if π are the real parameters that generated this realization. Thanks to (16) (and *invertibility*), this becomes

$$\left(\frac{\prod_{t=1}^T \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \hat{\pi})}{\prod_{t=1}^T \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \pi)} \right)^{1/T} \geq 1$$

with probability that tends to one as $T \rightarrow \infty$. Thinking about the random variables (rather than the realizations) now, it can be written

$$\begin{aligned} & \left(\prod_{t=1}^T \frac{p_t(\hat{\pi})}{p_t(\pi)} \right)^{1/T} \\ & = \prod_{x_1, \dots, x_T = 0, v_1, \dots, v_k} \left\{ \left(\frac{\prod_{t=1}^T \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \hat{p})}{\prod_{t=1}^T \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; p)} \right)^{1/T} \right\}^{f_{x_1}(X_1) \dots f_{x_T}(X_T)}, \end{aligned}$$

where X_1, \dots, X_T are the same random variables for both numerator/denominator

with realizations as in $f_{x_1}(X_1) \cdots f_{x_T}(X_T)$ (those have been generated by the real $\boldsymbol{\pi}_0$), and it is verified that

$$\lim_{T \rightarrow \infty} \mathbb{P} \left(\left(\prod_{t=1}^T \frac{p_t(\hat{\boldsymbol{\pi}})}{p_t(\boldsymbol{\pi}_0)} \right)^{1/T} \geq 1 \right) = 1 \text{ for any } \boldsymbol{\pi} \in \Theta. \quad (17)$$

In the other hand, it is true for any $\boldsymbol{\pi} \in \Theta$, that

$$\mathbb{E} \left(\prod_{t=1}^T \frac{p_t(\boldsymbol{\pi})}{p_t(\boldsymbol{\pi}_0)} \right) = \mathbb{E} \left(\mathbb{E} \left(\prod_{t=1}^T \frac{p_t(\boldsymbol{\pi})}{p_t(\boldsymbol{\pi}_0)} \mid X_{1-n}, n \in \mathbb{N}; \boldsymbol{\pi}_0 \right) \right)$$

and it is essential that it is the “real” $\boldsymbol{\pi}_0$ that generates the random variables and governs the (conditional) expectation, *i.e.*,

$$\begin{aligned} & \mathbb{E} \left(\prod_{t=1}^T \frac{p_t(\boldsymbol{\pi})}{p_t(\boldsymbol{\pi}_0)} \mid X_{1-n}, n \in \mathbb{N}; \boldsymbol{\pi}_0 \right) \\ &= \sum_{x_1, \dots, x_T=0, v_1, \dots, v_k} \mathbb{P}(X_1 = x_1, \dots, X_T = x_T \mid X_{1-n}, n \in \mathbb{N}; \boldsymbol{\pi}_0) \\ & \quad \cdot \prod_{t=1}^T \frac{\mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \boldsymbol{\pi})}{\mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_{1-n}, n \in \mathbb{N}; \boldsymbol{\pi}_0)} \\ & \equiv \sum_{x_1, \dots, x_T=0, v_1, \dots, v_k} \mathbb{P}(X_1 = x_1, \dots, X_T = x_T \mid X_{1-n}, n \in \mathbb{N}; \boldsymbol{\pi}) = 1: \end{aligned}$$

the last statement is true since these are probabilities adding to one (regardless of the $\boldsymbol{\pi} \in \Theta$). Then it holds that

$$\mathbb{E} \left(\prod_{t=1}^T \frac{p_t(\boldsymbol{\pi})}{p_t(\boldsymbol{\pi}_0)} \right) = 1 \text{ for any } \boldsymbol{\pi} \in \Theta.$$

As in Jensen’s inequality (use the function $g(y) = y^{1/T}$, $y > 0$ with $g''(y) = (1/T)((1-T)/T)y^{(1/T)-2} < 0$ for $T > 1$), this last statement can be transformed to

$$\mathbb{E} \left\{ \left(\prod_{t=1}^T \frac{p_t(\boldsymbol{\pi})}{p_t(\boldsymbol{\pi}_0)} \right)^{1/T} \right\} \leq 1 \text{ for any } \boldsymbol{\pi} \in \Theta, \quad (18)$$

with the equality holding if and only if $\boldsymbol{\pi} = \boldsymbol{\pi}_0$ (it is considered here that $p_t(\boldsymbol{\pi}_1) = p_t(\boldsymbol{\pi}_2)$, $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \Theta$ can only result from $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_2$).

In order to conclude, one can combine (17) and write

$$\lim_{T \rightarrow \infty} \mathbb{P} \left(\left(\prod_{t=1}^T \frac{p_t(\hat{\boldsymbol{\pi}})}{p_t(\boldsymbol{\pi}_0)} \right)^{1/T} \geq 1 \right) = 1$$

with (18) that writes

$$\mathbb{E} \left\{ \left(\prod_{t=1}^T \frac{p_t(\hat{\boldsymbol{\pi}})}{p_t(\boldsymbol{\pi}_0)} \right)^{1/T} \right\} \leq 1$$

and they are both satisfied when

$$\left(\prod_{t=1}^T p_t(\hat{\boldsymbol{\pi}}) \right)^{1/T} - \left(\prod_{t=1}^T p_t(\boldsymbol{\pi}_0) \right)^{1/T} \xrightarrow{P} 0 \text{ as } T \rightarrow \infty$$

or, in other words, when

$$\hat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}_0 \text{ as } T \rightarrow \infty. \tag{19}$$

4.2. Asymptotic Distribution

Thanks to a Taylor’s expansion, (15) can be turned to

$$\begin{aligned} & \sum_{t=1}^T \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \frac{\partial}{\partial \pi_i} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0} + \sum_{t=1}^T \sum_j \left\{ \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \frac{\partial^2}{\partial \pi_j \partial \pi_i} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0} \right. \\ & \left. - \frac{1}{\tilde{p}_{t,t-1}^2(\boldsymbol{\pi}_0)} \left(\frac{\partial}{\partial \pi_j} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0} \right) \left(\frac{\partial}{\partial \pi_i} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0} \right) \right\} (\hat{\pi}_j - \pi_{j,0}) \\ & + \sum_{t=1}^T E_{t,i}(\hat{\boldsymbol{\pi}}) = 0, \end{aligned}$$

where i (and j) are indexes that refer to all different scalar π in $\boldsymbol{\pi} \in \Theta$ and $E_{t,i}(\hat{\boldsymbol{\pi}})$ are scalars; to understand their role better, consider the expansion for fixed t and define the function

$$\begin{aligned} e_{t,i}(\boldsymbol{\pi}) := & \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi})} \frac{\partial}{\partial \pi_i} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \\ & - \sum_{x=0,1} \frac{\partial^x}{\partial \boldsymbol{\pi}^x} \left(\frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi})} \frac{\partial}{\partial \pi_i} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \right) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0} \frac{(\boldsymbol{\pi} - \boldsymbol{\pi}_0)^x}{x!} \end{aligned}$$

with “ $x=0$ ” applying no derivative at all to the function ($(\boldsymbol{\pi} - \boldsymbol{\pi}_0)^0 \equiv 1$) and “ $x=1$ ” being the usual row vector of first derivatives (times a column vector): then it is clear that $e(\boldsymbol{\pi}_0) = 0$. Due to (19) and the continuity of the function e at $\boldsymbol{\pi}_0$, it can be concluded that $e(\hat{\boldsymbol{\pi}}) \xrightarrow{P} 0$ as $T \rightarrow \infty$ (without worrying about the $(t, t-1)$ label, *i.e.*, the convergence to zero takes place anyway for $T \rightarrow \infty$). It will be taken for granted then that

$$\frac{\sum_{t=1}^T E_{t,i}(\hat{\boldsymbol{\pi}})}{\dots} \xrightarrow{P} 0 \text{ as } T \rightarrow \infty, \tag{20}$$

where “...” is the convenient divisor T or even $T^{1/2}$. Note that (20) has been justified without the presumption of existence of a higher than second derivative of the conditional probabilities under study.

After omitting the extra terms, all the equations are stacked to come up with

$$\sum_{t=1}^T \left(\frac{1}{\tilde{p}_{t,t-1}^2(\boldsymbol{\pi}_0)} \tilde{\delta}_t \tilde{\delta}_t^\tau - \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{D}_t \right) (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) = \sum_{t=1}^T \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{\delta}_t,$$

where $\tilde{\delta}_t$ is the column vector of $\frac{\partial}{\partial \pi_i} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0}$, $i = 1, 2, \dots$ and \tilde{D}_t is

the matrix with elements $\frac{\partial^2}{\partial \pi_i \partial \pi_j} \tilde{p}_{t,t-1}(\boldsymbol{\pi}) \Big|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0}$, $i, j = 1, 2, \dots$ and “ τ ” stands

for the transpose operator. It is re-written that

$$T^{1/2}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) = \left\{ \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{\tilde{p}_{t,t-1}^2(\boldsymbol{\pi}_0)} \tilde{\delta}_t \tilde{\delta}_t^\tau - \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{D}_t \right) \right\}^{-1} T^{-1/2} \sum_{t=1}^T \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{\delta}_t. \quad (21)$$

Next, consider δ_t to be the column vector of $\left. \frac{\partial}{\partial \pi_i} p_t(\boldsymbol{\pi}) \right|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0}$, $i=1,2,\dots$.

Lemma 4.1: It holds that

$$T^{-1/2} \sum_{t=1}^T \left(\frac{1}{p_t(\boldsymbol{\pi}_0)} \delta_t - \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{\delta}_t \right) \xrightarrow{P} \mathbf{0} \text{ as } T \rightarrow \infty.$$

Proof: This may be presented.

Once Lemma 4.1 has been established, Equation (22) should be replaced by

$$T^{1/2}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) = \left\{ \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{\tilde{p}_{t,t-1}^2(\boldsymbol{\pi}_0)} \tilde{\delta}_t \tilde{\delta}_t^\tau - \frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{D}_t \right) \right\}^{-1} T^{-1/2} \sum_{t=1}^T \frac{1}{p_t(\boldsymbol{\pi}_0)} \delta_t. \quad (22)$$

To proceed further, consider D_t to be the matrix with elements

$$\left. \frac{\partial^2}{\partial \pi_i \partial \pi_j} p_t(\boldsymbol{\pi}) \right|_{\boldsymbol{\pi}=\boldsymbol{\pi}_0}, \quad i, j=1, \dots.$$

Lemma 4.2: It holds that

$$\frac{1}{T} \sum_{t=1}^T \left(\left(\frac{1}{\tilde{p}_{t,t-1}(\boldsymbol{\pi}_0)} \tilde{D}_t - \frac{1}{\tilde{p}_{t,t-1}^2(\boldsymbol{\pi}_0)} \tilde{\delta}_t \tilde{\delta}_t^\tau \right) - \left(\frac{1}{p_t(\boldsymbol{\pi}_0)} D_t - \frac{1}{p_t^2(\boldsymbol{\pi}_0)} \delta_t \delta_t^\tau \right) \right) \xrightarrow{P} \mathbf{O}$$

as $T \rightarrow \infty$, where \mathbf{O} (use bold for emphasis) is the square matrix of zeros.

Proof: This may be presented.

Once Lemma 4.2 has been established, Equation (22) should be replaced by

$$T^{1/2}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) = \left\{ \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{p_t^2(\boldsymbol{\pi}_0)} \delta_t \delta_t^\tau - \frac{1}{p_t(\boldsymbol{\pi}_0)} D_t \right) \right\}^{-1} T^{-1/2} \sum_{t=1}^T \frac{1}{p_t(\boldsymbol{\pi}_0)} \delta_t. \quad (23)$$

Theorem 4.3: It holds that

$$T^{-1/2} \sum_{t=1}^T \frac{1}{p_t(\boldsymbol{\pi}_0)} \delta_t \xrightarrow{D} N \left(\mathbf{0}, \text{Var} \left(\frac{1}{p_t(\boldsymbol{\pi}_0)} \delta_t \right) \right) \text{ as } T \rightarrow \infty.$$

Proof: This may be presented.

Lemma 4.4: It holds that

$$\frac{1}{T} \sum_{t=1}^T \frac{1}{p_t^2(\boldsymbol{\pi}_0)} \delta_t \delta_t^\tau \xrightarrow{P} \text{Var} \left(\frac{1}{p_t(\boldsymbol{\pi}_0)} \delta_t \right) \text{ as } T \rightarrow \infty.$$

Proof: This may be presented.

Lemma 4.5: It holds that

$$\frac{1}{T} \sum_{t=1}^T \frac{1}{p_t(\boldsymbol{\pi}_0)} D_t \xrightarrow{P} \mathbf{O} \text{ as } T \rightarrow \infty.$$

Proof: This may be presented.

5. Empirical Illustrations

This section serves in practice, when for special cases of TARMA models the maximum likelihood estimation takes place, so that (i) it is refreshed how to compute joint probabilities to insert in the likelihood (and estimate) and, more importantly, (ii) it is examined how well the estimators perform for moderate sample sizes: for both (i), (ii) particular interest lies in $q \geq 1$, when “naïve” moments estimates cannot be computed directly from the data, as opposed to the more traditional TAR cases. At this point, it is reminded that the TAR is a special case of the TARMA model with AR parameters only and explicit results that have been obtained for the Markov chains. The TARMA with the inclusion of MA parts is a parsimonious way to succeed the infinite order of a TAR model and where the paper interest lies.

Hence the groundwork equation

$$\begin{aligned}
 X_t = & I_t(1 - X_{t-1})(1 - I_{t-1}) + I_t^{(10)}X_{t-1}(1 - I_{t-1}) \\
 & + I_t^{(01)}(1 - X_{t-1})I_{t-1} + I_t^{(11)}X_{t-1}I_{t-1}
 \end{aligned} \tag{24}$$

will, in general, govern a series $\{X_t\} \sim \text{TARMA}(1,1,1)$ (with realizations in $\{0,1\}$). By writing $\psi_{(ij)} := \mathbb{P}(X_t = i, I_t = j)$, $i, j = 0,1$, then from the parameters $\pi = \mathbb{P}(I_t = 1)$ and $\pi_{|v}^{*(ij)} = \mathbb{P}(I_t^{(ij)} = 1 | I_t = v)$, $(i, j) \neq (0,0)$, $v = 0,1$ ($\pi^{(ij)} = \mathbb{P}(I_t^{(ij)} = 1) \equiv \pi\pi_{|1}^{*(ij)} + (1 - \pi)\pi_{|0}^{*(ij)}$), [1] has contributed a methodology that eventually (under causality) computes the ψ as in

$$\begin{aligned}
 \psi_{(11)} = & \left\{ \pi \left[\left(1 - (1 - \pi)\pi_{|0}^{*(10)} \right) \left(1 - \pi \left(1 - \pi_{|1}^{*(01)} \right) \right) - (1 - \pi)\pi_{|0}^{*(01)}\pi \left(1 - \pi_{|1}^{*(10)} \right) \right] \right\} \\
 & / \left\{ \left(1 + \pi - \pi\pi_{|1}^{*(11)} \right) \left[\left(1 - (1 - \pi)\pi_{|0}^{*(10)} \right) \left(1 - \pi \left(1 - \pi_{|1}^{*(01)} \right) \right) - (1 - \pi)\pi_{|0}^{*(01)}\pi \left(1 - \pi_{|1}^{*(10)} \right) \right] \right. \\
 & + \left(\pi - \pi\pi_{|1}^{*(10)} \right) \left[\left(1 - \pi \left(1 - \pi_{|1}^{*(01)} \right) \right) \left(1 - \pi \right) \pi_{|0}^{*(11)} + (1 - \pi)\pi_{|0}^{*(01)}\pi \left(1 - \pi_{|1}^{*(11)} \right) \right] \\
 & \left. + \left(\pi - \pi\pi_{|1}^{*(01)} \right) \left[\pi \left(1 - \pi_{|1}^{*(10)} \right) \left(1 - \pi \right) \pi_{|0}^{*(11)} + \left(1 - (1 - \pi)\pi_{|0}^{*(10)} \right) \pi \left(1 - \pi_{|1}^{*(11)} \right) \right] \right\},
 \end{aligned} \tag{25}$$

followed by

$$\psi_{(10)} = \frac{\left(1 - \pi \left(1 - \pi_{|1}^{*(01)} \right) \right) \left(1 - \pi \right) \pi_{|0}^{*(11)} + (1 - \pi)\pi_{|0}^{*(01)}\pi \left(1 - \pi_{|1}^{*(11)} \right)}{\left(1 - (1 - \pi)\pi_{|0}^{*(10)} \right) \left(1 - \pi \left(1 - \pi_{|1}^{*(01)} \right) \right) - (1 - \pi)\pi_{|0}^{*(01)}\pi \left(1 - \pi_{|1}^{*(10)} \right)} \psi_{(11)} \tag{26}$$

and

$$\begin{aligned}
 \psi_{(01)} = & \frac{\pi \left(1 - \pi_{|1}^{*(10)} \right) \left(1 - \pi \right) \pi_{|0}^{*(11)} + \left(1 - (1 - \pi)\pi_{|0}^{*(10)} \right) \pi \left(1 - \pi_{|1}^{*(11)} \right)}{\left(1 - (1 - \pi)\pi_{|0}^{*(10)} \right) \left(1 - \pi \left(1 - \pi_{|1}^{*(01)} \right) \right) - (1 - \pi)\pi_{|0}^{*(01)}\pi \left(1 - \pi_{|1}^{*(10)} \right)} \psi_{(11)} \\
 & \text{(and } \psi_{(00)} = 1 - \psi_{(10)} - \psi_{(01)} - \psi_{(11)} \text{)}.
 \end{aligned} \tag{27}$$

Additionally, after writing for $h \in \mathbb{N}$ that $\psi_{((i_1, \dots, i_{h+1})|j)}(h) := \mathbb{P}(X_t = i_1, \dots, X_{t-h} = i_{h+1}, I_t = j)$ (and $\psi_{(ij)}(0) \equiv \psi_{(ij)}$), the methodology also contributes the recursive formulae

$$\psi_{((1,0,i_2,\dots,i_{h+1})|1)}(h+1) = \pi \psi_{((0,i_2,\dots,i_{h+1})|0)}(h) + \pi \pi_1^{*(0|1)} \psi_{((0,i_2,\dots,i_{h+1})|1)}(h), \quad (28)$$

$$\psi_{((1,1,i_2,\dots,i_{h+1})|1)}(h+1) = \sum_{j=0,1} \pi \pi_1^{*(1|j)} \psi_{((1,i_2,\dots,i_{h+1})|j)}(h),$$

$$\psi_{((1,0,i_2,\dots,i_{h+1})|0)}(h+1) = (1-\pi) \pi_0^{*(0|1)} \psi_{((0,i_2,\dots,i_{h+1})|1)}(h),$$

$$\psi_{((1,1,i_2,\dots,i_{h+1})|0)}(h+1) = \sum_{j=0,1} (1-\pi) \pi_0^{*(1|j)} \psi_{((1,i_2,\dots,i_{h+1})|j)}(h),$$

and the recursive formulae

$$\psi_{((0,0,i_2,\dots,i_{h+1})|1)}(h+1) = \pi \left(1 - \pi_1^{*(0|1)}\right) \psi_{((0,i_2,\dots,i_{h+1})|1)}(h),$$

$$\psi_{((0,1,i_2,\dots,i_{h+1})|1)}(h+1) = \sum_{j=0,1} \pi \left(1 - \pi_1^{*(1|j)}\right) \psi_{((1,i_2,\dots,i_{h+1})|j)}(h),$$

$$\psi_{((0,0,i_2,\dots,i_{h+1})|0)}(h+1)$$

$$= (1-\pi) \psi_{((0,i_2,\dots,i_{h+1})|0)}(h) + (1-\pi) \left(1 - \pi_0^{*(0|1)}\right) \psi_{((0,i_2,\dots,i_{h+1})|1)}(h),$$

$$\psi_{((0,1,i_2,\dots,i_{h+1})|0)}(h+1) = \sum_{j=0,1} (1-\pi) \left(1 - \pi_0^{*(1|j)}\right) \psi_{((1,i_2,\dots,i_{h+1})|j)}(h), \quad (29)$$

from which the probabilities

$$\mathbb{P}\left(X_t = i_0, X_{t-1} = i_1, \dots, X_{t-(h+1)} = i_{h+1}\right) = \sum_{j=0,1} \psi_{((i_0, i_1, \dots, i_{h+1})|j)}(h+1)$$

can be computed for $i_0, i_1, \dots, i_{h+1} = 0, 1$.

To generate X_t , $t = 1, \dots, T$, it has been used that

$$X_t = Y_t + \sum_{i \in \mathbb{N}} \left(\prod_{j=0}^{i-1} (DY)_{t-j} \right) Y_{t-i}, \quad Y_t := I_t^{(0|1)} I_{t-1} + I_t (1 - I_{t-1}),$$

$$(DY)_t := \left(I_t^{(1|1)} - I_t^{(0|1)} \right) I_{t-1} + \left(I_t^{(1|0)} - I_t \right) (1 - I_{t-1}),$$

with the sum being held up to a finite constant $c = 20$ for $t = 1$, and so on. The two presented conditions for causality and invertibility write together that

$$\pi + \mathbb{E} \left| df_1 \left(I_t^{(0|1)} \right) \right|, \mathbb{E} \left| df_1 \left(I_t^{(1|0)} \right) \right| + \mathbb{E} \left| d^2 f_1 \left(I_t^{(1|1)} \right) \right|,$$

$$\max_{i,j=0,1} \left\{ \pi^{(i|j)} \right\} + \mathbb{E} \left| df_1 \left(I_t^{(1|0)} \right) \right|, \mathbb{E} \left| df_1 \left(I_t^{(0|1)} \right) \right| + \mathbb{E} \left| d^2 f_1 \left(I_t^{(1|1)} \right) \right| < 1$$

(the $d^h f_1(\cdot)$, $h = 1, 2$ notation was re-introduced in the very beginning of Section 3). Hence by imposing the strong requirement $\max_{i,j=0,1} \left\{ \pi^{(i|j)} \right\} < 1/6$, the series is secured to be where it should be (without worrying about the interdependence scenario).

Firstly the case $\pi = 0.165$, $\pi^{(1|0)} = 0.152$, $\pi^{(0|1)} = 0.125$ and $\pi^{(1|1)} = 0.108$ (and under interdependence) has been studied, for a number of observations $T = 6, 10, 14$ (the maximum sample size 14 has been picked due to a purely technical reason of restoring in one vector 2^{T+1} probabilities for the last recursion, or otherwise a modification must take place in the code). A search of the four parameters $0 < \pi^{(\dots)} < 0.1667$, $\pi^{(\dots)} = \pi^{(\dots)} + 0.01$ has materialized and for each attempted value the probabilities (25), (26), (27) and recursions (28)...(29) must be computed: the likelihood is none other than one of these final probabili-

ties indicated by the position of zero-one(s) in the sample series (this is the same position regardless of the parameter values under search). Then the attempted parameter values that offered the biggest of those probabilities are taken to be the maximum likelihood estimates.

Meanwhile, an opportunity of a comparison with the TAR model (of order $p = 2$) with same parameters ($\pi = 0.165$, $\pi^{(10)} = 0.152$, $\pi^{(01)} = 0.125$ and $\pi^{(11)} = 0.108$) has arisen, *i.e.*, $\{X_1, \dots, X_T\}$ have now been born from

$$X_t = I_t(1 - X_{t-1})(1 - X_{t-2}) + I_t^{(10)}X_{t-1}(1 - X_{t-2}) + I_t^{(01)}(1 - X_{t-1})X_{t-2} + I_t^{(11)}X_{t-1}X_{t-2}.$$

Note that it can be derived $X_t = I_t + \sum_{n \in \mathbb{N}} C_{t,n}^{(n-1)} I_{t-n}$ with $C_{t,1}^{(0)} := DI_t^{(10)} \equiv I_t^{(10)} - I_t$, $C_{t,2}^{(0)} := DI_t^{(01)} \equiv I_t^{(01)} - I_t$, $C_{t,(1,2)}^{(0)} := D^2 I_t^{(11)} \equiv I_t^{(11)} - I_t^{(10)} - I_t^{(01)} + I_t$ and for $n \in \mathbb{N}$

$$C_{t,n+1}^{(n)} := C_{t,n+1}^{(n-1)} + C_{t,n}^{(n-1)} DI_{t-n}^{(10)} + C_{t,(n,n+1)}^{(n-1)} I_{t-n}^{(10)}, C_{t,n+2}^{(n)} := C_{t,n}^{(n-1)} DI_{t-n}^{(01)}, \text{ and}$$

$$C_{t,(n+1,n+2)}^{(n)} := C_{t,n}^{(n-1)} D^2 I_{t-n}^{(11)} + C_{t,(n,n+1)}^{(n-1)} D^* I_{t-n}^{(11)}, D^* I_t^{(11)} := I_t^{(11)} - I_t^{(10)}.$$

Computing the probabilities $\mathbb{P}(X_t = i, X_{t-1} = j)$, $i, j = 0, 1$ (from the same methodology and under causality) is straightforward, while the condition for causality now requires $\left[\left(1 + \beta^* \sum_{m=0}^1 \sum_{m_1=0}^m 1 \right)^2 - 1 \right] < 1$ or $\beta^* < \frac{\sqrt{2}-1}{3} \approx 0.138$, where

$$\beta^* := \max \left\{ \pi, \max \left\{ \mathbb{E} \left| DI_t^{(10)} \right|, \mathbb{E} \left| DI_t^{(01)} \right| \right\}, \mathbb{E} \left| D^2 I_t^{(11)} \right| \right\};$$

so it might be that it is violated (the search is no different than for the TARMA above).

According to **Table 1**, it is indeed remarkable how the performances for the two models almost coincide. That is such an encouraging sign for the TARMA, as it tempts the researcher to dare include moving-average parts and benefit from the rare flexibility, if they are willing to take the extra computational burden of the recursions: exact likelihoods *can* be computed as functions of the parameters and, even for small sample sizes, the estimators seem to perform quite satisfactorily. It is highlighted that the precision here is as returned by the code with no formal justification attached to it (for a smaller $R = 100$ repetitions, the differences to the table were still minimal). From **Table 1** it looks like there is a better bias (and MSE) performance of the π estimator over the other three ones, so this is investigated further.

Table 2 picture still favors the π estimator with a low absolute bias: the two $\pi^{(01)}$, $\pi^{(11)}$ estimators score the highest bias, naturally resulting from the real values being quite high. Similarly, for both numerical cases the $\pi^{(10)}$ estimator, with the lowest real value, escapes the high bias values without outperforming the π in the same department though. The conclusions for the variance can be reversed. At this stage, it is highlighted that the easily obtainable (conditional

Table 1. Approximate (from $R = 1000$ replications) bias, variance and Mean Squared Error of the Maximum Likelihood Estimators based on T consecutive observations from the TARMA (1, 1) and TAR (2) models ($k = 1$, Bernoulli variables) with real values $\pi = 0.165$, $\pi^{(10)} = 0.152$, $\pi^{(01)} = 0.125$ and $\pi^{(11)} = 0.108$ (under interdependence).

	TARMA			TAR		
	Bias	Var	MSE	Bias	Var	MSE
$T = 6$						
π	-0.05675	0.00508444	0.008305	-0.0545	0.00497475	0.007945
$\pi^{(10)}$	-0.14	0.0002724	0.0198724	-0.12595	0.0021499	0.0180133
$\pi^{(01)}$	-0.09865	0.00212518	0.011857	-0.1039	0.00154179	0.012337
$\pi^{(11)}$	-0.08375	0.00193444	0.0089485	-0.07445	0.0029779	0.0085207
$T = 10$						
π	-0.04738	0.00267194	0.0049168	-0.03844	0.00260857	0.0040862
$\pi^{(10)}$	-0.13786	0.00057666	0.019582	-0.11545	0.0032776	0.0166063
$\pi^{(01)}$	-0.09121	0.00297854	0.0112978	-0.09595	0.0024946	0.011701
$\pi^{(11)}$	-0.07685	0.00272518	0.0086311	-0.0782	0.00257796	0.0086932
$T = 14$						
π	-0.03761	0.002243061	0.0036575731	-0.04143	0.00233756	0.004054
$\pi^{(10)}$	-0.13783	0.000572841	0.0195699499	-0.10495	0.0041848	0.0151993
$\pi^{(01)}$	-0.08032	0.00394194	0.0103932424	-0.08935	0.00318958	0.011173
$\pi^{(11)}$	-0.0677	0.0035838	0.00816709	-0.0773	0.00267651	0.0086518

Table 2. Approximate (from $R = 10000$ replications) bias, variance and Mean Squared Error of the Maximum Likelihood Estimators based on $T = 6$ consecutive observations from the TARMA (1, 1) model ($k = 1$, Bernoulli variables) with real values, for the “Case 1.1”, $\pi = 0.105$, $\pi^{(10)} = 0.081$, $\pi^{(01)} = 0.148$, $\pi^{(11)} = 0.152$, or for the “Case 1.2”, $\pi = 0.165$, $\pi^{(10)} = 0.081$, $\pi^{(01)} = 0.152$ and $\pi^{(11)} = 0.148$ (both under interdependence).

	Case 1.1			Case 1.2		
	Bias	Var	MSE	Bias	Var	MSE
π	-0.02201	0.00562096	0.0061054	-0.056825	0.00508792	0.008317
$\pi^{(10)}$	-0.069541	0.000208601	0.00504455	-0.068298	0.000374559	0.00503918
$\pi^{(01)}$	-0.130132	0.00109877	0.0180331	-0.125859	0.00211538	0.0179559
$\pi^{(11)}$	-0.1312	0.00150336	0.0187168	-0.123285	0.001990072	0.0171899

likelihood) asymptotic result (for the estimators $\hat{\pi}^{(1)}$, $\hat{\pi}$ of $\pi^{(1)} := \mathbb{P}(X_t = 1 | X_{t-1} = 1)$, $\pi := \mathbb{P}(X_t = 1 | X_{t-1} = 0)$, respectively) regarding the $\{X_t\} \sim \text{TAR}(1)$ model, *i.e.*,

$$T^{1/2} \begin{pmatrix} \hat{\pi}^{(1)} - \pi^{(1)} \\ \hat{\pi} - \pi \end{pmatrix} \xrightarrow{D} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \pi^{(1)}(1 - \pi^{(1)})/\mathbb{E}\{X_t\} & 0 \\ 0 & \pi(1 - \pi)/\mathbb{E}\{1 - X_t\} \end{pmatrix} \right)$$

has also been approved by simulation indications (for sample sizes $T = 10, 20, 30$ and $R = 100$), but those will not be presented here.

The investigation is not over yet, as it is desirable to smooth out the differences in the bias of estimators for the TARMA (1, 1) model, and the solution to that could be an “agreement” between the different $I^{(\dots)}$ variables (at the same time t) via the presence of interdependence. Indeed according to **Table 3** and for the only small sample size ($T = 6$), “Case 2.1” exhibits low biases in all but the -0.11726 approximation from the 1000 estimates of $\pi_1^{*(01)}$; similarly, for the “Case 2.2” and excluding the highest (absolute) bias for π (its true value has been set high), almost all other estimates are approximated to be near their real value. The variance results for both cases are very impressive too. Note that the conditions for causality-invertibility might not have been followed faithfully in **Table 3** (hence the real value $\pi = 0.25$ in “Case 2.2”) and the “agreement” search has been set to ≥ 0.85 for $\pi_1^{*(\dots)}$, or ≤ 0.15 for $\pi_0^{*(\dots)}$ (together with a search $\pi \leq 0.1667$ that justifies the -0.11655 bias in “Case 2.2”).

Again it is reminded that the results obtained from fewer simulations (**Table 2**: $R = 100, 1000$, **Table 3**: $R = 100$) are hardly any different. The simulation indications of this section favor the addition of moving-average parts for the modelling of strict stationarity, if one is willing to slightly inconvenience themselves with a more sophisticated code for the likelihood computation. There

Table 3. Approximate (from $R = 1000$ replications) bias, variance and Mean Squared Error of the Maximum Likelihood Estimators based on $T = 6$ consecutive observations from the TARMA (1, 1) model ($k = 1$, Bernoulli variables) with real values, for the “Case 2.1”, $\pi = 0.165$, $\pi_1^{*(10)} = 0.95$, $\pi_0^{*(10)} = 0.03$, $\pi_1^{*(01)} = 0.98$, $\pi_0^{*(01)} = 0.05$, $\pi_1^{*(11)} = 0.966$, $\pi_0^{*(11)} = 0.044$, or for the “Case 2.2”, $\pi = 0.25$, $\pi_1^{*(10)} = 0.95$, $\pi_0^{*(10)} = 0.03$, $\pi_1^{*(01)} = 0.98$, $\pi_0^{*(01)} = 0.05$ and $\pi_1^{*(11)} = 0.966$, $\pi_0^{*(11)} = 0.044$ (the interdependence is implied for both).

	Case 2.1			Case 2.2		
	Bias	Var	MSE	Bias	Var	MSE
π	-0.0545	0.00497475	0.007945	-0.11655	0.0032776	0.0168615
$\pi_0^{*(10)}$	-0.01832	0.00019877	0.0005344	-0.01316	0.000774014	0.0009472
$\pi_1^{*(10)}$	-0.09804	0.00027055	0.0098824	-0.09202	0.00105352	0.0095212
$\pi_0^{*(01)}$	-0.02908	0.00119115	0.0020368	-0.02032	0.0019743	0.0023872
$\pi_1^{*(01)}$	-0.11726	0.00162129	0.0153712	-0.10704	0.00268724	0.0141448
$\pi_0^{*(11)}$	-0.01852	0.00161797	0.00196096	-0.00196	0.00281824	0.00282208
$\pi_1^{*(11)}$	-0.09794	0.00220224	0.0117945	-0.07862	0.00383594	0.010017

have been no signs of distinction between the auto-regressive and moving-average estimators' merits: not only does this strengthen the views of Sections 4.1 and 4.2, but also it brings to mind the classic Gaussian ARMA gem of [8], that the moving-average estimation transforms in theory to an auto-regression like situation.

6. Conclusions and Extending to the χ^2 Test for Stationarity

Straight from (23), Theorem 4.3 and Lemmas 4.4, 4.5, there is the important derivation

$$T^{1/2}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \xrightarrow{D} N\left(\mathbf{0}, \text{Var}\left(\frac{1}{p_t(\boldsymbol{\pi}_0)}\delta_t\right)^{-1}\right) \text{ as } T \rightarrow \infty \quad (30)$$

for the Maximum Likelihood Estimators of the parameters of a process that assigns its appearance to a causal and invertible TARMA model. The normal distribution convergence (30) implies the chi-square distribution convergence

$$T(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)^T \text{Var}\left(\frac{1}{p_t(\boldsymbol{\pi}_0)}\delta_t\right)(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \xrightarrow{D} \chi_{df}^2, \quad (31)$$

where df equals the size of the parameter vector.

The asymptotic result (31) can be a real asset when testing whether a sample series has been generated by a stationary (\mathbb{Z} -indexed) process: the null hypothesis will be of the form

$$H_0: \{X_t\} \text{ is stationary with "distribution" ...}$$

where "distribution" is being used in the wide sense of a hyper-model that might specify as much information is allowed for the researcher (for example, marginal distribution, pairwise distribution,... or marginal and conditional distributions,... with or without knowledge of parameters etc). Under H_0 , a proper identification of k (this might not be necessary if the variables are discrete with "Bernoulli" being the best of cases), as well as p and q (some faint suggestions for setting $p=0$ or $q=0$ might be found in [1]) must take place to proceed. Identification issues must still be resolved in the case that the researcher is not testing an assumption, but opts for the alternative task of point and interval estimation. For either inference route taken, further research is mandatory, in order to approximate the variance matrix $\text{Var}(p_t^{-1}\delta_t)$. The reader might wish to look for the special case of a causal TAR model, which due to its simplicity, is believed here that it *should be supported as the most appropriate test for the non-parametric time series stationarity*.

In a more structured context, an equivalent to the [14] ARMA methodology of identification, estimation, validation and forecasting can now become reality for the strictly stationary processes: any distribution (heavy-tailed, with extreme values, asymmetric or with levels of skewness) can be tamed by a proper categorization and then p and q need to be selected as well. This paper has contributed the complete guide for the second step with the best of all methods of estimation, by putting on the table all necessary properties of the TARMA maximum

likelihood estimators.

In order to use the TARMA equation, the sample series needs to be produced by a stationary one. Any sign of persistent tendency, such as trends, seasonality or cycles forbid the use of TARMA: joint distributions (of equal “windows”) must look alike in time. For the second-order stationarity [15] give answers on how the series can be fixed to ARMA; or, there is a whole philosophy of modelling those tendencies directly. For the strict stationarity, one can still resort to (31), testing and computing whether the sample series is close enough to a hypothesis of hyper-dependence: a large value of the statistic, as compared to a chi-square value, will suggest that the hypothesis collapses. Otherwise, there are parametric/theoretical approaches that work under specific distributional assumptions.

To wrap it up, it is strongly believed that a precious link has been solidified between the “multiplicative linear” (as called in [1])—strictly stationary- times series and the non-parametric inferential statistics. The well-known χ^2 test for the distribution of a variable that has generated a random sample, can be extended now to the χ^2 test for the stationary-principled hyper-model, dominating a \mathbb{Z} -indexed process that has generated a sample series: as in the former case, when extra (than the prespecified) parameters need to be estimated, it might be worth investigating the reduction of degrees of freedom in the statistic χ^2 distribution. Regardless of various such minor issues that might be studied further, the main achievement of this paper is that it has dealt quite satisfactorily with a complex problem: the inference for any stationary time series that can be clothed by a causal and invertible TARMA equation.

Acknowledgments

This is for the author’s beloved father (deceased).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Dimitriou-Fakalou, C. (2019) The Table Auto-Regressive Moving-Average Model for (Categorical) Stationary Series: Statistical Properties (Causality; from the All Random to the Conditional Random). *Journal of Nonparametric Statistics*, **31**, 31-63. <https://doi.org/10.1080/10485252.2018.1527912>
- [2] Jacobs, P.A. and Lewis, A.W. (1978) Discrete Time Series Generated by Mixtures II: Asymptotic Properties. *Journal of the Royal Statistical Society: Series B*, **40**, 222-228. <https://doi.org/10.1111/j.2517-6161.1978.tb01667.x>
- [3] Möller, T.A. and Wei, C.H. (2020) Generalized Discrete Autoregressive Moving-Average Models. *Applied Stochastic Models in Business and Industry*, **36**, 641-659. <https://doi.org/10.1002/asmb.2520>
- [4] Lomnicki, Z.A. and Zaremba, S.K. (1955) Some Applications of Zero-One Processes.

- Journal of the Royal Statistical Society: Series B*, **17**, 243-255.
<https://doi.org/10.1111/j.2517-6161.1955.tb00198.x>
- [5] Joe, H. (1996) Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class. *Journal of Applied Probability*, **33**, 664-677.
<https://doi.org/10.1017/S0021900200100105>
- [6] Kedem, B. (1980) Estimation of the Parameters in Stationary Autoregressive Processes after Hard Limiting. *Journal of the American Statistical Association*, **75**, 146-153.
<https://doi.org/10.1080/01621459.1980.10477445>
- [7] Azzalini, A. (1983) Maximum Likelihood Estimation of Order m for Stationary Stochastic Processes. *Biometrika*, **70**, 381-387.
<https://doi.org/10.1093/biomet/70.2.381>
- [8] Hannan, E.J. (1973) The Asymptotic Theory of Linear Time Series Models. *Journal of Applied Probability*, **10**, 130-145. <https://doi.org/10.1017/S0021900200042145>
- [9] Cui, Y. and Lund, R.B. (2009) A New Look at Time Series of Counts. *Biometrika*, **96**, 781-792. <https://doi.org/10.1093/biomet/asp057>
- [10] Roitershtein, A. and Zhong, Z. (2013) On Random Coefficient INAR(1) Processes. *Science China Mathematics*, **56**, 177-200.
<https://doi.org/10.1007/s11425-012-4547-z>
- [11] Davis, R.A., Fokianos, K., Holan, S.H., Joe, H., Livsey, J., Lund, R., Pipiras, V. and Ravishanker, N. (2021) Count Time Series: A Methodological Review. *Journal of the American Statistical Association*, **116**, 1533-1547.
<https://doi.org/10.1080/01621459.2021.1904957>
- [12] Anderson, T.W. and Goodman, L.A. (1957) Statistical Inference about Markov Chains. *The Annals of Mathematical Statistics*, **28**, 89-110.
<https://doi.org/10.1214/aoms/1177707039>
- [13] Klotz, J. (1973) Statistical Inference in Bernoulli Trials with Dependence. *The Annals of Statistics*, **1**, 373-379. <https://doi.org/10.1214/aos/1176342377>
- [14] Box, G.E.P. and Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
- [15] Brockwell, P.J. and Davis, R.A. (1991) Time Series: Theory and Methods. 2nd Edition, Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4419-0320-4>