
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Voutilainen, Marko; Viitasaari, Lauri; Ilmonen, Pauliina
On model fitting and estimation of strictly stationary processes

Published in:
Modern Stochastics: Theory and Applications

DOI:
[10.15559/17-VMSTA91](https://doi.org/10.15559/17-VMSTA91)

Published: 01/01/2017

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Voutilainen, M., Viitasaari, L., & Ilmonen, P. (2017). On model fitting and estimation of strictly stationary processes. *Modern Stochastics: Theory and Applications*, 4(4), 381-406. <https://doi.org/10.15559/17-VMSTA91>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

On model fitting and estimation of strictly stationary processes

Marko Voutilainen^{a,*}, Lauri Viitasaari^b, Pauliina Ilmonen^a

^a*Department of Mathematics and Systems Analysis
Aalto University School of Science
P.O. Box 11100, FI-00076 Aalto, Finland*

^b*Department of Mathematics and Statistics
University of Helsinki
P.O. Box 68, FI-00014 University of Helsinki, Finland*

marko.voutilainen@aalto.fi (M. Voutilainen), lauri.viitasaari@iki.fi (L. Viitasaari),
pauliina.ilmonen@aalto.fi (P. Ilmonen)

Received: 13 September 2017, Revised: 22 November 2017, Accepted: 25 November 2017,
Published online: 22 December 2017

Abstract Stationary processes have been extensively studied in the literature. Their applications include modeling and forecasting numerous real life phenomena such as natural disasters, sales and market movements. When stationary processes are considered, modeling is traditionally based on fitting an autoregressive moving average (ARMA) process. However, we challenge this conventional approach. Instead of fitting an ARMA model, we apply an AR(1) characterization in modeling any strictly stationary processes. Moreover, we derive consistent and asymptotically normal estimators of the corresponding model parameter.

Keywords AR(1) representation, asymptotic normality, consistency, estimation, strictly stationary processes

2010 MSC 60G10, 62M09, 62M10, 60G18

1 Introduction

Stochastic processes are widely used in modeling and forecasting numerous real life phenomena such as natural disasters, activity of the sun, sales of a company and mar-

*Corresponding author.

ket movements, to mention a few. When stationary processes are considered, modeling is traditionally based on fitting an autoregressive moving average (ARMA) process. However, in this paper, we challenge this conventional approach. Instead of fitting an ARMA model, we apply the AR(1) characterization in modeling any strictly stationary processes. Moreover, we derive consistent and asymptotically normal estimators of the corresponding model parameter.

One of the reasons why ARMA processes have been in a central role in modeling of time-series data is that for every autocovariance function $\gamma(\cdot)$ vanishing at infinity and for every $n \in \mathbb{N}$ there exists an ARMA process X such that $\gamma(k) = \gamma_X(k)$ for every $k = 0, 1, \dots, n$. For a general overview of the theory of stationary ARMA processes and their estimation, the reader may consult for example [1] or [5].

ARMA processes, and their extensions, have been studied extensively in the literature. A direct proof of consistency and asymptotic normality of Gaussian maximum likelihood estimators for causal and invertible ARMA processes was given in [18]. The result was originally obtained, using asymptotic properties of the Whittle estimator, in [7]. The estimation of the parameters of strictly stationary ARMA processes with infinite variances was studied in [16], again, by using Whittle estimators. Portmanteau tests for ARMA models with stable Paretian errors with infinite variance were introduced in [12]. An efficient method for evaluating the maximum likelihood function of stationary vector ARMA models was presented in [14]. Fractionally integrated ARMA models with a GARCH noise process, where the variance of the error terms is also of ARMA form, was studied in [13]. Consistency and asymptotic normality of the quasi-maximum likelihood estimators of ARMA models with the noise process driven by a GARCH model was shown in [3]. A least squares approach for ARMA parameter estimation has been studied at least in [9] by contrasting its efficiency with the maximum likelihood estimation. Also estimators of autocovariance and their limiting behavior have been addressed in numerous papers. See for example [2, 8, 11] and [15].

Modeling an observed time-series with an ARMA process starts by fixing the orders of the model. This is often done by an educated guess, but there also exists methods for estimating the orders, see e.g. [6]. After the orders are fixed, the related parameters can be estimated, for example, by using the maximum likelihood or least squares estimators. These estimators are expressed in terms of optimization problems and do not generally admit closed form representations. The final step is to conduct various diagnostic tests to determine whether the estimated model is sufficiently good or not. These tests are often designed to recognize whether the residuals of the model support the underlying assumptions about the error terms. Depending on whether one considers strict or weak stationarity, the error process is usually assumed to be an IID process or white noise, respectively. If the tests do not support the assumptions about the noise process, then one has to start all over again. Tests for the goodness of fit of ARMA models have been suggested e.g. in [4].

The approach taken in this paper is based on the discrete version of the main theorem of [17] leading to an AR(1) characterization for (any) strictly stationary processes. Note that this approach covers, but is not limited to, strictly stationary ARMA processes. It was stated in [17] that a process is strictly stationary if and only if for every fixed $0 < H < 1$ it can be represented in the AR(1) form with $\phi = e^{-H}$

and a unique, possibly correlated, noise term. Although the representation is unique only after H is fixed, we show that in most of the cases, given just one value of the autocovariance function of the noise, one is able to determine the AR(1) parameter and, consequently, the entire autocovariance function of the noise process. It is worth emphasizing that since the parameter–noise pair in the AR(1) characterization is not unique, it is natural that some information about the noise has to be assumed. Note that conventionally, when applying ARMA models, we have assumptions about the noise process much stronger than being IID or white noise. That is, the autocovariance function of the noise is assumed to be identically zero except at the origin. When founding estimation on the AR(1) characterization, one does not have to select between different complicated models. In addition, there is only one parameter left to be estimated. Yet another advantage over classical ARMA estimation is that we obtain closed form expressions for the estimators.

The paper is organized as follows. We begin Section 2 by introducing some terminology and notation. After that, we give a characterization of discrete time strictly stationary processes as AR(1) processes with a possibly correlated noise term together with some illustrative examples. The AR(1) characterization leads to Yule–Walker type equations for the AR(1) parameter ϕ . In this case, due to the correlated noise process, the equations are of quadratic form in ϕ . For the rest of the section, we study the quadratic equations and determine ϕ with as little information about the noise process as possible. The approach taken in Section 2 leads to an estimator of the AR(1) parameter. We consider estimation in detail in Section 3. The end of Section 3 is dedicated to testing the assumptions we make when constructing the estimators. A simulation study to assess finite sample properties of the estimators is presented in Section 4. Finally, we end the paper with three appendices containing a technical proof, detailed discussion on some special cases and tabulated simulation results.

2 On AR(1) characterization in modeling strictly stationary processes

Throughout the paper we consider strictly stationary processes.

Definition 1. Assume that $X = (X_t)_{t \in \mathbb{Z}}$ is a stochastic process. If

$$(X_{t+n_1}, X_{t+n_2}, \dots, X_{t+n_k}) \stackrel{law}{=} (X_{n_1}, X_{n_2}, \dots, X_{n_k})$$

for all $k \in \mathbb{N}$ and $t, n_1, n_2, \dots, n_k \in \mathbb{Z}$, then X is strictly stationary.

Definition 2. Assume that $G = (G_t)_{t \in \mathbb{Z}}$ is a stochastic process and denote $\Delta_t G = G_t - G_{t-1}$. If $(\Delta_t G)_{t \in \mathbb{Z}}$ is strictly stationary, then the process G is a strictly stationary increment process.

The following class of stochastic processes was originally introduced in [17].

Definition 3. Let $H > 0$ be fixed and let $G = (G_t)_{t \in \mathbb{Z}}$ be a stochastic process. If G is a strictly stationary increment process with $G_0 = 0$ and if the limit

$$\lim_{k \rightarrow -\infty} \sum_{t=k}^0 e^{tH} \Delta_t G \tag{1}$$

exists in probability and defines an almost surely finite random variable, then G belongs to the class of converging strictly stationary increment processes, and we denote $G \in \mathcal{G}_H$.

Next, we consider the AR(1) characterization of strictly stationary processes. The continuous time analogy was proved in [17] together with a sketch of a proof for the discrete case. For the reader’s convenience, a detailed proof of the discrete case is presented in Appendix A.

Theorem 1. *Let $H > 0$ be fixed and let $X = (X_t)_{t \in \mathbb{Z}}$ be a stochastic process. Then X is strictly stationary if and only if $\lim_{t \rightarrow -\infty} e^{tH} X_t = 0$ in probability and*

$$\Delta_t X = (e^{-H} - 1)X_{t-1} + \Delta_t G \tag{2}$$

for a unique $G \in \mathcal{G}_H$.

Corollary 1. *Let $H > 0$ be fixed. Then every discrete time strictly stationary process $(X_t)_{t \in \mathbb{Z}}$ can be represented as*

$$X_t - \phi^{(H)} X_{t-1} = Z_t^{(H)}, \tag{3}$$

where $\phi^{(H)} = e^{-H}$ and $Z_t^{(H)} = \Delta_t G$ is another strictly stationary process.

It is worth to note that the noise Z in Corollary 1 is unique only after the parameter H is fixed. The message of this result is that every strictly stationary process is an AR(1) process with a strictly stationary noise that may have a non-zero auto-covariance function. The following examples show how some conventional ARMA processes can be represented as an AR(1) process.

Example 1. *Let X be a strictly stationary AR(1) process defined by*

$$X_t - \varphi X_{t-1} = \epsilon_t, \quad (\epsilon_t) \sim IID(0, \sigma^2)$$

with $\varphi > 0$. Then we may simply choose $\phi^{(H)} = \varphi$ and $Z_t^{(H)} = \epsilon_t$.

Example 2. *Let X be a strictly stationary ARMA(1, q) process defined by*

$$X_t - \varphi X_{t-1} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad (\epsilon_t) \sim IID(0, \sigma^2)$$

with $\varphi > 0$. Then we may set $\phi^{(H)} = \varphi$, and $Z_t^{(H)}$ then equals to the MA(q) process.

Example 3. *Consider a strictly stationary AR(1) process X with $\varphi < 0$. Then X admits an MA(∞) representation*

$$X_t = \sum_{k=0}^{\infty} \varphi^k \epsilon_{t-k}.$$

From this it follows that

$$Z_t^{(H)} = \epsilon_t + \sum_{k=0}^{\infty} \varphi^k (\varphi - \phi^{(H)}) \epsilon_{t-1-k}$$

and

$$\text{cov}(Z_t^{(H)}, Z_0^{(H)}) = \phi^{t-2}(\phi - \phi^{(H)})\sigma^2 \left(\phi + (\phi - \phi^{(H)}) \sum_{n=1}^{\infty} (\phi^2)^n \right).$$

Hence in the case of an AR(1) process with a negative parameter, the autocovariance function of the noise Z of the representation (3) is non-zero everywhere.

Next we show how to determine the AR(1) parameter $\phi^{(H)}$ in (3) provided that the observed process X is known. In what follows, we omit the superindices in (3). We assume that the second moments of the considered processes are finite and that the processes are centered. That is, $\mathbb{E}(X_t) = \mathbb{E}(Z_t) = 0$ for every $t \in \mathbb{Z}$. Throughout the rest of the paper, we use the notation $\text{cov}(X_t, X_{t+n}) = \gamma(n)$ and $\text{cov}(Z_t, Z_{t+n}) = r(n)$ for every $t, n \in \mathbb{Z}$.

Lemma 1. *Let centered $(X_t)_{t \in \mathbb{Z}}$ be of the form (3). Then*

$$\phi^2 \gamma(n) - \phi(\gamma(n + 1) + \gamma(n - 1)) + \gamma(n) - r(n) = 0 \tag{4}$$

for every $n \in \mathbb{Z}$.

Proof. Let $n \in \mathbb{Z}$. By multiplying both sides of

$$X_n - \phi X_{n-1} = Z_n$$

with $Z_0 = X_0 - \phi X_{-1}$ and taking expectations, we obtain

$$\begin{aligned} & \mathbb{E}(X_n(X_0 - \phi X_{-1})) - \phi \mathbb{E}(X_{n-1}(X_0 - \phi X_{-1})) \\ & = \phi^2 \gamma(n) - \phi(\gamma(n + 1) + \gamma(n - 1)) + \gamma(n) = r(n). \end{aligned} \quad \square$$

Corollary 2. *Let centered $(X_t)_{t \in \mathbb{Z}}$ be of the form (3) and let $N \in \mathbb{N}$ be fixed.*

(1) *If $\gamma(N) \neq 0$, then either*

$$\phi = \frac{\gamma(N + 1) + \gamma(N - 1) + \sqrt{(\gamma(N + 1) + \gamma(N - 1))^2 - 4\gamma(N)(\gamma(N) - r(N))}}{2\gamma(N)} \tag{5}$$

or

$$\phi = \frac{\gamma(N + 1) + \gamma(N - 1) - \sqrt{(\gamma(N + 1) + \gamma(N - 1))^2 - 4\gamma(N)(\gamma(N) - r(N))}}{2\gamma(N)}. \tag{6}$$

(2) *If $\gamma(N) = 0$ and $r(N) \neq 0$, then*

$$\phi = -\frac{r(N)}{\gamma(N + 1) + \gamma(N - 1)}.$$

Note that if $\gamma(N) = r(N) = 0$, then Lemma 1 yields only $\gamma(N + 1) + \gamma(N - 1) = 0$ providing no information about the parameter ϕ . As such, in order to determine the parameter ϕ , we require that either $\gamma(N) \neq 0$ or $r(N) \neq 0$.

Remark 1. *If the variance $r(0)$ of the noise is known, then (5) and (6) reduces to*

$$\phi = \frac{\gamma(1) \pm \sqrt{\gamma(1)^2 - \gamma(0)(\gamma(0) - r(0))}}{\gamma(0)}.$$

At first glimpse it seems that Corollary 2 is not directly applicable. Indeed, in principle it seems like there could be complex-valued solutions although representation (3) together with (4) implies that there exists a solution $\phi \in (0, 1)$. Furthermore, it is not clear whether the true value is given by (5) or (6). We next address these issues. We start by proving that the solutions to (4) cannot be complex. At the same time we are able to determine which one of the solutions one should choose.

Lemma 2. *The discriminants of (5) and (6) are always non-negative.*

Proof. Let $k \in \mathbb{Z}$. By multiplying both sides of (3) with X_{t-k} , taking expectations, and applying (3) repeatedly we obtain

$$\begin{aligned} \gamma(k) - \phi\gamma(k - 1) &= \mathbb{E}(Z_t X_{t-k}) = \mathbb{E}(Z_t (Z_{t-k} + \phi X_{t-k-1})) \\ &= r(k) + \phi \mathbb{E}(Z_t X_{t-k-1}) \\ &= r(k) + \phi \mathbb{E}(Z_t (Z_{t-k-1} + \phi X_{t-k-2})) \\ &= r(k) + \phi r(k + 1) + \phi^2 \mathbb{E}(Z_t X_{t-k-2}). \end{aligned}$$

Proceeding as above l times we get

$$\gamma(k) - \phi\gamma(k - 1) = \sum_{i=0}^{l-1} \phi^i r(k + i) + \phi^l \mathbb{E}(Z_t (\phi X_{t-k-l-2})).$$

Letting l approach infinity leads to

$$\gamma(k) - \phi\gamma(k - 1) = \sum_{i=0}^{\infty} \phi^i r(k + i), \tag{7}$$

where the series converges as $r(k + i) \leq r(0)$ and $0 < \phi < 1$. It now follows from (7) that

$$\begin{aligned} \gamma(N) &= \phi\gamma(N - 1) + \sum_{i=0}^{\infty} \phi^i r(N + i) \\ &= \phi\gamma(N - 1) + r(N) + \phi \sum_{i=1}^{\infty} \phi^{i-1} r(N + i) \\ &= \phi\gamma(N - 1) + r(N) + \phi \sum_{i=0}^{\infty} \phi^i r(N + i + 1) \\ &= \phi\gamma(N - 1) + \phi(\gamma(N + 1) - \phi\gamma(N)) + r(N). \end{aligned}$$

Denote the discriminant of (5) and (6) by D . That is,

$$D = (\gamma(N - 1) + \gamma(N + 1))^2 - 4\gamma(N)(\gamma(N) - r(N)).$$

By using the equation above we observe that

$$D = \left(\frac{\gamma(N) + \phi^2\gamma(N) - r(N)}{\phi} \right)^2 - 4\gamma(N)(\gamma(N) - r(N)).$$

Denoting $a_N = \frac{r(N)}{\gamma(N)}$, multiplying by $\frac{\phi^2}{\gamma(N)^2}$, and using the identity

$$(a + b)^2 - 4ab = (a - b)^2$$

yields

$$\frac{\phi^2}{\gamma(N)^2} D = (1 + \phi^2 - a_N)^2 - 4\phi^2(1 - a_N) = (\phi^2 - 1 + a_N)^2 \geq 0.$$

This concludes the proof. □

Note that if $r(N) = 0$, as $\phi < 1$, the discriminant is always positive. Let $a_N = \frac{r(N)}{\gamma(N)}$. The proof above now gives us the following identity

$$\phi = \frac{1}{2\phi} \left(1 + \phi^2 - a_N \pm \frac{|\gamma(N)|}{\gamma(N)} |\phi^2 - 1 + a_N| \right).$$

This enables us to consider the choice between (5) and (6). Assume that $\gamma(N) > 0$. If $\phi^2 - 1 + a_N > 0$, then ϕ is given by (5) (as $\phi \in (0, 1)$). Similarly, if $\phi^2 - 1 + a_N < 0$, then ϕ is determined by (6). Finally, contrary conclusions hold in the case $\gamma(N) < 0$. In particular, we can always choose between (5) and (6) provided that either $a_N \leq 0$ or $a_N \geq 1$. Moreover, from (4) it follows that

$$\frac{r(N)}{\gamma(N)} = \frac{r(N+k)}{\gamma(N+k)}$$

if and only if

$$\frac{\gamma(N+1) + \gamma(N-1)}{\gamma(N)} = \frac{\gamma(N+1+k) + \gamma(N-1+k)}{\gamma(N+k)},$$

provided that the denominators differ from zero. Since (5) and (6) can be written as

$$\begin{aligned} \phi &= \frac{\gamma(N+1) + \gamma(N-1)}{2\gamma(N)} \\ &\pm \frac{1}{2} \operatorname{sgn}(\gamma(N)) \sqrt{\left(\frac{\gamma(N+1) + \gamma(N-1)}{\gamma(N)} \right)^2 - 4 \left(1 - \frac{r(N)}{\gamma(N)} \right)}, \end{aligned} \quad (8)$$

we observe that one can always rule out one of the solutions (5) and (6) provided that $a_N \neq a_{N+k}$. Therefore, it always suffices to know two values of the autocovariance r such that $a_N \neq a_{N+k}$, except the worst case scenario where $a_j = a \in (0, 1)$ for every $j \in \mathbb{Z}$. A detailed analysis of this particular case is given in Appendix B.

Remark 2. Consider a fixed strictly stationary process X . If we fix one value of the autocovariance function of the noise such that Corollary 2 yields an unambiguous AR(1) parameter, then the quadratic equations (4) will unravel the entire autocovariance function of the noise process. In comparison, conventionally, the noise is assumed to be white — meaning that the entire autocovariance function of the noise is assumed to be known a priori.

We end this section by observing that in the case of vanishing autocovariance function of the noise, we get the following simplified form for the AR(1) parameter.

Theorem 2. Let centered $(X_t)_{t \in \mathbb{Z}}$ be of the form (3) and let $N \in \mathbb{N}$ be fixed. Assume that $r(m) = 0$ for every $m \geq N$. If $\gamma(N - 1) \neq 0$, then for every $n \geq N$, we have

$$\phi = \frac{\gamma(n)}{\gamma(n - 1)}.$$

In particular, γ admits an exponential decay for $n \geq N$.

Proof. Let $\gamma(N - 1) \neq 0$. It follows directly from (7) and the assumptions that

$$\gamma(n) = \phi\gamma(n - 1) \quad \text{for every } n \geq N.$$

The condition $\gamma(N - 1) \neq 0$ now implies the claim. \square

Recall that the representation (2) is unique only after H is fixed. As a simple corollary for Theorem 2 we obtain the following result giving some new information about the uniqueness of the representation (2).

Corollary 3. Let X be a strictly stationary process with a non-vanishing autocovariance. Then there exists at most one pair (H, G) satisfying (2) such that the non-zero part of the autocovariance function of the increment process $(\Delta_t G)_{t \in \mathbb{Z}}$ is finite.

Proof. Assume that there exists $H_1, H_2 > 0$, and $G_1 \in \mathcal{G}_{H_1}$ and $G_2 \in \mathcal{G}_{H_2}$ such that the pairs (H_1, G_1) and (H_2, G_2) satisfy (2) and the autocovariances of $(\Delta_t G_1)_{t \in \mathbb{Z}}$ and $(\Delta_t G_2)_{t \in \mathbb{Z}}$ have cut-off points. From Theorem 2 it follows that $H_1 = H_2$ and since for a fixed H the process G in (2) is unique, we get $G_1 = G_2$. \square

3 Estimation

Corollary 2 gives natural estimators for ϕ provided that we have been able to choose between (5) and (6), and that a value of $r(n)$ is known. We emphasize that in our model it is sufficient to know only one (or in some cases two) of the values $r(n)$, whereas in conventional ARMA modeling much stronger assumptions are required. (In fact, in conventional ARMA modeling the noise process is assumed to be white noise.) It is also worth to mention that, generally, estimators of the parameters of stationary processes are not expressible in a closed form. For example, this is the case with the maximum likelihood and least squares estimators of conventionally modeled ARMA processes, see [1]. Within our method, the model fitting is simpler. Finally, it is worth to note that assumption of one known value of $r(n)$ is a natural one and cannot be avoided. Indeed, this is a direct consequence of the fact that the pair

(ϕ, Z) in representation (3) is not unique. In fact, for practitioner, it is not absolutely necessary to know any values of $r(n)$. The practitioner may make an educated guess and proceed in estimation. If the obtained estimate then turns out to be feasible, the practitioner can stop there. If the obtained estimate turns out to be unreasonable (not on the interval $(0, 1)$), then the practitioner have to make another educated guess. The process is similar to selecting p and q in traditional $ARMA(p, q)$ modeling.

Throughout this section, we assume that (X_1, \dots, X_T) is an observed series from a centered strictly stationary process that is modeled using the representation (3). We use $\hat{\gamma}_T(n)$ to denote an estimator of the corresponding autocovariance $\gamma(n)$. For example, $\hat{\gamma}_T(n)$ can be given by

$$\hat{\gamma}_T(n) = \frac{1}{T} \sum_{t=1}^{T-n} X_t X_{t+n},$$

or more generally

$$\hat{\gamma}_T(n) = \frac{1}{T} \sum_{t=1}^{T-n} (X_t - \bar{X})(X_{t+n} - \bar{X}),$$

where \bar{X} is the sample mean of the observations. For this estimator the corresponding sample covariance (function) matrix is positive semidefinite. On the other hand, the estimator is biased while it is asymptotically unbiased. Another option is to use $T - n - 1$ as a denominator. In this case one has an unbiased estimator, but the sample covariance (function) matrix is no longer positive definite. Obviously, both estimators have the same asymptotic properties. Furthermore, for our purposes it is irrelevant how the estimators $\hat{\gamma}_T(n)$ are defined, as long as they are consistent, and the asymptotic distribution is known.

We next consider estimators of the parameter ϕ arising from characterization (3). In this context, we pose some assumptions related to the autocovariance function of the observed process X . The justification and testing of these assumptions are discussed in Section 3.1. From *a priori* knowledge that $\phi \in (0, 1)$ we enforce also the estimators to the corresponding closed interval. However, if one prefers to use unbounded versions of the estimators, one may very well do that. The asymptotic properties are the same in both cases. We begin by defining an estimator corresponding to the second part (2) of Corollary 2.

Definition 4. Assume that $\gamma(N) = 0$. Then we define

$$\hat{\phi}_T = -\frac{r(N)}{\hat{\gamma}_T(N+1) + \hat{\gamma}_T(N-1)} \mathbb{1}_{\hat{\gamma}_T(N+1) + \hat{\gamma}_T(N-1) \neq 0} \tag{9}$$

whenever the right-hand side lies on the interval $[0, 1]$. If the right-hand side is below zero, we set $\hat{\phi}_T = 0$ and if the right-hand side is above one, we set $\hat{\phi}_T = 1$.

Theorem 3. Assume that $\gamma(N) = 0$ and $r(N) \neq 0$. If the vector-valued estimator $[\hat{\gamma}_T(N+1), \hat{\gamma}_T(N-1)]^\top$ is consistent, then $\hat{\phi}_T$ is consistent.

Proof. Since $\gamma(N) = 0$ and $r(N) \neq 0$, Equation (4) guarantees that $\gamma(N+1) + \gamma(N-1) \neq 0$. Therefore consistency of $\hat{\phi}_T$ follows directly from the continuous mapping theorem. □

Theorem 4. Let $\hat{\phi}_T$ be given by (4), and assume that $\gamma(N) = 0$ and $r(N) \neq 0$. Set $\boldsymbol{\gamma} = [\gamma(N + 1), \gamma(N - 1)]^\top$ and $\hat{\boldsymbol{\gamma}}_T = [\hat{\gamma}_T(N + 1), \hat{\gamma}_T(N - 1)]^\top$. If

$$l(T)(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \xrightarrow{\text{law}} \mathcal{N}(\mathbf{0}, \Sigma)$$

for some covariance matrix Σ and some rate function $l(T)$, then

$$l(T)(\hat{\phi}_T - \phi) \xrightarrow{\text{law}} \mathcal{N}(\mathbf{0}, \nabla f(\boldsymbol{\gamma})^\top \Sigma \nabla f(\boldsymbol{\gamma})),$$

where $\nabla f(\boldsymbol{\gamma})$ is given by

$$\nabla f(\boldsymbol{\gamma}) = -\frac{r(N)}{(\gamma(N + 1) + \gamma(N - 1))^2} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{10}$$

Proof. For the simplicity of notation, in the proof we use the unbounded version of the estimator $\hat{\phi}_T$. Since the true value of ϕ lies strictly between 0 and 1, the very same result holds also for the bounded estimator of Definition 2. Indeed, this is a simple consequence of the Slutsky’s theorem. To begin with, let us define an auxiliary function f by

$$f(\mathbf{x}) = f(x_1, x_2) = \frac{r(N)}{x_1 + x_2} \mathbb{1}_{x_1 + x_2 \neq 0}.$$

If $x_1 + x_2 \neq 0$, the function f is smooth in a neighborhood of \mathbf{x} . Since $\gamma(N) = 0$ together with $r(N) \neq 0$ implies that $\gamma(N + 1) + \gamma(N - 1) \neq 0$, we may apply the delta method at $\mathbf{x} = \boldsymbol{\gamma}$ to obtain

$$l(T)(\hat{\phi}_T - \phi) = -l(T)(f(\hat{\boldsymbol{\gamma}}_T) - f(\boldsymbol{\gamma})) \xrightarrow{\text{law}} \mathcal{N}(\mathbf{0}, \nabla f(\boldsymbol{\gamma})^\top \Sigma \nabla f(\boldsymbol{\gamma})),$$

where $\nabla f(\boldsymbol{\gamma})$ is given by (10). This concludes the proof. □

Remark 3. By writing

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

the variance of the limiting random variable reads

$$\frac{r(N)^2}{(\gamma(N + 1) + \gamma(N - 1))^4} (\sigma_X^2 + 2\sigma_{XY} + \sigma_Y^2).$$

Remark 4. In many cases the convergency rate is the best possible, that is $l(T) = \sqrt{T}$. However, our results are valid with any rate function. One might, for example in the case of many long memory processes, have other convergency rates for the estimators $\hat{\gamma}_T(n)$.

We continue by defining an estimator corresponding to the first part (1) of the Corollary 2. For this we assume that, for reasons discussed in Section 2, we have chosen the solution (5) (cf. Remark 5 and Section 3.1). As above, we show that consistency and asymptotic normality follow from the same properties of the autocovariance estimators. In the sequel we use a short notation

$$g(\mathbf{x}) = g(x_1, x_2, x_3) = (x_1 + x_3)^2 - 4x_2(x_2 - r(N)). \tag{11}$$

In addition, we denote

$$\boldsymbol{\gamma} = [\gamma(N + 1), \gamma(N), \gamma(N - 1)]^\top$$

and

$$\hat{\boldsymbol{\gamma}}_T = [\hat{\gamma}_T(N + 1), \hat{\gamma}_T(N), \hat{\gamma}_T(N - 1)]^\top.$$

Definition 5. Assume that $\gamma(N) \neq 0$. We define an estimator for ϕ associated to (5) by

$$\hat{\phi}_T = \frac{\hat{\gamma}_T(N + 1) + \hat{\gamma}_T(N - 1) + \sqrt{g(\hat{\boldsymbol{\gamma}}_T)} \mathbb{1}_{g(\hat{\boldsymbol{\gamma}}_T) > 0}}{2\hat{\gamma}_T(N)} \mathbb{1}_{\hat{\gamma}_T(N) \neq 0} \quad (12)$$

whenever the right-hand side lies on the interval $[0, 1]$. If the right-hand side is below zero, we set $\hat{\phi}_T = 0$ and if the right-hand side is above one, we set $\hat{\phi}_T = 1$.

Theorem 5. Assume that $\gamma(N) \neq 0$ and $g(\boldsymbol{\gamma}) > 0$. Furthermore, assume that ϕ is given by (5). If $\hat{\boldsymbol{\gamma}}_T$ is consistent, then $\hat{\phi}_T$ is consistent.

Proof. As $g(\boldsymbol{\gamma}) > 0$, the result is again a simple consequence of the continuous mapping theorem. \square

Before proving the asymptotic normality, we present some short notation. We set

$$C_N = \frac{\gamma(N + 1) + \gamma(N - 1) + \sqrt{g(\boldsymbol{\gamma})}}{\gamma(N)} \quad (13)$$

and

$$\begin{aligned} \Sigma_\phi = \frac{1}{4\gamma(N)^2} & \left((\nabla\sqrt{g(\boldsymbol{\gamma})})^\top \Sigma \nabla\sqrt{g(\boldsymbol{\gamma})} + 2 \begin{bmatrix} 1 \\ -C_N \\ 1 \end{bmatrix}^\top \Sigma \nabla\sqrt{g(\boldsymbol{\gamma})} \right. \\ & \left. + \begin{bmatrix} 1 \\ -C_N \\ 1 \end{bmatrix}^\top \Sigma \begin{bmatrix} 1 \\ -C_N \\ 1 \end{bmatrix} \right), \end{aligned} \quad (14)$$

where

$$\nabla\sqrt{g(\boldsymbol{\gamma})} = \frac{1}{\sqrt{g(\boldsymbol{\gamma})}} \begin{bmatrix} \gamma(N + 1) + \gamma(N - 1) \\ 2(r(N) - 2\gamma(N)) \\ \gamma(N + 1) + \gamma(N - 1) \end{bmatrix}.$$

Theorem 6. Let the assumptions of Theorem 5 prevail. If

$$l(T)(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \xrightarrow{law} \mathcal{N}(\mathbf{0}, \Sigma)$$

for some covariance matrix Σ and some rate function $l(T)$, then $l(T)(\hat{\phi}_T - \phi)$ is asymptotically normal with zero mean and variance given by (14).

Proof. The proof follows the same lines as the proof of Theorem 4 but for the reader's convenience, we present the details. Furthermore, as in the proof of Theorem 4, since the true value of ϕ lies strictly between 0 and 1, for the notational simplicity, we may and will use the unbounded version of the estimator. Indeed, the asymptotics for the bounded version then follow directly from the Slutsky's theorem. We have

$$\begin{aligned} & \left(\frac{\hat{\gamma}_T(N+1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N+1)}{\hat{\gamma}_T(N)} - \frac{\gamma(N+1)}{\gamma(N)} \right) \\ &= \frac{1}{\hat{\gamma}_T(N)} (\hat{\gamma}_T(N+1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N+1)) + \left(\frac{\gamma(N+1)}{\hat{\gamma}_T(N)} - \frac{\gamma(N+1)}{\gamma(N)} \right) \\ &= \frac{1}{\hat{\gamma}_T(N)} \left(\hat{\gamma}_T(N+1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N+1) - \frac{\gamma(N+1)}{\gamma(N)} (\hat{\gamma}_T(N) - \gamma(N)) \right). \end{aligned}$$

Similarly

$$\begin{aligned} & \left(\frac{\hat{\gamma}_T(N-1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N-1)}{\hat{\gamma}_T(N)} - \frac{\gamma(N-1)}{\gamma(N)} \right) \\ &= \frac{1}{\hat{\gamma}_T(N)} \left(\hat{\gamma}_T(N-1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N-1) - \frac{\gamma(N-1)}{\gamma(N)} (\hat{\gamma}_T(N) - \gamma(N)) \right) \end{aligned}$$

and

$$\begin{aligned} & \left(\frac{\sqrt{g(\hat{\boldsymbol{\gamma}}_T)}\mathbb{1}_{g(\hat{\boldsymbol{\gamma}}_T)>0}\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \sqrt{g(\boldsymbol{\gamma})}}{\hat{\gamma}_T(N)} - \frac{\sqrt{g(\boldsymbol{\gamma})}}{\gamma(N)} \right) \\ &= \frac{1}{\hat{\gamma}_T(N)} \left(\sqrt{g(\hat{\boldsymbol{\gamma}}_T)}\mathbb{1}_{g(\hat{\boldsymbol{\gamma}}_T)>0}\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \sqrt{g(\boldsymbol{\gamma})} - \frac{\sqrt{g(\boldsymbol{\gamma})}}{\gamma(N)} (\hat{\gamma}_T(N) - \gamma(N)) \right). \end{aligned}$$

For C_N given in (13) we have

$$\begin{aligned} l(T)(\hat{\phi}_T - \phi) &= \frac{l(T)}{2\hat{\gamma}_T(N)} (\hat{\gamma}_T(N+1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N+1) \\ &\quad + \hat{\gamma}_T(N-1)\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \gamma(N-1) - C_N(\hat{\gamma}_T(N) - \gamma(N)) \\ &\quad + \sqrt{g(\hat{\boldsymbol{\gamma}}_T)}\mathbb{1}_{g(\hat{\boldsymbol{\gamma}}_T)>0}\mathbb{1}_{\hat{\gamma}_T(N)\neq 0} - \sqrt{g(\boldsymbol{\gamma})}). \end{aligned}$$

By defining

$$h(\mathbf{x}) = h(x_1, x_2, x_3) = (x_1 + x_3 + \sqrt{g(\mathbf{x})}\mathbb{1}_{g(\mathbf{x})>0})\mathbb{1}_{x_2\neq 0} - C_N x_2$$

we have

$$l(T)(\hat{\phi}_T - \phi) = \frac{l(T)}{2\hat{\gamma}_T(N)} (h(\hat{\boldsymbol{\gamma}}_T) - h(\boldsymbol{\gamma})). \tag{15}$$

If $x_2 \neq 0$ and $g(\mathbf{x}) > 0$, the function h is smooth in a neighborhood of \mathbf{x} . Therefore we may apply the delta method at $\mathbf{x} = \boldsymbol{\gamma}$ to obtain

$$l(T)(h(\hat{\boldsymbol{\gamma}}_T) - h(\boldsymbol{\gamma})) \xrightarrow{\text{law}} \mathcal{N}(\mathbf{0}, \nabla h(\boldsymbol{\gamma})^\top \Sigma \nabla h(\boldsymbol{\gamma})),$$

where

$$\nabla h(\boldsymbol{\gamma})^\top \Sigma \nabla h(\boldsymbol{\gamma}) = \left(\left[\begin{array}{c} 1 \\ -C_N \\ 1 \end{array} \right] + \nabla \sqrt{g(\boldsymbol{\gamma})} \right)^\top \Sigma \left(\left[\begin{array}{c} 1 \\ -C_N \\ 1 \end{array} \right] + \nabla \sqrt{g(\boldsymbol{\gamma})} \right)$$

$$\begin{aligned}
 &= (\nabla \sqrt{g(\boldsymbol{\gamma})})^\top \Sigma \nabla \sqrt{g(\boldsymbol{\gamma})} + 2 \begin{bmatrix} 1 \\ -C_N \\ 1 \end{bmatrix}^\top \Sigma \nabla \sqrt{g(\boldsymbol{\gamma})} \\
 &\quad + \begin{bmatrix} 1 \\ -C_N \\ 1 \end{bmatrix}^\top \Sigma \begin{bmatrix} 1 \\ -C_N \\ 1 \end{bmatrix}.
 \end{aligned}$$

Hence (15) and Slutsky’s theorem imply that $l(T)(\hat{\phi}_T - \phi)$ is asymptotically normal with zero mean and variance given by (14). \square

Remark 5. One straightforwardly observes the same limiting behavior as in Theorems 5 and 6 for the estimator related to (6). This fact also can be used to determine which one of Equations (5) and (6) gives the correct ϕ (cf. Section 3.1).

Remark 6. If $\gamma(N) \neq 0$ and $g(\boldsymbol{\gamma}) = 0$ we may define an estimator

$$\hat{\phi}_T = \frac{\hat{\gamma}_T(N+1) + \hat{\gamma}_T(N-1)}{2\hat{\gamma}_T(N)} \mathbb{1}_{\hat{\gamma}_T(N) \neq 0}.$$

Assuming that

$$l(T)(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \xrightarrow{law} \mathcal{N}(\mathbf{0}, \Sigma)$$

it can be shown similarly as in the proofs of Theorems 4 and 6 that

$$l(T)(\hat{\phi}_T - \phi) \xrightarrow{law} \mathcal{N}\left(0, \frac{1}{4\gamma(N)^2} \begin{bmatrix} 1 \\ -\frac{\gamma(N+1)+\gamma(N-1)}{\gamma(N)} \\ 1 \end{bmatrix}^\top \Sigma \begin{bmatrix} 1 \\ -\frac{\gamma(N+1)+\gamma(N-1)}{\gamma(N)} \\ 1 \end{bmatrix}\right)$$

Remark 7. The estimator related to Theorem 2 reads

$$\hat{\phi}_T = \frac{\hat{\gamma}_T(n+1)}{\hat{\gamma}_T(n)} \mathbb{1}_{\hat{\gamma}_T(n) \neq 0},$$

where we assume that $\gamma(n) \neq 0$. By using the same techniques as earlier, it can be shown that if

$$l(T)(\hat{\gamma}_T(n+1) - \gamma(n+1), \hat{\gamma}_T(n) - \gamma(n)) \xrightarrow{law} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}\right),$$

then

$$l(T)(\hat{\phi}_T - \phi) \xrightarrow{law} \mathcal{N}\left(0, \frac{\sigma_X^2}{\gamma(n)^2} + \frac{\gamma(n+1)^2}{\gamma(n)^4} \sigma_Y^2 - 2\frac{\gamma(n+1)}{\gamma(n)^3} \sigma_{XY}\right).$$

Note that the asymptotics given in Remarks 6 and 7 hold also if one forces the corresponding estimators to the interval $[0, 1]$ as we did in Definitions 4 and 5.

3.1 Testing the underlying assumptions

When choosing the estimator that corresponds the situation at hand, we have to make assumptions related to the values of $\gamma(N)$ (for some N) and $g(\boldsymbol{\gamma})$. In addition, we have to consider the question of the choice between (5) and (6).

Let us first discuss how to test the null hypothesis that $\gamma(N) = 0$. If the null hypothesis holds, then by asymptotic normality of the autocovariances, we have that

$$l(T)\hat{\gamma}_T(N) \xrightarrow{\text{law}} \mathcal{N}(0, \sigma^2) \tag{16}$$

with some σ^2 . Hence we may use

$$\hat{\gamma}_T(N) \sim_a \mathcal{N}\left(0, \frac{\sigma^2}{l(T)^2}\right)$$

as a test statistics. A similar approach can be applied also when testing the null hypothesis that $g(\boldsymbol{\gamma}) = 0$, where g is defined by (11). The alternative hypothesis is of the form $g(\boldsymbol{\gamma}) > 0$. Assuming that the null hypothesis holds, we obtain by the delta method that

$$l(T)(g(\hat{\boldsymbol{\gamma}}_T) - g(\boldsymbol{\gamma})) \xrightarrow{\text{law}} \mathcal{N}(0, \tilde{\sigma}^2)$$

for some $\tilde{\sigma}^2$ justifying the use of

$$g(\hat{\boldsymbol{\gamma}}_T) \sim_a \mathcal{N}\left(0, \frac{\tilde{\sigma}^2}{l(T)^2}\right)$$

as a test statistics. If the tests above suggest that $\gamma(N) \neq 0$ and $g(\boldsymbol{\gamma}) > 0$, then the choice of the sign can be based on the discussion in Section 2. Namely, if for the ratio $a_N = \frac{r(N)}{\gamma(N)}$ it holds that $a_N \leq 0$ or $a_N \geq 1$, then the sign is unambiguous. The sign of $\gamma(N)$ can be deduced from the previous testing of the null hypothesis $\gamma(N) = 0$. By (16), if necessary, one can test the null hypothesis $\gamma(N) = r(N)$ using the test statistics

$$\hat{\gamma}_T(N) \sim_a \mathcal{N}\left(r(N), \frac{\sigma^2}{l(T)^2}\right),$$

where the alternative hypothesis is of the form $\frac{r(N)}{\gamma(N)} < 1$. Finally, assume that one wants to test if the null hypothesis $a_N = a_k$ holds. By the delta method we obtain that

$$l(T)(\hat{a}_N - \hat{a}_k - a_N + a_k) \xrightarrow{\text{law}} \mathcal{N}(0, \bar{\sigma}^2)$$

for some $\bar{\sigma}^2$ suggesting that

$$\hat{a}_N - \hat{a}_k \sim_a \mathcal{N}\left(0, \frac{\bar{\sigma}^2}{l(T)^2}\right)$$

could be utilized as a test statistics.

4 Simulations

We present a simulation study to assess the finite sample performance of the estimators. In the simulations, we apply the estimator corresponding to the first part (1) of Corollary 2. We simulate data from AR(1) processes and ARMA(1, 2) processes with $\theta_1 = 0.8$ and $\theta_2 = 0.3$ as the MA parameters. (Note that these processes correspond

to Examples 1 and 2.) We assess the effects of the sample size T , AR(1) parameter φ , and the chosen lag N . We consider the sample sizes $T = 50, 500, 5000, 50000$, lags $N = 1, 2, 3, \dots, 10$, and the true parameter values $\varphi = 0.1, 0.2, 0.3, \dots, 0.9$. For each combination, we simulate 1000 draws. The sample means of the obtained estimates are tabulated in Appendix C.

Histograms given in Figures 1, 2 and 3 reflect the effects of the sample size T , AR(1) parameter φ , and the chosen lag N , respectively. In Figure 1, the parameter $\varphi = 0.5$ and the lag $N = 3$. In Figure 2, the sample size $T = 5000$ and the lag $N = 3$. In Figure 3, the parameter $\varphi = 0.5$ and the sample size $T = 5000$. The summary statistics corresponding to the data displayed in the histograms are given in Appendix C.

Figure 1 exemplifies the rate of convergence of the estimator as the number of observations grows. One can see that with the smallest sample size, the lower bound is hit numerous times due to the large variance of the estimator. In the upper series of the histograms, the standard deviation reduces from 0.326 to 0.019, whereas in the lower series it reduces from 0.250 to 0.008. The faster convergence in the case of ARMA(1, 2) can be explained with the larger value of $\gamma(3)$ reducing the variance in comparison to the AR(1) case. The same phenomenon recurs also in the other two figures.

Figure 2 reflects the effect of the AR(1) parameter on the value of $\gamma(3)$ and consequently on the variance of the estimator. The standard deviation reduces from 0.322 to 0.020 in the case of AR(1) and from 0.067 to 0.009 in the case of ARMA(1, 2).

In Figure 3 one can see how an increase in the lag increases the variance of the estimator. In the topmost sequence, the standard deviation increases from 0.014 to 0.326 and in the bottom sequence from 0.015 to 0.282.

We wish to emphasize that in general smaller lag does not imply smaller variance, since the autocovariance function of the observed process is not necessarily decreasing. In addition, although the autocovariance $\gamma(N)$ appears to be the dominant factor when it comes to the speed of convergence, there are also other possibly significant terms involved in the limit distribution of Theorem 6.

A Proof of Theorem 1

We provide here a detailed proof of Theorem 1. The continuous time version of the theorem was recently proved in [17] and we loosely follow the same lines in our proof for the discrete time version.

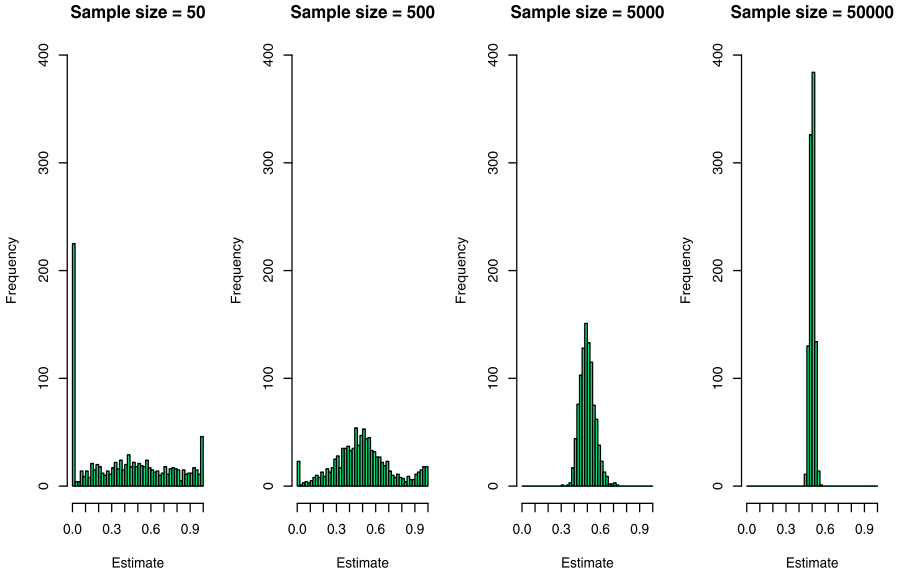
Definition 6. Let $H > 0$. A discrete time stochastic process $Y = (Y_{e^t})_{t \in \mathbb{Z}}$ with $\lim_{t \rightarrow -\infty} Y_{e^t} = 0$ is H -self-similar if

$$(Y_{e^{t+s}})_{t \in \mathbb{Z}} \stackrel{\text{law}}{=} (e^{sH} Y_{e^t})_{t \in \mathbb{Z}}$$

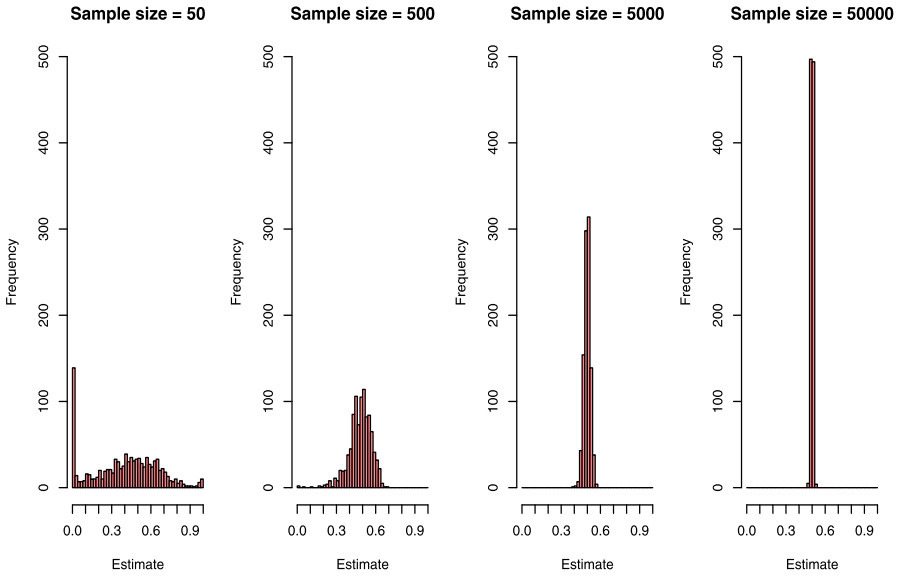
for every $s \in \mathbb{Z}$ in the sense of finite-dimensional distributions.

Definition 7. Let $H > 0$. In addition, let $X = (X_t)_{t \in \mathbb{Z}}$ and $Y = (Y_{e^t})_{t \in \mathbb{Z}}$ be stochastic processes. We define the discrete Lamperti transform by

$$(\mathcal{L}_H X)_{e^t} = e^{tH} X_t$$

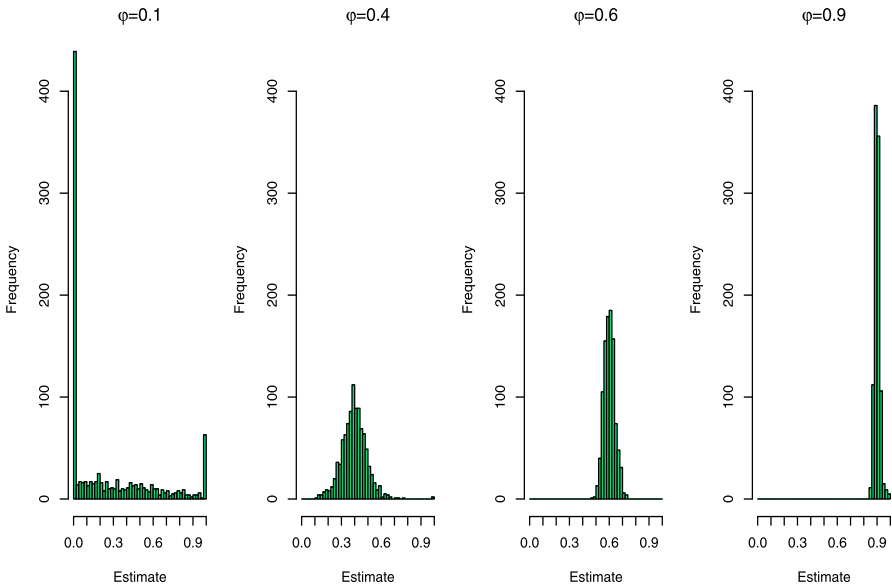


(a) AR(1) process.

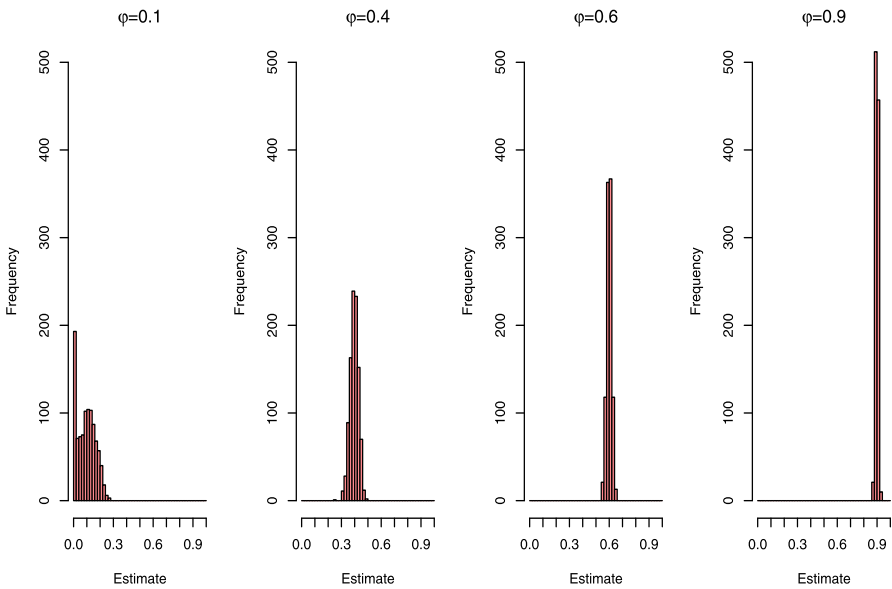


(b) ARMA(1,2) process.

Fig. 1. The effect of the sample size T on the estimates $\hat{\varphi} = \hat{\phi}$. The true parameter value $\varphi = 0.5$ and the lag $N = 3$. The number of iterations is 1000

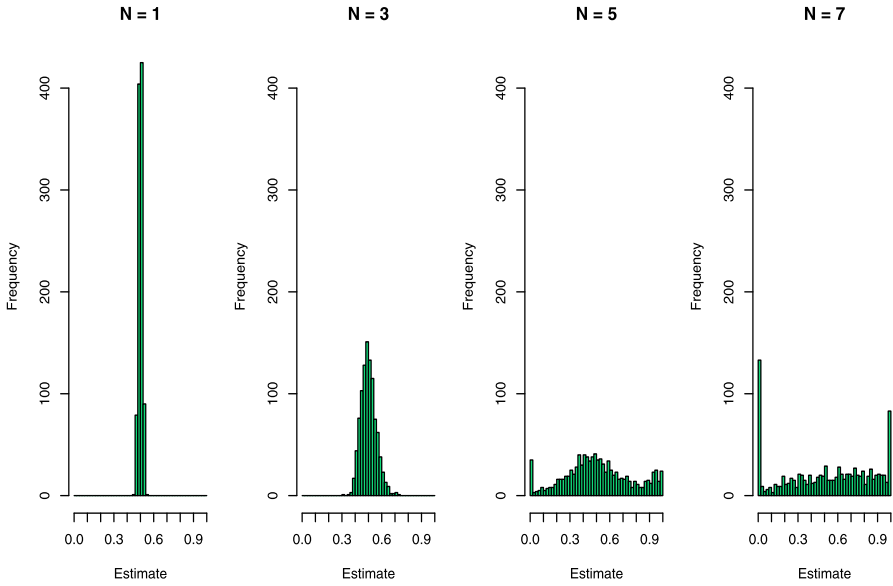


(a) AR(1) process.

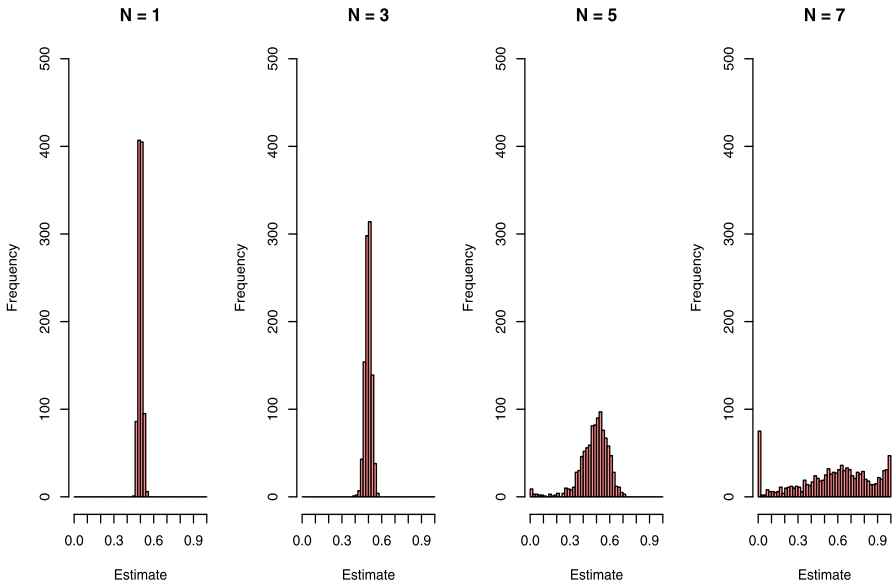


(b) ARMA(1,2) process.

Fig. 2. The effect of the true parameter value φ on the estimates $\hat{\varphi} = \hat{\phi}$. The sample size $T = 5000$ and the lag $N = 3$. The number of iterations is 1000



(a) AR(1) process.



(b) ARMA(1,2) process.

Fig. 3. The effect of the lag N on the estimates $\hat{\varphi} = \hat{\phi}$. The sample size $T = 5000$ and the true parameter value $\varphi = 0.5$. The number of iterations is 1000

and its inverse by

$$(\mathcal{L}_H^{-1}Y)_t = e^{-tH}Y_{e^t}.$$

Theorem 7 (Lamperti [10]). *If $X = (X_t)_{t \in \mathbb{Z}}$ is strictly stationary, then $(\mathcal{L}_H X)_{e^t}$ is H -self-similar. Conversely, if $Y = (Y_{e^t})_{t \in \mathbb{Z}}$ is H -self-similar, then $(\mathcal{L}_H^{-1}Y)_t$ is strictly stationary.*

Lemma 3. *Let $H > 0$ and assume that $(Y_{e^t})_{t \in \mathbb{Z}}$ is H -self-similar. Let us denote $\Delta_t Y_{e^t} = Y_{e^t} - Y_{e^{t-1}}$. Then the process $(G_t)_{t \in \mathbb{Z}}$ defined by*

$$G_t = \begin{cases} \sum_{k=1}^t e^{-kH} \Delta_k Y_{e^k}, & t \geq 1 \\ 0, & t = 0 \\ -\sum_{k=t+1}^0 e^{-kH} \Delta_k Y_{e^k}, & t \leq -1 \end{cases} \tag{17}$$

belongs to \mathcal{G}_H .

Proof. By studying the cases $t \geq 2, t = 1, t = 0$ and $t \leq -1$ separately, it is straightforward to see that

$$\Delta_t G = e^{-tH} \Delta_t Y_{e^t} \quad \text{for every } t \in \mathbb{Z}. \tag{18}$$

Now

$$\lim_{k \rightarrow -\infty} \sum_{t=k}^0 e^{tH} \Delta_t G = \lim_{k \rightarrow -\infty} \sum_{t=k}^0 \Delta_t Y_{e^t} = Y_{e^0} - \lim_{k \rightarrow -\infty} Y_{e^k}$$

and since Y is self-similar, we have

$$Y_{e^k} \stackrel{\text{law}}{=} e^{kH} Y_{e^0}.$$

Thus

$$\lim_{k \rightarrow -\infty} Y_{e^k} = 0$$

in distribution, and hence also in probability. This implies that

$$\sum_{t=-\infty}^0 e^{tH} \Delta_t G$$

is an almost surely finite random variable. Next we show that G has strictly stationary increments. For this, assume that $t, s, l \in \mathbb{Z}$ with $t > s$ are arbitrary. Then

$$\begin{aligned} G_t - G_s &= \sum_{k=s+1}^t \Delta_k G = \sum_{k=s+1}^t e^{-kH} \Delta_k Y_{e^k} = \sum_{j=s+l+1}^{t+l} e^{-(j-l)H} \Delta_{j-l} Y_{e^{j-l}} \\ &\stackrel{\text{law}}{=} \sum_{j=s+l+1}^{t+l} e^{-jH} \Delta_j Y_{e^j} = G_{t+l} - G_{s+l}, \end{aligned}$$

where the equality in law follows from H -self-similarity of $(Y_{e^t})_{t \in \mathbb{Z}}$. Treating n -dimensional vectors similarly concludes the proof. □

Proof of Theorem 1. Assume first that X is strictly stationary. In this case X clearly satisfies the limit condition. In addition, there exists a H -self-similar Y such that

$$\begin{aligned} \Delta_t X &= e^{-tH} Y_{e^t} - e^{-(t-1)H} Y_{e^{t-1}} \\ &= (e^{-H} - 1)e^{-(t-1)H} Y_{e^{t-1}} + e^{-tH} (Y_{e^t} - Y_{e^{t-1}}) \\ &= (e^{-H} - 1)X_{t-1} + e^{-tH} \Delta_t Y_{e^t}. \end{aligned}$$

Defining the process G as in Lemma 3 completes the proof of the ‘if’ part. For the proof of the ‘only if’ part, assume that $G \in \mathcal{G}_H$. From (2) it follows that

$$\begin{aligned} X_t &= e^{-H} X_{t-1} + \Delta_t G = e^{-2H} X_{t-2} + e^{-H} \Delta_{t-1} G + \Delta_t G \\ &= \sum_{j=0}^n e^{-jH} \Delta_{t-j} G + e^{-(n+1)H} X_{t-n-1} \\ &= e^{-tH} \left(\sum_{k=t-n}^t e^{kH} \Delta_k G + e^{(t-n-1)H} X_{t-n-1} \right) \end{aligned}$$

for every $n \in \mathbb{N}$. Since $G \in \mathcal{G}_H$ and $\lim_{m \rightarrow -\infty} e^{mH} X_m = 0$ in probability, we obtain that

$$X_t = e^{-tH} \sum_{k=-\infty}^t e^{kH} \Delta_k G$$

for every $t \in \mathbb{Z}$. Now, by strictly stationary increments of G , we have

$$e^{-tH} \sum_{j=-M}^t e^{jH} \Delta_{j+s} G \stackrel{\text{law}}{=} e^{-tH} \sum_{j=-M}^t e^{jH} \Delta_j G.$$

for every $t, M \in \mathbb{Z}$ such that $-M \leq t$. Since the sums above converge as M tends to infinity, we obtain

$$X_{t+s} = e^{-(t+s)H} \sum_{j=-\infty}^t e^{(j+s)H} \Delta_{j+s} G \stackrel{\text{law}}{=} e^{-tH} \sum_{j=-\infty}^t e^{jH} \Delta_j G = X_t.$$

Treating multidimensional distributions similarly we thus observe that X is strictly stationary. Finally, to prove the uniqueness assume there exist $G_1, G_2 \in \mathcal{G}_H$ such that

$$e^{tH} X_t = \sum_{k=-\infty}^t e^{kH} \Delta_k G_1 = \sum_{k=-\infty}^t e^{kH} \Delta_k G_2$$

for every $t \in \mathbb{Z}$. Then

$$e^{tH} X_t - e^{(t-1)H} X_{t-1} = e^{tH} \Delta_t G_1 = e^{tH} \Delta_t G_2.$$

Hence $\Delta_t G_1 = \Delta_t G_2$ for every $t \in \mathbb{Z}$ implying that $G_1 = G_2 + c$. Since both processes are zero at $t = 0$, it must hold that $c = 0$. \square

Remark 8. Corollary 1 is almost trivial. However, it is well motivated by Theorem 1. On the other hand, Theorem 1 is far away from trivial as it states both sufficient

and necessary conditions. We prove Theorem 1 using discrete Lamperti transform. In principle, one could consider proving Theorem 1 by starting from Corollary 1. However, at this point, we have not assumed any moment conditions, and thus it is not clear whether a process G constructed from $Z^{(H)}$ of Corollary 1 would satisfy $G \in \mathcal{G}_H$. Indeed, a counter example is provided in [17, Proposition 2.1.]. See also [17, Theorem 2.2.], where moment conditions are discussed.

B Discussion on special cases

In this appendix we take a closer look at “worst case scenario” processes related to the choice between (5) and (6). These are such processes that, for some $0 < a < 1$, $a_j = a$ for every $j \in \mathbb{Z}$. By (4) this is equivalent to

$$\frac{\gamma(j + 1) + \gamma(j - 1)}{\gamma(j)} = b \tag{19}$$

for every $j \in \mathbb{Z}$, where $\phi < b < \phi + \frac{1}{\phi}$. In order to study processes of this form, we consider formal power series.

Definition 8. *Let*

$$f(x) = \sum_{n=0}^{\infty} c_n x^n$$

be a formal power series in x . We now define the coefficient extractor operator $[\cdot]\{\}$ by*

$$[x^m]\{f(x)\} = c_m$$

Setting $j = 0$ in (19) we obtain that $\gamma(1) = \frac{b}{2}\gamma(0)$. This leads to the following recursion.

$$\gamma(n) = b\gamma(n - 1) - \gamma(n - 2) \quad \text{for } n \geq 2. \tag{20}$$

It follows immediately from the first step of the recursion that $b > 2$ does not define an autocovariance function of a stationary process. Note also that for $b = 2$ Equation (20) implies that $\gamma(n) = \gamma(0)$ for every $n \in \mathbb{Z}$. This corresponds to the completely degenerate process $X_n = X_0$. We next study the case $0 < b < 2$. For this, we define a generating function regarded as a formal power series by

$$f(x) = \sum_{n=0}^{\infty} \gamma(n)x^n. \tag{21}$$

Then the coefficients of $f(x)$ satisfy

$$\begin{aligned} [x^n]\{f(x)\} &= b[x^{n-1}]\{f(x)\} - [x^{n-2}]\{f(x)\} \\ &= [x^n]\{bx f(x)\} - [x^n]\{x^2 f(x)\} \\ &= [x^n]\{bx f(x) - x^2 f(x)\} \end{aligned}$$

for $n \geq 2$. For simplicity, we assume that $\gamma(0) = 1$. By taking the constant and the first order terms into account we obtain

$$f(x) = bx f(x) - x^2 f(x) - bx + 1 + \frac{b}{2}x,$$

which implies

$$f(x) = \frac{1 - \frac{b}{2}x}{x^2 - bx + 1}.$$

Since the function above is analytic at $x = 0$, the corresponding power series expansion is (21). Furthermore, since the recursion formula is linear, for a general $\gamma(0)$ it holds that

$$\gamma(n) = \gamma(0)[x^n] \left\{ \left(1 - \frac{b}{2}x\right) \sum_{n=0}^{\infty} (bx - x^2)^n \right\}.$$

C Tables

The simulation results highlighted in Section 4 are chosen from a more extensive set of simulations. All the simulation results are given in a tabulated form in this appendix. The two processes considered in the simulations are AR(1) and ARMA(1, 2). The used MA parameters are $\theta_1 = 0.8$ and $\theta_2 = 0.3$. The tables represent the efficiency dependence of the estimator on the AR(1) parameter φ and the used lag N . We have varied the column variable AR(1) parameter from 0.1 to 0.9 and the row variable lag from 1 to 10. The tables display the sample means of the estimates from 1000 iterations with different sample sizes. At the end of this appendix, we provide summary statistics tables corresponding to the histograms presented in Section 4.

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.10	0.18	0.27	0.38	0.48	0.57	0.67	0.77	0.85
2	0.25	0.26	0.30	0.35	0.45	0.54	0.64	0.74	0.82
3	0.32	0.35	0.35	0.40	0.41	0.48	0.57	0.69	0.80
4	0.30	0.37	0.42	0.47	0.50	0.52	0.55	0.66	0.77
5	0.33	0.39	0.42	0.50	0.53	0.57	0.61	0.65	0.75
6	0.34	0.37	0.42	0.47	0.56	0.60	0.66	0.69	0.74
7	0.32	0.37	0.43	0.49	0.57	0.60	0.68	0.69	0.75
8	0.32	0.34	0.45	0.51	0.57	0.64	0.69	0.72	0.76
9	0.31	0.37	0.44	0.50	0.59	0.64	0.70	0.73	0.78
10	0.34	0.35	0.43	0.51	0.58	0.64	0.70	0.75	0.78

Table 1. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for AR(1) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The sample size is 50 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
2	0.23	0.24	0.30	0.40	0.51	0.60	0.70	0.80	0.91
3	0.29	0.31	0.34	0.40	0.50	0.61	0.71	0.81	0.91
4	0.32	0.37	0.40	0.40	0.49	0.61	0.70	0.81	0.90
5	0.30	0.37	0.42	0.48	0.50	0.58	0.70	0.81	0.90
6	0.30	0.36	0.44	0.47	0.53	0.58	0.68	0.80	0.90
7	0.30	0.37	0.44	0.49	0.53	0.57	0.65	0.79	0.90
8	0.32	0.39	0.44	0.51	0.57	0.61	0.68	0.76	0.90
9	0.30	0.38	0.45	0.51	0.59	0.63	0.68	0.75	0.89
10	0.32	0.39	0.46	0.52	0.58	0.64	0.70	0.75	0.89

Table 2. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for AR(1) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The sample size is 500 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
2	0.13	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
3	0.26	0.27	0.32	0.40	0.50	0.60	0.70	0.80	0.90
4	0.30	0.32	0.34	0.42	0.51	0.61	0.70	0.80	0.90
5	0.29	0.37	0.38	0.43	0.51	0.62	0.71	0.80	0.90
6	0.31	0.37	0.41	0.45	0.49	0.62	0.71	0.81	0.90
7	0.29	0.38	0.40	0.47	0.52	0.59	0.72	0.81	0.90
8	0.29	0.40	0.45	0.51	0.54	0.58	0.71	0.81	0.91
9	0.32	0.37	0.41	0.50	0.54	0.60	0.68	0.82	0.91
10	0.29	0.37	0.41	0.51	0.57	0.61	0.68	0.82	0.91

Table 3. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for AR(1) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The sample size is 5000 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
2	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
3	0.21	0.21	0.30	0.40	0.50	0.60	0.70	0.80	0.90
4	0.28	0.30	0.33	0.40	0.50	0.60	0.70	0.80	0.90
5	0.29	0.34	0.36	0.41	0.51	0.60	0.70	0.80	0.90
6	0.29	0.37	0.39	0.42	0.52	0.60	0.70	0.80	0.90
7	0.29	0.37	0.44	0.45	0.51	0.61	0.70	0.80	0.90
8	0.31	0.37	0.43	0.48	0.49	0.62	0.70	0.80	0.90
9	0.31	0.35	0.43	0.49	0.53	0.60	0.71	0.80	0.90
10	0.32	0.37	0.42	0.48	0.53	0.58	0.72	0.80	0.90

Table 4. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for AR(1) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The sample size is 50000 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.08	0.14	0.22	0.32	0.41	0.52	0.61	0.72	0.81
2	0.09	0.13	0.20	0.30	0.39	0.50	0.60	0.72	0.82
3	0.32	0.33	0.32	0.34	0.40	0.46	0.58	0.71	0.81
4	0.65	0.66	0.62	0.60	0.60	0.56	0.60	0.68	0.78
5	0.64	0.67	0.69	0.70	0.69	0.69	0.69	0.70	0.77
6	0.64	0.68	0.69	0.72	0.74	0.75	0.76	0.75	0.78
7	0.64	0.67	0.70	0.72	0.77	0.78	0.79	0.79	0.80
8	0.65	0.67	0.71	0.72	0.76	0.79	0.81	0.80	0.83
9	0.63	0.68	0.72	0.74	0.78	0.80	0.82	0.83	0.84
10	0.65	0.68	0.70	0.74	0.78	0.80	0.83	0.85	0.85

Table 5. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for ARMA(1, 2) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the sample size is 50 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.09	0.19	0.29	0.39	0.49	0.59	0.69	0.80	0.90
2	0.09	0.19	0.28	0.39	0.49	0.59	0.69	0.79	0.90
3	0.12	0.18	0.26	0.37	0.48	0.59	0.69	0.79	0.90
4	0.58	0.49	0.38	0.37	0.45	0.57	0.69	0.79	0.90
5	0.64	0.65	0.62	0.57	0.52	0.56	0.68	0.79	0.90
6	0.65	0.68	0.67	0.68	0.66	0.61	0.67	0.79	0.90
7	0.66	0.68	0.69	0.71	0.72	0.69	0.69	0.78	0.90
8	0.66	0.68	0.71	0.72	0.75	0.73	0.72	0.78	0.90
9	0.66	0.68	0.71	0.74	0.76	0.75	0.76	0.77	0.90
10	0.65	0.68	0.71	0.74	0.77	0.78	0.78	0.78	0.89

Table 6. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for ARMA(1, 2) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the sample size is 500 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
2	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
3	0.10	0.19	0.30	0.40	0.50	0.60	0.70	0.80	0.90
4	0.34	0.21	0.27	0.39	0.50	0.60	0.70	0.80	0.90
5	0.61	0.55	0.40	0.37	0.48	0.60	0.70	0.80	0.90
6	0.65	0.65	0.62	0.50	0.48	0.59	0.70	0.80	0.90
7	0.63	0.68	0.67	0.63	0.58	0.57	0.70	0.80	0.90
8	0.64	0.68	0.68	0.71	0.69	0.60	0.69	0.80	0.90
9	0.65	0.69	0.69	0.73	0.71	0.67	0.68	0.80	0.90
10	0.64	0.67	0.71	0.75	0.74	0.74	0.69	0.80	0.90

Table 7. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for ARMA(1, 2) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the sample size is 5000 and the number of iterations is 1000

N/φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
2	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
3	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
4	0.13	0.19	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0.56	0.30	0.28	0.39	0.50	0.60	0.70	0.80	0.90
6	0.62	0.60	0.41	0.37	0.50	0.60	0.70	0.80	0.90
7	0.63	0.65	0.63	0.46	0.47	0.60	0.70	0.80	0.90
8	0.64	0.66	0.68	0.63	0.49	0.59	0.70	0.80	0.90
9	0.62	0.67	0.69	0.71	0.60	0.58	0.70	0.80	0.90
10	0.65	0.67	0.71	0.73	0.70	0.59	0.70	0.80	0.90

Table 8. The sample means of the parameter estimates $\hat{\varphi} = \hat{\phi}$ for ARMA(1, 2) processes with different parameter values φ using lags $N = 1, 2, 3, \dots, 10$. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the sample size is 50000 and the number of iterations is 1000

T	max	min	mean	median	sd	mad	skewness
50	1.00	0.00	0.413	0.409	0.326	0.436	0.222
500	0.999	0.00	0.502	0.495	0.218	0.187	0.207
5000	0.726	0.319	0.501	0.497	0.058	0.056	0.456
50000	0.561	0.443	0.501	0.502	0.019	0.019	-0.058

Table 9. The effect of the sample size T on the estimates $\hat{\varphi} = \hat{\phi}$ for an AR(1) process. The true parameter value $\varphi = 0.5$ and the lag $N = 3$. The number of iterations is 1000

T	max	min	mean	median	sd	mad	skewness
50	0.999	0.00	0.399	0.425	0.250	0.264	-0.062
500	0.681	0.00	0.481	0.491	0.086	0.078	-1.020
5000	0.570	0.395	0.499	0.500	0.024	0.023	-0.201
50000	0.527	0.474	0.500	0.500	0.008	0.007	0.036

Table 10. The effect of the sample size T on the estimates $\hat{\varphi} = \hat{\phi}$ for an ARMA(1, 2) process. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the true parameter value $\varphi = 0.5$ and the lag $N = 3$. The number of iterations is 1000

φ	max	min	mean	median	sd	mad	skewness
0.1	1.00	0.00	0.257	0.097	0.322	0.144	1.056
0.4	0.989	0.111	0.396	0.395	0.096	0.083	0.500
0.6	0.738	0.476	0.602	0.601	0.041	0.043	0.211
0.9	1.00	0.852	0.901	0.899	0.020	0.018	0.943

Table 11. The effect of the true parameter value φ on the estimates $\hat{\varphi} = \hat{\phi}$ for AR(1) processes. The sample size $T = 5000$ and the lag $N = 3$. The number of iterations is 1000

φ	max	min	mean	median	sd	mad	skewness
0.1	0.273	0.00	0.096	0.098	0.067	0.082	0.144
0.4	0.496	0.254	0.396	0.397	0.032	0.032	-0.198
0.6	0.650	0.540	0.600	0.600	0.018	0.019	-0.061
0.9	0.929	0.868	0.899	0.899	0.009	0.009	-0.076

Table 12. The effect of the true parameter value φ on the estimates $\hat{\varphi} = \hat{\phi}$ for ARMA(1, 2) processes. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the sample size $T = 5000$ and the lag $N = 3$. The number of iterations is 1000

N	max	min	mean	median	sd	mad	skewness
1	0.550	0.457	0.501	0.501	0.014	0.015	0.017
3	0.726	0.319	0.501	0.497	0.058	0.056	0.456
5	1.00	0.00	0.513	0.493	0.246	0.226	0.098
7	1.00	0.00	0.525	0.558	0.326	0.395	-0.216

Table 13. The effect of the lag N on the estimates $\hat{\varphi} = \hat{\phi}$ for an AR(1) process. The sample size $T = 5000$ and the true parameter value $\varphi = 0.5$. The number of iterations is 1000

N	max	min	mean	median	sd	mad	skewness
1	0.548	0.455	0.500	0.500	0.015	0.016	0.134
3	0.570	0.395	0.499	0.500	0.024	0.023	-0.201
5	0.710	0.00	0.482	0.499	0.112	0.092	-1.456
7	1.00	0.00	0.576	0.613	0.282	0.275	-0.488

Table 14. The effect of the lag N on the estimates $\hat{\varphi} = \hat{\phi}$ for an ARMA(1, 2) process. The MA parameters $\theta_1 = 0.8$ and $\theta_2 = 0.3$, the sample size $T = 5000$ and the true parameter value $\varphi = 0.5$. The number of iterations is 1000

References

[1] Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1991). [MR1093459](#)

- [2] Davis, R., Resnick, S.: Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics* **14**(2), 533–558 (1986). [MR0840513](#)
- [3] Francq, C., Zakoïan, J.-M.: Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**(4), 605–637 (2004)
- [4] Francq, C., Roy, R., Zakoïan, J.-M.: Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* **100**(470), 532–544 (2005)
- [5] Hamilton, J.D.: *Time Series Analysis*, 1st edn. Princeton university press, Princeton (1994). [MR1278033](#)
- [6] Hannan, E.J.: The estimation of the order of an ARMA process. *The Annals of Statistics* **8**(5), 1071–1081 (1980). [MR0585705](#)
- [7] Hannan, E.J.: The asymptotic theory of linear time-series models. *Journal of Applied Probability* **10**(1), 130–145 (1973)
- [8] Horváth, L., Kokoszka, P.: Sample autocovariances of long-memory time series. *Bernoulli* **14**(2), 405–418 (2008). [MR2544094](#)
- [9] Koreisha, S., Pukkila, T.: A generalized least-squares approach for estimation of autoregressive moving-average models. *Journal of Time Series Analysis* **11**(2), 139–151 (1990)
- [10] Lamperti, J.: Semi-stable stochastic processes. *Transactions of the American mathematical Society* **104**(1), 62–78 (1962)
- [11] Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M.S., Reisen, V.A.: Robust estimation of the scale and of the autocovariance function of Gaussian short-and long-range dependent processes. *Journal of Time Series Analysis* **32**(2), 135–156 (2011)
- [12] Lin, J.-W., McLeod, A.I.: Portmanteau tests for ARMA models with infinite variance. *Journal of Time Series Analysis* **29**(3), 600–617 (2008)
- [13] Ling, S., Li, W.: On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* **92**(439), 1184–1194 (1997)
- [14] Mauricio, J.A.: Exact maximum likelihood estimation of stationary vector ARMA models. *Journal of the American Statistical Association* **90**(429), 282–291 (1995)
- [15] McElroy, T., Jach, A.: Subsampling inference for the autocovariances and autocorrelations of long-memory heavy-tailed linear time series. *Journal of Time Series Analysis* **33**(6), 935–953 (2012)
- [16] Mikosch, T., Gadrich, T., Kluppelberg, C., Adler, R.J.: Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics* **23**(1), 305–326 (1995)
- [17] Viitasaari, L.: Representation of stationary and stationary increment processes via Langevin equation and self-similar processes. *Statistics & Probability Letters* **115**, 45–53 (2016)
- [18] Yao, Q., Brockwell, P.J.: Gaussian maximum likelihood estimation for ARMA models. I. time series. *Journal of Time Series Analysis* **27**(6), 857–875 (2006)