Autoregressive Models with Time-dependent Coefficients
A Comparison between Several Approaches

Rajae Azrak
Université Mohammed V

Guy Mélard
ECARES, SBS-EM, Université libre de Bruxelles

December 2017

ECARES working paper 2017-48
Autoregressive models with time-dependent coefficients - A comparison between several approaches

Rajae AZRAK 2 and Guy MÉLARD 3

Running title: tdAR models comparison

1 Correspondence to be sent to the second author.

2 Université Mohammed V - Rabat, Faculté des Sciences Juridiques Economiques et Sociales, Rabat, Morocco (E-mail address: raja.e.azrak@gmail.com).

3 Université Libre de Bruxelles, Solvay Brussels School of Economics and Management and ECARES, CP 114/4, Avenue Franklin Roosevelt, 50, B-1050 Brussels, Belgium (E-mail address: gmelard@ulb.ac.be). This work has benefited from an IAP-network in Statistics grant, contract P5/24, financially supported by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC), and a grant from Fonds de la Recherche Scientifique - FNRS, Belgium.
Autoregressive models with time-dependent coefficients -
A comparison between several approaches

Abstract
Autoregressive-moving average (ARMA) models with time-dependent (td) coefficients and marginally heteroscedastic innovation variance provide a natural alternative to stationary ARMA models. Several theories have been developed in the last fifteen years for parametric estimation in that context. In this paper, we focus on time-dependent autoregressive (tdAR) models and consider our theory in that case. We provide also an alternative theory for tdAR($p$) processes which relies on a $\rho$-mixing property. We compare the Dahlhaus theory for locally stationary processes and the Bibi and Francq theory, made essentially for cyclically time-dependent models, with our own theory. With respect to existing theories, there are differences in the basic assumptions (e.g. on derivability with respect to time or with respect to parameters) that are better seen on specific cases like the tdAR(1) process. There are also differences in terms of asymptotics as shown by an example. Our opinion is that the field of application can play a role here. The paper is completed by simulation results that show that the asymptotic theory can be used even for short series (less than 50 observations).

Key words and phrases. Nonstationary process; time series; time-dependent model; time-varying model; $\rho$-mixing property; locally stationary processes.
1 Introduction

Autoregressive-moving average (ARMA) models with time-dependent (td) coefficients and marginally heteroscedastic innovation variance provide a natural alternative to stationary ARMA time series models. Several theories have been developed in the last fifteen years for parametric estimation in that context.

To simplify our presentation, let us consider the case of the tdAR(1) model with a time-dependent coefficient $\phi_t^{(n)}$, which depends on time $t$ and also on $n$, the length of the series. Let also $g_t^{(n)} > 0$ and $(e_t, t \in \mathbb{N})$ be a white noise process, consisting of independent random variables, not necessarily identically distributed, with mean zero and with standard deviation $\sigma > 0$, and 4-th order cumulant $\kappa_{4t}$. The model is defined by

$$w_t^{(n)} = \phi_t^{(n)} w_{t-1}^{(n)} + g_t^{(n)} e_t.$$

The coefficient $\phi_t^{(n)}$ and $g_t^{(n)}$ depend on $t$ and sometimes, not always, on $n$, but also on parameters. We denote $\sigma_t^{(n)} = g_t^{(n)} \sigma$, the innovation standard deviation. For given $n$, consider a sequence of observations $w^{(n)} = (w_1^{(n)}, w_2^{(n)}, \ldots, w_n^{(n)})$ of the process. When $\phi_t^{(n)}$ or $g_t^{(n)}$ depend on $n$ we should speak of a triangular array process, not of a stochastic process. Note that we use $g_t^{(n)}$ instead of $\{h_t^{(n)}\}_{t=1}^{n}$ in [1] to comply with the notations introduced in multivariate models, see [2].

The AR(1) process with a time-dependent coefficient has been considered by [3, 4, 5], [6, 7] have extended the results to autoregressive processes of order $p$, [1], denoted AM, and [8], denoted BF, have considered tdARMA processes. Contrarily to AM, in BF the coefficients depend only on $t$, not on $n$. Besides, although the basic assumptions of BF and AM are different, their asymptotics are somewhat similar but differ considerably from those of Dahlhaus [7] based on locally stationary processes (LSP). See the nice overview in [9]. For simplicity, we will compare these approaches on autoregressive models.

Two approaches can be sketched for asymptotic theories within nonstationary processes, see [10]. Approach 1 consists in analyzing the behavior of the process when $n$ tends to infinity. That assumes some generating mechanism in the background that remains the same over time. Two examples can be mentioned: processes with periodically changing coefficients and cointegrated processes. It is in the former context that BF have established asymptotic properties for parameter estimates in the case where $n$ goes to infinity. Approach 2 for asymptotics within nonstationary processes consists in determining how estimates that are obtained for a finite and fixed sample size behave. This is the setting for describing in general the properties of a test under local alternatives (where the parameter space is rescaled by $1/\sqrt{n}$), or in nonparametric regression. Approach 2 is the framework considered in [7] for LSP that we will briefly summarize now. First, there is an assumption of local stationarity that imposes continuity with respect to time and even differentiability. But also, $n$ is not simply increased to infinity. The coefficients, like $\phi_t^{(n)}$ are considered as a function of rescaled time $t/n$. Therefore, everything happens as if time is rescaled to the interval $[0; 1]$. Suppose $\phi_t^{(n)} = \tilde{\phi}_{t/n}$ and $g_t^{(n)} = \tilde{g}_{t/n}$, where $\tilde{\phi}_u$ and $\tilde{g}_u$, $0 \leq u \leq 1$, depend on a finite number of parameters, are differentiable functions of $u$ and such that $|\tilde{\phi}_u| < 1$ for all $u$. The model is written as
\[ w^{(n)}_t = \tilde{\phi}_{t/n} w^{(n)}_{t-1} + \tilde{g}_{t/n} e_t. \]  

(1.2)

As a consequence, the assumptions made in the LSP theory are quite different from those of AM and BF, for example, because of the different nature of the asymptotics. The AM approach is somewhere between these two approaches 1 and 2, sharing parts of their characteristics but not all of them. In Section 2, we specialize the assumptions of AM to AR(\(p\)) processes and consider the special case of a tdAR(1) process. In Appendix B, we provide an alternative theory for tdAR(\(p\)) processes which relies on a \(\rho\)-mixing property.

The AM theory is further illustrated in Appendix C by simulation results on a tdAR(2) process. The LSP and BF theories are summarized in Sections 3 and 4, respectively. In Section 5, we compare the Dahlhaus LSP theory with our own AM theory. This is partly explained thanks to examples. The differences in the basic assumptions are emphasized. Similarly, in Section 6, a comparison is presented between the AM and BF approaches [8], before the conclusions in Section 7.

2 The AM theory for time-dependent autoregressive processes

Let us consider the AM theory in the special case of tdAR(\(p\)) processes. We want also to see if simpler conditions can be derived for the treatment of pure autoregressive processes.

We consider a triangular array of random variables \(w = (w^{(n)}_t, t = 1, \ldots, n, n \in \mathbb{N})\) defined on a probability space \((\Omega, F, P_\beta)\), with values in \(\mathbb{R}\), whose distribution depends on a vector \(\beta = (\beta_1, \ldots, \beta_r)\) of unknown parameters to be estimated, with \(\beta\) lying in an open set \(B\) of an Euclidean space \(\mathbb{R}^r\). The true value of \(\beta\) is denoted by \(\beta^0\). By abuse of language, we will nevertheless talk about the process \(w\).

Definition 1

The process \(w\) is called an autoregressive process of order \(p\), with time-dependent coefficients, if, and only if, it satisfies the equation

\[ w^{(n)}_t = \sum_{k=1}^{p} \phi^{(n)}_{tk} w^{(n)}_{t-k} + g^{(n)}_t e_t, \]  

(2.1)

where \((e_t, t \in \mathbb{N})\) and \(g^{(n)}_t\) is a deterministic strictly positive function of time.

We denote again \(\sigma_t^{(n)} = \sigma g^{(n)}_t\). The initial values \(w_t, t < 1\), are supposed to be equal to zero. The \(r\)-dimensional vector \(\beta\) contains all the parameters to be estimated, those in \(\phi^{(n)}_{tk}, k = 1, \ldots, p\), and those in \(g^{(n)}_t\) but not the scale factor \(\sigma\) which is estimated separately.

We suppose a specific deterministic parameterization in function of \(t\) and \(n\). Let \(\phi^{(n)}_{tk}(\beta)\) be the parametric coefficient with \(\phi^{(n)}_{tk}(\beta^0) = \phi^{(n)}_{tk}\), and similarly \(g^{(n)}_t(\beta^0) = g^{(n)}_t\). Let \(e^{(n)}_t(\beta)\) be the residual for a given \(\beta\):

\[ e^{(n)}_t(\beta) = w^{(n)}_t - \sum_{k=1}^{p} \phi^{(n)}_{tk}(\beta) w^{(n)}_{t-k}. \]  

(2.2)
Note that \( e_t^{(n)}(\beta^0) = g_t^{(n)}(\beta^0)e_t \).

Thanks to the assumption about initial values and by using (2.1) recurrently, it is possible to write the pure moving average representation of the process:

\[
  w_t^{(n)} = \sum_{k=0}^{t-1} \psi_{tk}^{(n)}(\beta^0)g_{t-k}^{(n)}(\beta^0)e_{t-k}
\]  

(see [1] for a recurrence formula). Let \( F_t \) be the \( \sigma \)-field generated by the \((w_s^{(n)}, s \leq t)\), hence by \((e_s, s \leq t)\), which explains why a superscript \(^{(n)}\) is suppressed, and \( F_0 = \{\emptyset, \Omega\} \). To simplify the presentation, we denote \( E_{\beta^0}(\cdot)(\beta) = \{E_{\beta^0}(\cdot)(\beta)\}_{\beta = \beta^0} \) and similarly \( \text{var}_{\beta^0}(\cdot) \) and \( \text{cov}_{\beta^0}(\cdot) \). We are interested in the Gaussian quasi-maximum likelihood estimator

\[
  \hat{\beta}^{(n)} = \arg\min_{\beta \in \mathbb{R}^r} \frac{1}{2} \sum_{t=1}^n \left\{ \log[\sigma_t^{(n)}(\beta)]^2 + \left( \frac{e_t^{(n)}(\beta)}{\sigma_t^{(n)}(\beta)} \right)^2 \right\}.
\]  

Denote \( a_t^{(n)}(\beta) \) the expression between curved brackets in (2.4). Note that the first term of \( a_t^{(n)}(\beta) \) will sometimes be omitted, corresponding to a weighted least squares method, especially when \( \sigma_t^{(n)}(\beta) \) does not depend on the parameters, or even ordinary least squares, when \( \sigma_t^{(n)}(\beta) \) does not depend on \( t \). BF considers that estimation method, and also a variant where the denominator is replaced by a consistent estimator. Other estimators are also used in the LSP theory, see Section 3.

We need expressions for the derivatives of \( e_t^{(n)}(\beta) \) with respect to \( \beta \) using (2.2). The first derivative is

\[
  \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} = -\sum_{k=1}^P \frac{\partial \phi_{tk}^{(n)}(\beta)}{\partial \beta_i} w_t^{(n)} - \sum_{k=1}^{t-1} \psi_{tk}^{(n)}(\beta, \beta^0)g_{t-k}^{(n)}(\beta^0)e_{t-k},
\]  

(2.5)

\[ i = 1, \ldots, r. \]  

It will be convenient to write it as a pure moving average\(^4\)

\[
  \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} = -\sum_{k=1}^{t-1} \psi_{tk}^{(n)}(\beta, \beta^0)g_{t-k}^{(n)}(\beta^0)e_{t-k},
\]  

(2.6)

for \( i = 1, \ldots, r \), where the coefficients \( \psi_{tk}^{(n)}(\beta, \beta^0) \) are obtained by the following relations

\[
  \psi_{tk}^{(n)}(\beta, \beta^0) = \sum_{u=1}^k \frac{\partial \phi_{tu}^{(n)}(\beta)}{\partial \beta_i} \psi_{t-k-u}^{(n)}(\beta^0).
\]

\(^4\)We have improved the presentation of AM in the light of [8], especially by making it clear that some quantities like \( \psi_{tk}^{(n)}(\beta, \beta^0) \) depend on both \( \beta \) and \( \beta^0 \). The notations for the innovations are also changed to emphasize that \( F_t \) doesn’t depend on \( n \).
Similarly, we introduce \( \psi_{tijk}^{(n)}(\beta, \beta^0) \) and \( \psi_{tijlk}^{(n)}(\beta, \beta^0) \) using the second and third derivatives of \( e_t^{(n)}(\beta) \), for \( i, j, l = 1, \ldots, r \).

Under all the assumptions of Theorem 2' of [1], see Appendix A, the estimator \( \hat{\beta}^{(n)} \) converges in probability to \( \beta^0 \) and, furthermore, \( \sqrt{n}(\hat{\beta}^{(n)} - \beta^0) \overset{L}{\to} N(0, V^{-1}WV^{-1}) \) when \( n \to \infty \), where, with \( ^T \) denoting transposition,

\[
W = \lim_{n \to \infty} \frac{1}{4n} \sum_{t=1}^{n} E_{\beta^0} \left( \frac{\partial \alpha_t^{(n)}(\beta) \partial \alpha_t^{(n)}(\beta)}{\partial \beta^T} \right),
\]

and

\[
V = \lim_{n \to \infty} \frac{1}{2n} \sum_{t=1}^{n} E_{\beta^0} \left( \frac{\partial^2 \alpha_t^{(n)}(\beta)}{\partial \beta \partial \beta^T} | F_{t-1} \right).
\]

**Example. The tdAR(1) process**

Let us consider an tdAR(1) process defined by (1.1). We have for the \( \psi_{tk}^{(n)}(\beta^0) \) in (2.3)

\[
\psi_{tk}^{(n)}(\beta^0) = \prod_{l=0}^{k-1} \phi_{t-l}^{(n)}(\beta^0), \quad k = 1, \ldots, t-1,
\]

where a product for an empty set of indices is set to one. Similarly

\[
\psi_{tik}^{(n)}(\beta^0) = \frac{\partial \phi_t^{(n)}(\beta)}{\partial \beta_i} \psi_{t-i,k-1}^{(n)}(\beta^0) = \left. \frac{\partial \phi_t^{(n)}(\beta)}{\partial \beta_i} \right|_{\beta = \beta^0} \prod_{l=1}^{k-1} \phi_{t-l}^{(n)}(\beta^0),
\]

and analogous expressions for second and third derivatives. The following is an application of Theorem 2' of [1].

**Theorem 2A**

Consider a tdAR(1) process defined by (1.1) under the assumptions of Theorem 2' in Appendix A except that \( H_{2',1} \) is replaced by \( H_{2',1A} \):

\( H_{2',1A} \) Let us suppose that there exist constants \( C, \Psi \) \((0 < \Psi < 1)\), \( M_1, M_2, \) and \( M_3 \) such that the following inequalities hold for all \( i, j, l, k \):

\[
\left| \psi_{tk}^{(n)}(\beta^0) \right| < C \psi^k, \quad \left| \left\{ \frac{\partial \phi_t^{(n)}(\beta)}{\partial \beta_i} \right\}_{\beta = \beta^0} \right| < M_1,
\]

and analogous bounds \( M_2 \) and \( M_3 \) for the second and third-order derivatives.

Then the results of Theorem 2' of [1] are still valid.
Proof

Let us show the first of the inequalities in $H_{2,1}$ since the others are similar. Consider for $\nu = 1, \ldots, t - 1$

\[
\sum_{k=v}^{t-1} \{ \psi_{tik}^{(n)}(\beta^0, \beta^0) \}^2 = \left( \frac{\partial \phi_t^{(n)}(\beta)}{\partial \beta_i} \right)^2 \sum_{k=v}^{t-1} \{ \psi_{t,i,k-1}^{(n)}(\beta^0) \}^2 < \frac{M_1^2 C^2 (\psi^2)^{v-1}}{1 - \psi^2},
\]

hence $N_1 = M_1^2 C^2 (1 - \psi^2)^{-1}$ and $\Phi = \psi^2 < 1$. \hfill \Box

Remark

Note that the first inequality of $H_{2,1A}$ is true when $|\phi_t^{(n)}(\beta^0)| < 1$ for all $t$ and $n$ but this is not a necessity. A finite number of those $\phi_t^{(n)}(\beta^0)$ can be greater than 1 without any problem. For example $\phi_t^{(n)}(\beta) = (4 + \beta/n)(t/n)(1 - t/n)$, with $0 < \beta < 1$ would be acceptable because the interval around $t/n = 0.5$ where the coefficient is greater than 1 shrinks when $n \to \infty$. With this in mind, Example 3 of [1] can be slightly modified in order to allow that the upper bound of the $|\phi_t^{(n)}(\beta^0)|$'s be greater than one. This will be illustrated in Section 5. Note also that the other inequalities of $H_{2,1A}$ are easy to check.

One of the assumptions of Theorem 2' in Appendix A, $H_{2,7}$, is particularly strange at the first sight, although it could be checked in the examples of [1, Section 4]. It is interesting to note that, at least in the framework of autoregressive processes, it can be replaced by a more standard $\rho$-mixing condition. This is done in Appendix B. We were unfortunately not able to generalize it to time-dependent moving average (MA) or ARMA models.

In [1] the few simulations were presented for simple tdAR(1) and tdMA(1) models, moreover when the generated series were stationary. It allowed us to assess empirically the theory but not to put it in jeopardy. In Appendix C, we show simulations for a tdAR(2) model where the true coefficients $\phi_t^{(n)}$ and $\phi_t^{(n)}$ vary in an extreme way. The parameterization used is

\[
\phi_{t,k}^{(n)}(\beta) = \phi_k + \frac{1}{n-1} (t - \frac{n + 1}{2}) \phi_k^n, \quad k = 1,2. \tag{2.9}
\]

The results are good, although the model cannot be fitted on some of the generated series, especially for short series of length $n = 50$. Except for the behavior at the extreme end of the series, and with a small approximation of $t/(n - 1)$ by $t/n$, it will be seen in Section 5 that the data generator process fulfills the assumptions of the LSP theory, so these simulations can also be seen as illustrative of that theory. It should be noted that there are few simulation experiments mentioned in the LSP literature. The word “simulation” is mentioned twice in [9] but each time to express a request.

In [1, Section 6], there are illustrative numerical examples of the AM theory. They are all based on Box-Jenkins series A, B and G, see [11]. The last two series are closing IBM stock prices and international airline passenger numbers. In [12], there is an example of vector tdAR and tdMA models on monthly log returns of IBM stock and S&P500 index from January 1926 to December 1999, treated first in [13].
3 The theory of locally stationary processes

We have given in Section 1 some elements of the theory of Dahlhaus. It is based on a class of locally stationary processes (LSP), that means a sequence of stationary processes, based on a stochastic integral representation

\[ w_t^{(n)} = \int_{-\pi}^{\pi} e^{i\lambda t} A_t^{(n)}(\lambda) d\xi(\lambda), \]  

where \( \xi(\lambda) \) is a process with independent increments and \( A_t^{(n)}(\lambda) \) fulfills a condition so as to be called a slowly varying function with respect to \( t \).

In the case of autoregressive processes, which are emphasized in this paper, for example an AR(1) process, that means that the observations around time \( t \) are supposed to be generated by a stationary AR(1) process with some coefficient \( \phi_t \). Stationarity implies that \(-1 < \phi_t < 1\). Around time \( t \), fitting is done using the process at time \( t \). More generally, for AR(\( p \)) processes, the autoregressive coefficients are such that the roots of the autoregressive polynomial are greater than 1 in modulus.

The estimation method is based either on a spectral approach or on a Whittle approximation of the Gaussian likelihood. [14] also sketches an exact maximum likelihood estimation method like the one used here based on [15] or [16].

As mentioned above, Dahlhaus approach of doing asymptotics relies on rescaling time \( t \) in \( u = t/n \). That does not mean that the process is considered in continuous time but at least that its coefficients are considered in continuous time. Asymptotics are done by assuming an increasing number of observations between 0 and 1. That means that the coefficients are considered as a function of \( t/n \), not separately as a function of \( t \) and \( n \). This is nearly the same as was assumed in (2.9), since \( t/(n-1) \) is close to \( t/n \) for large \( n \).

Note however that Example 1 of [1] is not in that class of processes. More generally, processes where the coefficients are periodic functions of \( t \) are excluded from the class of processes under consideration. Of course, what was said about the coefficients is also valid for the innovation standard deviation. If the latter is a periodic function of time \( t \), with a given period \( s \), the process is not compatible with time rescaling. We will compare the LSP theory with the AM theory in Section 5.

4 The theory of cyclically time-dependent models

Here we will focus on BF, [8], but part of the discussion is also appropriate for older approaches like [4, 5, 6]. BF have developed a general theory of estimation for linear models with time-dependent coefficients particularly aimed at the case of cyclically time-dependent coefficients. See also [17, 18, 19, 20].

The linear models include autoregressive but also moving average (MA) or ARMA models like AM. The coefficients can depend on \( t \) in a general way but not on \( n \). Hence, \( \phi_t^{(n)} \) is written \( \phi_{tk} \) in Definition 1. Heteroscedasticity is allowed in a similar way in the sense that the innovation variance can depend on \( t \) (but not on \( n \)). The estimation method is a quasi-generalized least squares method.
The BF theory supports several classes of models. The periodic ARMA or PARMA models, where the coefficients are periodic functions of time, is an important class. Note that the period does not need to be an integer. But [8, Section 3] also consider a switching model based on $\Delta$, a subset of integers in $\{1, 2, \ldots, n\}$ and its complement $\Delta^c$. For example, $\Delta$ can be associated to weekdays and $\Delta^c$ to the weekend. Then, the coefficient, e.g. $\phi_t$ in (1.1), depends on a parameter $\beta = (a, \tilde{a})$ in the following way: $\phi_t = a1_{\Delta}(t) + \tilde{a}1_{\Delta^c}(t)$, where $1_{\Delta}(t)$ denotes the indicator function, equal to 1 if $t$ belongs to $\Delta$, and 0 otherwise. Consequently there are two different regimes. But the composition of $\Delta$ and $\Delta^c$ can also be generated by an i.i.d. sequence of Bernoulli experiments, with some parameter $\pi$, provided they are independent from the white noise process $(e_t, t \in \mathbb{N})$.

Under appropriate assumptions, there are a theorem of almost sure consistency and a theorem of convergence in law, somewhat like in Theorem 2' of Appendix A. Note that strong consistency is proved here, not just in probability. We will compare the BF theory with the AM theory in Section 6.

5 A comparison with the theory of locally stationary processes

In this Section, we compare the AM approach described in Section 2 with the LSP approach described in Section 3. The basic model is (2.1) although the coefficients $\phi_{tk}^{(n)}$, $k = 1, \ldots, p$, and $g_t^{(n)}$ depend on $t$ and $n$ through $t/n$ only. Although LSP can be moving averages, see [7], the latter are rarely mentioned in the LSP literature. The overview paper [9] mentions “moving average” only once, and it is in a reference.

Dependency with respect to $u = t/n$ of the model coefficients as well as the innovation standard deviation is assumed to be continuous and even differentiable in the Dahlhaus theory. In comparison, the other theories including AM and BF, see [8], accept discrete values of the coefficients with respect to time, without requiring a slow variation. They make instead assumptions of differentiability with respect to the parameters.

Another point of discussion is as follows. In order to handle economic and social data with an annual seasonality, [11] have proposed the so-called seasonal ARMA processes, where the autoregressive and moving average polynomials are products of polynomials in the lag operator $B$ and polynomials in $B^s$ for some $s > 1$, for example $s = 12$, for monthly data, or $s = 4$, for quarterly data. Although series generated by these processes are not periodic, with suitably initial values, they can show a pseudo-periodic behavior with period $s$. Let us consider such ARMA processes with time-dependent coefficients, for example an AR(12) defined by the equation $y_t = \phi_{tk}^{(n)}(\beta)y_{t-12} + e_t$, with the same notations as in Section 1. There are exactly 11 observations between times $t$ and $t - 12$ and an increase of the total number of observations would not affect that. For such processes, Approach 1 of doing asymptotics, described in Section 1, seems to be the most appropriate, assuming that there is a larger number of years, not that there is a larger number of months within a year. Of course Approach 2 of doing asymptotics is perfectly valid in all cases where the frequency of observation is more or less arbitrary.

To conclude, the AM approach is better suited for economic time series, where we can imagine that more years will become available, see the left hand part of Figure 1. In other contexts, like in biology and engineering, we can imagine that more data become available
with an increasing sampling rate, see the right hand side of Figure 1. Then the LSP theory seems more appropriate.

![Figure 1. Schematic presentation on how to interpret asymptotics in AM and Dahlhaus LSP theories (see text for details).](image)

In the following example, we will consider a tdAR(1) process but with the innovation standard deviation being a periodic function of time. Let us first show a unique artificial series of length 128 generated by (1.1) with

$$\phi_t^{(n)} = \phi' + (t - \frac{n + 1}{2})\phi'',$$

(5.2)

with $\phi' = 0.15$, $\phi'' = 0.015$ and the $e_t$ are normally and independently distributed with mean 0 and variance $g_t$, where $g_t$ is a periodic function of $t$ with period 12, simulating a seasonal heteroscedasticity for monthly data. Furthermore $g_t$, which doesn’t depend on $n$, assumes values $g = 0.5$ and $1/g = 2$, each during six consecutive time points. We have omitted the factor $1/(n - 1)$ here since only one series length is considered. The series plotted in Figure 2 clearly shows a nonstationary pattern. The choice of $\phi' = 0.15$ and $\phi'' = 0.015$ is such that the autoregressive coefficient follows a straight line which goes slightly above +1 at the end of the series (see Figure 3). The parameters are estimated using the exact Gaussian maximum likelihood method which provides the following estimates (with the standard errors): $\hat{\phi}' = 0.157 \pm 0.062$, $\hat{\phi}'' = 0.0159 \pm 0.0014$, and $\hat{g} = 0.344 \pm 0.044$, which are compatible with the true values. For $n = 128$, we provide the fit of $\phi_t^{(n)}$ and $g_t$, respectively, in Figures 3 and 4. Figures 5 and 6 give a better insight on the relationship between the observations, showing broadly a negative autocorrelation during the first half of the series and a positive autocorrelation during the second half, as well as a small scatter during half of the year and a large scatter during the other half. Note finally that this example is not compatible with the LSP theory since $\phi_t^{(n)} > 1$ for some $t$, and $g_t$ being piecewise constant is not a differentiable function of time. Also the asymptotics related to that theory will also be difficult to interpret since $g_t$ is periodic with a fixed period.
Figure 2. Artificial series produced using the process defined by (1.1) and (5.2) (see text for details).

Figure 3. True value of $\phi_t^n$ (which goes above 1!) (solid line) and its fit (discontinuous line).

We have run Monte Carlo simulations using the same setup except that polynomials of degree 2 were fitted for $\phi_t^n$ instead of a linear function of time. The parameterization is

$$\phi_t^n(\beta) = \phi' + (t - \frac{n + 1}{2})\phi'' + (t - \frac{n + 1}{2})^2\phi'''$$

(5.3)

and $g_t$ is a periodic function which oscillates between the two values $g$ and $1/g$, defined like above. The estimation program is the same as in AM but extended to cover polynomials of time of degree up to 3 as well as for AR (or similarly MA) coefficients as for $g_t^n$. The latter capability is not used here but well an older implementation for intervention analysis [21]. Estimates are obtained by numerically maximizing the exact Gaussian likelihood.
A number of 1000 series of length 128 were generated using a program written in Matlab with Gaussian innovations and without warm-up. Note that results were obtained for 964 series only. They are provided in Table 1. Unfortunately, some estimates of the standard errors were unreliable so their averages were useless and replaced by medians. The estimates of the standard errors are quite close to the empirical standard deviations. The fact that the results are not as good as the simulation experiments described by [1, Section 5], at least for series of 100 observations or more, may be due to the fact that the basic assumptions are only barely satisfied with \( \phi_t(n) \) going nearly from about \(-1\) to \(1\). In Table 2, we have fitted the more adequate and simpler model with a linear function of time instead of a quadratic function of time. Now results were obtained for 999 series and the estimated standard error were always reliable so that their average across simulations are displayed.
Figure 6. $w_t^{(128)}$ as a function of $w_{t-1}^{(128)}$ (plusses: high scatter, when $g_t = 2$, circles: small scatter, when $g_t = 0.5$).

To end the Section on a positive side, let us say that the three illustrative examples in [1, Section 6] would give approximatively the same results under the LSP theory, provided that the same estimation method is used.

Table 1. Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and medians across simulations of estimated standard errors $\phi'$ (true value: 0.15), $\phi''$ (true value: 0.015), and $\phi'''$ (true value: 0), and $g$ (true value 0.5) for the tdAR(1) model described above, for $n = 128$; 964 replications (out of 1000).

<table>
<thead>
<tr>
<th>Parameter $n$</th>
<th>true value</th>
<th>Average</th>
<th>standard deviation</th>
<th>median of standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi'$</td>
<td>0.15</td>
<td>0.23554</td>
<td>0.14611</td>
<td>0.10380</td>
</tr>
<tr>
<td>$\phi''$</td>
<td>0.015</td>
<td>0.01282</td>
<td>0.00222</td>
<td>0.00160</td>
</tr>
<tr>
<td>$\phi'''$</td>
<td>0.0</td>
<td>-0.00000</td>
<td>0.00005</td>
<td>0.00005</td>
</tr>
<tr>
<td>$g$</td>
<td>0.5</td>
<td>0.54054</td>
<td>0.07857</td>
<td>0.08157</td>
</tr>
</tbody>
</table>

Table 2. Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and averages across simulations of estimated standard errors $\phi'$ (true value: 0.15), $\phi''$ (true value: 0.015), and $g$ (true value 0.5) for the tdAR(1) model described above, for $n = 128$; 999 replications (out of 1000).

<table>
<thead>
<tr>
<th>Parameter $n$</th>
<th>true value</th>
<th>average</th>
<th>standard deviation</th>
<th>average of standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi'$</td>
<td>0.15</td>
<td>0.22023</td>
<td>0.12577</td>
<td>0.06683</td>
</tr>
<tr>
<td>$\phi''$</td>
<td>0.015</td>
<td>0.01305</td>
<td>0.00202</td>
<td>0.00146</td>
</tr>
<tr>
<td>$g$</td>
<td>0.5</td>
<td>0.54290</td>
<td>0.07847</td>
<td>0.06984</td>
</tr>
</tbody>
</table>
6 A comparison with the theory of cyclically time-dependent models

In this Section, we compare the AM approach described in Section 2 with the BF approach described in Section 4. The basic model is (2.1) without the superscripts \( ^{(n)} \) since the coefficients \( \phi_{tk}, k = 1, \ldots, p, \) and \( g_t \) do not depend on \( n \). Therefore, it is clear that there is no sense to compare the LSP and the BF theories which act on disjoint classes of processes.

We have mentioned in Section 4 a few classes of processes to which the BF theory is applicable. The PARMA processes are surely compatible with the AM theory even with an irrational period for the periodic functions of time. On the contrary, the switching model based on an i.i.d. sequence of Bernoulli experiments is not a particular case of the models treated in [1]. This is characteristic of the BF theory that the two other theories cannot handle, at the first sight. When the switching model is based on a fixed subset of integers \( \Delta \), the AM theory can be adapted especially in the weekdays versus weekend example. On the other side, Examples 2 to 5 of [1] are incompatible with the BF theory since the coefficients depend on \( n \).

The basic assumptions of BF are different from those of AM. A comparison is difficult here but it is interesting to note a less restrictive assumption of existence of fourth order moments, not eighth order like in AM. Note that [22] has removed that requirement for the AM theory. Note that the expression for \( I \) in [8], which corresponds to our \( W \) in (2.7), does not involve 4th order moments since no parameter is involved in heteroscedasticity.

The process considered in Section 5 was an example for which the theory of locally stationary process would not apply because of the periodicity in the innovation standard deviation. That process is also an example for which the BF theory would not apply but this time because the autoregressive coefficient is a function of \( t \) and \( n \); not only of \( t \).

7 Conclusions

This paper was motivated by suggestions to see if the results in [1] simplify much in the case of autoregressive or even tdAR(1) processes, and by requests to compare more deeply the AM approach with others and push it in harder situations. We have recalled the main result in Appendix A.

We have shown that there are not many simplifications for tdAR processes, perhaps due to the intrinsically complex nature of ARMA processes with timedependent coefficients. Nevertheless, we have been able to simplify one of the assumptions for tdAR(1) processes. We have taken the opportunity of this study on autoregressive processes with time dependent coefficients in order to develop in Appendix B an alternative approach based on a \( \rho \)-mixing condition instead of the strange assumption \( H_{2t,7} \) made in AM. At least we could check that assumption in some examples which is not the case for the mixing condition, at the present time. Note that a mixing approach was the first we tried, before preferring \( H_{2t,7} \). The latter could be extended to tdMA and tdARMA processes, which was not the case for the mixing condition. Although theoretical results for tdAR(2) processes could not be shown in closed form expressions, the simulations in Appendix C indicate that the method is robust when causality becomes questionable. We have shown more stressing simulations than in AM. ARIMA models could have been possible for these simulations and examples but this paper has focused on autoregressive processes. Practical examples of tdARIMA models were already given in [1] and [12].
We have also compared the AM approach to others, more especially Dahlhaus LSP theory and the BF approach aimed at cyclically time-dependent linear processes. Let us comment on this more deeply.

Like in the Dahlhaus LSP theory, a different process is considered by AM for each \( n \) in the asymptotics. There are however several differences in the two approaches: (a) AM can cope with periodic evolutions with a fixed period, either in the coefficients or in the variance; (b) AM does not assume differentiability with respect to time but well with respect to the parameters, (c) to compensate, AM makes other assumptions which are more difficult to check; (d) which may explain why Dahlhaus theory is more widely applicable: other models than just ARMA models, other estimation methods than maximum likelihood, even semi-parametric methods, existence of a LAN approach, etc; (e) AM is purely time-domain oriented whereas Dahlhaus theory is based on a spectral representation. An example with an economic inspiration and its associated simulation experiments have shown that some of these assumptions of AM are less restrictive but there is no doubt that others are more stringent. In our opinion, the field of applications can have an influence on the kind of asymptotics. Dahlhaus LSP approach is surely well adapted to signal measurements in biology and engineering where the sample span of time is fixed and the sampling interval is more or less arbitrary. This is not true in economics and management where (a) time series models are primarily used to forecast a flow variable like sales or production, obtained by accumulating data over a given period of time, a month or a quarter, so (b) that the sampling period is fixed, and (c) moreover, some degree of periodicity is induced by seasonality. Here, it is difficult to assume that more observations become available during one year without affecting strongly the model. For that reason, even if the so-called seasonal ARMA processes, which are nearly the rule for economic data, are formally special cases of locally stationary processes, the way of doing asymptotics is not really adequate. For the same reason, rescaling time is not natural when the coefficients are periodic function of time.

Going now to a comparison of AM with the BF approach mainly aimed at cyclically time-dependent linear processes, we see a first fundamental difference in the fact that a different process is considered for each \( n \) in AM, not in BF. That assumption of dependency on \( n \) as well as on \( t \) was introduced in order to be able to do asymptotics in cases that would not have been possible otherwise (except in adopting the Dahlhaus approach, of course) but, at the same time making it possible to represent a periodic behavior. When the coefficients are only dependent on \( t \), not on \( n \), the AM and BF approaches come close in the sense that (a) the estimation methods are close; (b) the assumptions are quite similar. The example and simulations shown to distinguish AM from LSP is also illuminating the difference between AM and BF. There remains that the switching model based on an i.i.d. Bernoulli process is not feasible in the AM approach.

In some sense, AM can be seen as partly taking some features of both Dahlhaus LSP and BF approaches. Some features, like periodicity of the innovation variance, can be handled well in BF while others, like slowly time varying coefficients are in the scope of LSP. But cyclical behavior of some innovation variance and slowly varying coefficients together (or the contrary: cyclical behavior of some coefficients and slowly varying innovation variance) are not covered by Dahlhaus and BF theories but well by AM. The example of Section 5 may look artificial but includes all the characteristics that are not covered well by locally stationary processes and the corresponding asymptotic theory. It includes a time dependent first-order autoregressive coefficient \( \phi_t^{(n)} \) which is very realistic for an
I(0) (i.e. not integrated) economic time series and an innovation variance $\sigma^2_t$ which is a periodic function of time (this can be explained by seasonality like a winter/summer effect). To emphasize the difference with LSP approach, we have assumed that $\phi_t^{(n)}$ goes slightly outside of the causality (or stationarity, in Dahlhaus terminology) region for some time and that $\sigma^2_t$ is piecewise constant, hence not compatible with differentiability at each time.

**Acknowledgements**

We thank the two anonymous referees who made very useful suggestions, the editor and the editorial office. We thank those who have made comments to our previous paper, including Christian Francq, Marc Hallin, and mainly Rainer Dahlhaus and Denis Bosq.

**Appendix A. Theorem 2′ of [1]**

Consider an autoregressive-moving average process with time dependent coefficients (tdARMA) and suppose that the functions $\phi_t^{(n)}(\beta)$, $\theta_t^{(n)}(\beta)$ and $g_t^{(n)}(\beta)$ are three times continuously differentiable with respect to $\beta$, in the open set $B$ containing the true value $\beta^0$ of $\beta$, that there exist positive constants $\Phi < 1$, $N_1$, $N_2$, $N_3$, $N_4$, $N_5$, $K_1$, $K_2$, $K_3$, $m$, $M$, $m_1$ and $K$, such that $\forall t = 1, \ldots, n$ and uniformly with respect to $n$:

$$H_{2.1} \sum_{k=1}^{t-1} \psi_t^{(n)}(\beta^0, \beta^0)^2 < N_4 \Phi^{\nu-1}, \quad \sum_{k=1}^{t-1} \psi_t^{(n)}(\beta^0, \beta^0)^4 < N_2 \Phi^{\nu-1},$$

$$\sum_{k=1}^{t-1} \psi_t^{(n)}(\beta^0, \beta^0)^2 < N_5, \quad \nu = 1, \ldots, t - 1, \quad i, j, l = 1, \ldots, r;$$

$$H_{2.2} \left| \frac{\partial g_t^{(n)}(\beta)}{\partial \beta_i} \right| \bigg|_{\beta = \beta^0} \leq K_1, \quad \left| \frac{\partial^2 g_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j} \right| \bigg|_{\beta = \beta^0} \leq K_2,$$

$$\left| \frac{\partial^3 g_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} \right| \bigg|_{\beta = \beta^0} \leq K_3, \quad i, j, l = 1, \ldots, r;$$

$$H_{2.3} 0 < m \leq g_t^{(n)}(\beta^0) \leq m_1;$$

$$H_{2.4} E(\{\nu_t^{(n)}(\beta)\}^4) \leq M;$$

$$H_{2.5} E(e_t^8) \leq K.$$

Suppose furthermore that


\[ H_{2',6} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \left[ \sigma^{-2} E_{\beta^0} \left( \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} \left( g_t^{(n)}(\beta) \right)^2 \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j} \right) + \frac{1}{2} \left( \frac{\partial g_t^{(n)2}(\beta)}{\partial \beta_i} \left( g_t^{(n)}(\beta) \right)^2 \frac{\partial g_t^{(n)2}(\beta)}{\partial \beta_j} \right) \right] = V_{ij}, \]

\( i, j = 1, \ldots, r \), where the matrix \( V = \left( V_{ij} \right)_{1 \leq i, j \leq r} \) is a strictly definite positive matrix;

\[ H_{2',7} \frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-d} \psi_{t \ell k}^{(n)}(\beta^0) \psi_{t+d,j,k+d}(\beta^0) g_{t-k}^{(n)2}(\beta^0) = O \left( \frac{1}{n} \right), \]

and

\[ \frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-d} \psi_{t \ell k}^{(n)}(\beta^0) \psi_{t+d,i,k+d}(\beta^0) \psi_{t+d,j,k+d}(\beta^0) g_{t-k}^{(n)4}(\beta^0) \kappa_{4,t-k} = O \left( \frac{1}{n} \right), \]

and where \( \kappa_{4,t} \) is the fourth-order cumulant of \( e_t \). Then, when \( n \to \infty \),

- there exists an estimator \( \hat{\beta}_n \) such that \( \hat{\beta}_n \to \beta^0 \) in probability;
- \( n^{1/2} (\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V^{-1} W V^{-1}) \) where there exists a matrix \( W \) whose elements are defined by (2.7).

**Appendix B. Alternative assumptions under a mixing condition**

In this appendix, we shall need that the processes satisfy a mixing condition. The definition we use, e.g. [23], proposed by [24] in the context of stationary processes is the \( \rho \)-mixing condition.

**Definition 3**

Let \( (w_s, t \in \mathbb{Z}) \) be a process (not necessarily stationary) of random variables defined on a probability space \( (\Omega, F, P) \). We say that the process is \( \rho \)-mixing, if there exists a sequence of positive real numbers \( \rho(d), d > 1 \), such that \( \rho(d) \to 0 \) as \( n \to \infty \), where

\[ \rho(d) = \sup_{t \in \mathbb{Z}} \sup_{U \in \mathcal{L}^2(F_{t+\infty}^d)} \sup_{V \in \mathcal{L}^2(F_{t+\infty}^d)} |\text{corr}(U, V)|, \]  

\( F_{t+\infty}^d \) is the \( \sigma \)-field spanned by \( (w_s, s \leq t) \), and \( F_{t+\infty}^d \) is the \( \sigma \)-field spanned by \( (w_s, s \geq t + d) \). Then \( \rho(d) \) is called the \( \rho \)-mixing coefficient of the process.

Of course, if the process is strictly stationary, the supremum over \( t \) disappears and the definition coincides with the standard definition. The definition easily extends in our case of a triangular array process since the \( \sigma \)-fields are generated by the innovations (and innovations in reverse time).

**Lemma 4**

Let \( (w_t, t \in \mathbb{Z}) \) be a process (not necessarily stationary) which satisfies the \( \rho \)-mixing condition. Let a random variable \( U \in \mathcal{L}^2(F_{t+\infty}^d) \) and a random variable \( V \in \mathcal{L}^2(F_{t+\infty}^d) \); then,
\[ \text{cov}(U, V) \leq \rho(d) \{ \text{var}(U) \text{var}(V) \}^{1/2}. \]  

(B.2)

This is obvious taking into account (B.1), see [25].

**Theorem 5**

Consider a pure autoregressive process under the assumptions of Theorem 2' except that \( H_{2,7} \) is replaced by \( H_{2,74} \):

\( H_{2,74} \) For \( \beta = \beta^0 \), let the process be \( \rho \)-mixing with mixing coefficient \( \rho(d) \) bounded by an exponentially decreasing function, such that \(|\rho(d)| < \rho^0\), with \( 0 < \rho < 1 \).

Then the results of Theorem 2' of [1] are still valid.

**Proof**

\( H_{2,7} \) is used to prove two assumptions, \( H_{2,3} \) and \( H_{2,5} \), of [1, Theorem 1'], but the former is more demanding. We have to show, see equation (A.13) of [1], that

\[
\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \text{cov}_{\beta^0} \left( \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} \{ g_t^{(n)}(\beta) \}^{-2} \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j} \{ g_t^{(n)}(\beta) \}^{-2} \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j} \right). \tag{B.3}
\]

We decompose the external sum in two sums, one for \( d = 1, \ldots, p \) and one for \( d = p + 1, \ldots, n - 1 \) and we will show that both sums are \( O(1/n) \). Using Cauchy-Schwarz inequality and the fact that the proof of [1, Theorem 2] has shown that

\[
E_{\beta^0} \left( \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} \{ g_t^{(n)}(\beta) \}^{-2} \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j} \right)^2
\]

is bounded, uniformly in \( t \), using only \( H_{2,1} - H_{2,6} \), the first sum is indeed \( O(1/n) \).

The general term of the second sum can be written \( \{ g_t^{(n)}(\beta^0) \}^{-2} \{ g_{t+d}^{(n)}(\beta^0) \}^{-2} h_{t,i,j}(\beta^0) \), where

\[
h_{t,i,j}(\beta^0) = \text{cov}_{\beta^0} \left( g_{t,i}^{(n)}(\beta) g_{t,j}^{(n)}(\beta), g_{t+d,i}^{(n)}(\beta) g_{t+d,j}^{(n)}(\beta) \right),
\]

and \( g_{t,i}^{(n)}(\beta) = \partial e_t^{(n)}(\beta) / \partial \beta_i \). Given (2.5), \( U = g_{t,i}^{(n)}(\beta^0) g_{t,j}^{(n)}(\beta^0) \in L^2(F_{t+i}^\infty) \) and, provided \( d > p \), \( V = g_{t+d,i}^{(n)}(\beta^0) g_{t+d,j}^{(n)}(\beta^0) \in L^2(F_{t+d-p}^\infty) \), for all \( t \) and all \( i, j \). Indeed the right hand sides have finite variances by application of Cauchy-Schwarz inequality, and using the fact that \( E_{\beta^0} \left( g_{t,i}^{(n)}(\beta) \right)^4 \leq m_2^2 (N_2 K^{1/2} + 3N_2) \sigma^4 \), using \( H_{2,1}, H_{2,3} \) and \( H_{2,5} \), uniformly in \( n \), see (A.9) of [1]. Also \( \{ g_t^{(n)}(\beta^0) \}^{-2} \{ g_{t+d}^{(n)}(\beta^0) \}^{-2} \leq m^{-2} \), using \( H_{2,3} \). By \( H_{2,74} \) and using Lemma 4, \( h_{t,i,j}(\beta^0) \) is bounded by

\[
2 \rho(d-p) \left\{ E_{\beta^0} \left( g_{t,i}^{(n)}(\beta) g_{t,j}^{(n)}(\beta) \right)^{1/2} \right\}^{2} \left\{ E_{\beta^0} \left( g_{t+d,i}^{(n)}(\beta) g_{t+d,j}^{(n)}(\beta) \right)^{1/2} \right\}^{1/2}.
\]

Hence that expression is uniformly bounded, with respect to \( t \). Since \( H_{2,74} \) implies

\[
\sum_{d=p+1}^{n-1} |\rho(d-p)| \leq \sum_{d=p+1}^{n-1} \rho^{d-p} < \infty,
\]

hence (B.3) is \( O(1/n) \). The argument is similar for checking \( H_{3,5} \) but the expression to consider is
\[ \frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \text{cov}_{\beta_0} \left( K_t^{(n)j} (\beta) \frac{\partial e_t^{(n)} (\beta)}{\partial \beta_i}, K_{t+d}^{(n)j} (\beta) \frac{\partial e_{t+d}^{(n)} (\beta)}{\partial \beta_j} \right), \]

where

\[ K_t^{(n)j} (\beta) = 4 \frac{E \{ e_t^{(n)} (\beta) \}^3}{\sigma^4 g_t^{(n)6} (\beta)} \left( \frac{\partial g_t^{(n)2} (\beta)}{\partial \beta_i} \right). \]

The proof continues like in Theorem 2 and 2’ of AM, using a weak law of large numbers for a mixingale array [26] and referring to Theorems 1 and 1’ of AM, which make use of a central limit theorem for a martingale difference array [27] modified with a Lyapounov condition.

**Remarks**

Strong mixing should be a nice requirement. However, on the one hand, even stationary AR(1) processes can be non-strong mixing and, on the other hand, the covariance inequalities which are implied are not applicable in our context without stronger assumptions.

In an earlier version uniformly strong mixing or \( \varphi \)-mixing was used. But, as [23] points out, for Gaussian stationary processes, \( \varphi \)-mixing implies \( m \)-dependence for some \( m \). So the AR processes should behave like MA processes, leaving just white noise. Finally, we opted for \( \rho \)-mixing. There are results for stationary linear processes [28] and for ARMA processes [29], but apparently none for the non-stationary processes considered here. In practice, even if \( \rho \)-mixing condition is more appealing, checking \( H_{2r,7}^A \) is more challenging than checking \( H_{2r,7} \). For instance, in [1, Example 3] it is possible to check \( H_{2r,7}^A \).

**Appendix C. tdAR(2) Monte Carlo simulations**

The purpose of this appendix is to illustrate the procedure described in Section 2 on further, more stressing, simulation results than in [1]. In [1], Monte-Carlo simulations had been shown for nonstationary AR(1) and MA(1) models, with a time-dependent coefficient and a time-dependent innovation variance, for several series lengths between 25 and 400, in order to show convergence in an empirical way. The purpose was mainly to illustrate the theoretical results for these models, particularly the derivation of the asymptotic standard errors, and investigate the sensibility of the innovation distribution on the conclusions.

Here, we consider tdAR(2) models described by (2.1) and (2.9) in nearly the same setup as in AM except that the innovation variance is assumed constant but the series are generated using a process with linearly time-dependent coefficients, not stationary processes. Only Gaussian innovations are simulated, so the inverse of \( V \) in (2.8) is used in order to produce standard errors. Since we are only interested in autoregressive models, it does not seem necessary to compare the exact maximum likelihood and the approximate or conditional maximum likelihood methods. Numerical optimization was used.
In the parametrization in (2.9), the two coefficients $\phi_{t1}^{(n)}$ and $\phi_{t2}^{(n)}$ vary with respect to time between $-0.5$ and $0.5$ for the former, and between $-0.9$ and $0.5$ for the latter. If we consider the roots of the polynomials $1 - \phi_{t1}^{(n)} z - \phi_{t2}^{(n)} z^2$, for the different $t$, that means they are complex till well after the middle of the series, where their modulus is large (about 8) whereas it is close to 1 at the beginning and the smallest root is equal to 1 at the end of the series, so at the causality frontier. This is illustrated by Figure 7. It will therefore not be surprising if the empirical results are not as bright as in [1]. A plot of a sample series is shown in Figure 8 which illustrates the behavior: complex roots at the beginning correspond to oscillations, and a root close to 1 at the end corresponds to strong positive autocorrelation.

Figure 7. Variations of $\phi_{t1}^{(n)}$ (horizontal) and $\phi_{t2}^{(n)}$ (vertical) with respect to time $t$ for $n = 50$; inside the triangle corresponds to the causality condition and the curve separates complex roots (below) from real roots (above).

Figure 8. Plot of the data for one of the simulated tdAR(2) series for $n = 400$. 
Table 3. Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and averages across simulations of estimated standard errors $\phi_1^t$, $\phi_1^s$, $\phi_2^t$, and $\phi_2^s$ for the tdAR(2) model described above, for $n = 400$; 999 replications (out of 1000).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Average</th>
<th>Standard deviation</th>
<th>Average of standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_1^t$</td>
<td>0.0</td>
<td>0.007306</td>
<td>0.050587</td>
<td>0.043869</td>
</tr>
<tr>
<td>$\phi_1^s$</td>
<td>0.002551</td>
<td>0.002422</td>
<td>0.000322</td>
<td>0.000333</td>
</tr>
<tr>
<td>$\phi_2^t$</td>
<td>-0.2</td>
<td>-0.193960</td>
<td>0.048853</td>
<td>0.043537</td>
</tr>
<tr>
<td>$\phi_2^s$</td>
<td>0.003571</td>
<td>0.003421</td>
<td>0.000332</td>
<td>0.000325</td>
</tr>
</tbody>
</table>

Table 4. Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and averages across simulations of estimated standard errors $\phi_1^t$, $\phi_1^s$, $\phi_2^t$, and $\phi_2^s$ for the tdAR(2) model described above, for $n = 50$; 934 replications (out of 1000).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Average</th>
<th>Standard deviation</th>
<th>Average of standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_1^t$</td>
<td>0.0</td>
<td>0.01419</td>
<td>0.14260</td>
<td>0.13620</td>
</tr>
<tr>
<td>$\phi_1^s$</td>
<td>0.020408</td>
<td>0.01640</td>
<td>0.00900</td>
<td>0.00957</td>
</tr>
<tr>
<td>$\phi_2^t$</td>
<td>-0.2</td>
<td>-0.19510</td>
<td>0.13972</td>
<td>0.12436</td>
</tr>
<tr>
<td>$\phi_2^s$</td>
<td>0.028571</td>
<td>0.02355</td>
<td>0.00852</td>
<td>0.00749</td>
</tr>
</tbody>
</table>

Table 3 for $n = 400$ shows that the estimates are close to the true values of the parameters and that the asymptotic standard errors are well estimated, since the average of these estimates agrees more or less with the empirical standard deviation. When $n = 50$, note however, by comparison of the last two columns of Table 4 that the asymptotic standard errors are not badly estimated even if a larger proportion of fits have failed. We have seen in Section 5 an example which is still more extreme.

References


