AR MODELS WITH TIME-DEPENDENT COEFFICIENTS - A COMPARISON BETWEEN SEVERAL APPROACHES[†]

Rajae AZRAK [‡] and Guy MÉLARD [§]

September 22, 2011

Abstract

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

In this paper, we focus on autoregressive (AR) models and consider our theory in that case. We provide also an alternative theory for AR(p) processes which relies on a ρ-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 AR(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 examples on real series and by simulation results that show that the asymptotic theory can be used even for short series (less than 50 observations).

[*] *AMS 1991 subject classification.* Primary 62M15; secondary 62M05.
Key words and phrases. Nonstationary process, time series, time-dependent model.

[†] Correspondence to be sent to the second author.

We thank those who have made comments to our previous paper, including Christian Francq, Marc Hallin, and mainly Rainer Dahlhaus and Denis Bosq, as well as its anonymous referees.

[‡] Université Mohammed V-Souissi, Faculté des Sciences Juridiques Economiques et Sociales, Salé, Morocco (E-mail address: razrak96@yahoo.fr).

[§] 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.

1. Introduction

Autoregressive-moving average (ARMA) models with (td) time-dependent coefficients and marginally heteroscedastic innovation variance provide a natural alternative to stationary ARMA models. Several theories have been developed in the last ten 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 $h_t^{(n)} > 0$ 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 $\sigma > 0$. The model is defined by

$$w_t = \phi_t^{(n)} w_{t-1} + \{h_t^{(n)}\}^{1/2} e_t. \quad (1.1)$$

The coefficient $\phi_t^{(n)}$ and $\sigma_t^{(n)} = \{h_t^{(n)}\}^{1/2}\sigma$ depend t and n, but also on parameters. Consider a triangular sequence of observations $w^{(n)} = (w_1^{(n)}, w_2^{(n)}, ..., w_n^{(n)})$ of the process.

The AR(1) process with a time-dependent coefficient has been considered by Wegman (1974), Kwoun and Yajima (1986), Tjøstheim (1984). Hamdoune (1995) and Dahlhaus (1997) have extended the results to autoregressive processes of order p. Azrak and Mélard (2006), denoted AM, and Bibi and Francq (2003), denoted BF, have considered tdARMA processes. 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. We will therefore compare these approaches on autoregressive models.

Two approaches can be sketched for asymptotics within nonstationary processes, see Dahlhaus (1996b). Approach 1 consists in analyzing the behavior of the process when n tends to infinity. That assumes some generating mechanism in the background which remains the same over time. Two examples can be mentioned: processes with periodically changing coefficients and cointegrated processes. It is in that context that BF have established asymptotic properties for parameters 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 mostly also 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 the papers of Dahlhaus (1997) that we will briefly summarize now. First, there is an assumption of local stationarity which 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 $\sigma_t^{(n)} = \tilde{\sigma}_{t/n}$, where $\tilde{\phi}_u$ and $\tilde{\sigma}_u$, $0 \leq u \leq 1$, 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

$$w_t^{(n)} = \tilde{\phi}_{t/n} w_{t-1}^{(n)} + \tilde{\sigma}_{t/n} e_t. \quad (1.2)$$

As a consequence, the assumptions made by Dahlhaus in the locally stationary theory are quite different from those of AM and BF, for example because of the different nature of the asymptotics. Our 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. This is illustrated in Section 3 by examples on real series and by simulation results. In Section 4, we provide an alternative theory for tdAR(p) processes which relies on a ρ-mixing property. In Section 5, we compare the Dahlhaus theory for locally stationary processes with our own theory. This is partly explained thanks to examples. The differences in the basic assumptions are emphasized. Similarly, in Section 6, a comparison is presented with the BF (Bibi and Francq, 2003) approach.

2. Autoregressive processes with time-dependent coefficients

Let us consider the AM theory[1] in the special case of AR(p) processes. We want to see if simpler conditions can be derived for the treatment of pure autoregressive processes.

We consider a triangular array of random variables $w = (w_t^{(n)}, t = 1, ..., n, n \in \mathbb{N})$ defined on a probability space (Ω, F, P_β), with values in $\mathbb{R}$, whose distribution depends on a vector $\beta = (\beta_1, ..., \beta_r)$ of unknown parameters to be estimated, with β lying in an open set B of an Euclidean space $\mathbb{R}^r$. The true value of β is denoted by β^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

$$w_t^{(n)} = \sum_{k=1}^{p} \phi_{tk}^{(n)} w_{t-k}^{(n)} + \{h_t^{(n)}\}^{1/2} e_t, \tag{2.1}$$

where $(e_t, t \in \mathbb{N})$ are independent random variables with mean 0, standard deviation $\sigma > 0$, and 4-th order cumulant κ_{4t}, and $\{h_t^{(n)}\}^{1/2}$ is a deterministic strictly positive function of time. We denote $\sigma_t^{(n)} = \sigma\{h_t^{(n)}\}^{1/2}$, the innovation standard deviation. The initial values $w_t, t < 1$, are supposed to be equal to zero.

More precisely, the r-dimensional vector β contains all the parameters to be estimated, those in $\phi_{tk}^{(n)}, k = 1, ..., p$, and those in $\{\sigma_t^{(n)}\}^2 = \sigma^2 h_t^{(n)} > 0$ hence in $h_t^{(n)}$ but not the scale factor σ^2 which is estimated separately. Because of the possible dependency of $\phi_{tk}^{(n)}$ and $h_t^{(n)}$ with respect to n, we consider a triangular sequence $(w_t^{(n)}, t = 1, ..., n)$ which is solution of (2.1). Let $\phi_{tk}^{(n)}(\beta)$ be

[1] We have improved the presentation in the light of the BF paper, especially by making it clear that some coefficients depend on both β and β^0. The notations for the innovations are also changed to emphasize that F_t doesn't depend on n.

the parametric coefficient with $\phi_{tk}^{(n)}(\beta^0) = \phi_{tk}^{(n)}$, and similarly $h_t^{(n)}(\beta^0) = h_t^{(n)}$. Let $e_t^{(n)}(\beta)$ be the residual for a given β:

$$e_t^{(n)}(\beta) = w_t^{(n)} - \sum_{k=1}^{p} \phi_{tk}^{(n)}(\beta) w_{t-k}^{(n)}. \tag{2.2}$$

Note that $e_t^{(n)}(\beta^0) = \{h_t^{(n)}(\beta^0)\}^{1/2} e_{t-k}$

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:

$$w_t^{(n)} = \sum_{k=0}^{t-1} \psi_{tk}^{(n)}(\beta^0)\{h_{t-k}^{(n)}(\beta^0)\}^{1/2} e_{t-k}, \tag{2.3}$$

(see Azrak and Mélard, 2006, for a recurrence formula).

Let F_t be the σ-field generated by the $(w_s^{(n)}, s \leq t)$, hence by $(e_s, s \leq t)$, which explains why a superscript $^{(n)}$ is missing, and $F_0 = \{\emptyset, \Omega\}$. To simplify the presentation, we denote $E_{\beta^0}(\cdot(\beta)) = \{E_\beta(\cdot(\beta))\}_{\beta=\beta^0}$ and similarly $\mathrm{var}_{\beta^0}(\cdot)$ and $\mathrm{cov}_{\beta^0}(\cdot)$.

We are interested in the Gaussian quasi-maximum likelihood estimator

$$\hat{\beta}^{(n)} = \mathrm{argmin}_{\beta \in \mathbb{R}^r} \sum_{t=1}^{n} \Big[\log\{\sigma_t^{(n)}(\beta)\}^2 + \Big(\frac{e_t^{(n)}(\beta)}{\sigma_t^{(n)}(\beta)} \Big)^2 \Big]. \tag{2.4}$$

Denote $\alpha_t^{(n)}(\beta)$ the expression between square brackets in (2.4). Note that the first term of $\alpha_t^{(n)}(\beta)$ 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.

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

$$\frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} = -\sum_{k=1}^{p} \frac{\partial \phi_{tk}^{(n)}(\beta)}{\partial \beta_i} w_{t-k}^{(n)}, \tag{2.5}$$

$i = 1, ..., r$. It will be convenient to write it as a pure moving average

$$\frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} = -\sum_{k=1}^{t-1} \psi_{tik}^{(n)}(\beta, \beta^0)\{h_{t-k}^{(n)}(\beta^0)\}^{1/2} e_{t-k}, \tag{2.6}$$

and similarly for the second and third order derivatives

$$\frac{\partial^2 e_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j} = -\sum_{k=1}^{t-1} \psi_{tijk}^{(n)}(\beta, \beta^0)\{h_{t-k}^{(n)}(\beta^0)\}^{1/2} e_{t-k}, \tag{2.7}$$

$$\frac{\partial^3 e_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} = -\sum_{k=1}^{t-1} \psi_{tijlk}^{(n)}(\beta, \beta^0)\{h_{t-k}^{(n)}(\beta^0)\}^{1/2} e_{t-k}, \tag{2.8}$$

for $i, j, l = 1, ..., r$, where the coefficients $\psi_{tik}^{(n)}(\beta, \beta^0)$, $\psi_{tijk}^{(n)}(\beta, \beta^0)$ and $\psi_{tijlk}^{(n)}(\beta, \beta^0)$ are obtained by the following relations

$$\psi_{tik}^{(n)}(\beta, \beta^0) = \sum_{u=1}^{k} \frac{\partial \phi_{tu}^{(n)}(\beta)}{\partial \beta_i} \psi_{t-u,k-u}^{(n)}(\beta^0), \quad \psi_{tijk}^{(n)}(\beta, \beta^0) = \sum_{u=1}^{k} \frac{\partial^2 \phi_{tu}^{(n)}(\beta)}{\partial \beta_i \partial \beta_j} \psi_{t-u,k-u}^{(n)}(\beta^0),$$

$$\psi_{tijlk}^{(n)}(\beta, \beta^0) = \sum_{u=1}^{k} \frac{\partial^3 \phi_{tu}^{(n)}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} \psi_{t-u,k-u}^{(n)}(\beta^0).$$

Under all the assumptions of Theorem 2' of Azrak and Mélard (2006), see Appendix 1, the estimator $\hat{\beta}^{(n)}$ converges in probability to β^0 and, furthermore, $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0) \xrightarrow{L} N(0, V(\beta^0)^{-1} W(\beta^0) V(\beta^0)^{-1})$ when $n \to \infty$, where, with ^{T} denoting transposition,

$$W(\beta^0) = \lim_{n \to \infty} \frac{1}{4n} \sum_{t=1}^{n} E_{\beta^0} \left(\frac{\partial \alpha_t^{(n)}(\beta)}{\partial \beta} \frac{\partial \alpha_t^{(n)}(\beta)}{\partial \beta^T} \right), \tag{2.9}$$

and

$$V(\beta^0) = \lim_{n \to \infty} \frac{1}{2n} \sum_{t=1}^{n} E_{\beta^0} \left(\frac{\partial^2 \alpha_t^{(n)}(\beta)}{\partial \beta \partial \beta^T} / F_{t-1} \right). \tag{2.10}$$

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)

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

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

$$\psi_{tik}^{(n)}(\beta, \beta^0) = \frac{\partial \phi_t^{(n)}(\beta)}{\partial \beta_i} \psi_{t-l,k-1}^{(n)}(\beta^0) = \frac{\partial \phi_t^{(n)}(\beta)}{\partial \beta_i} \prod_{l=1}^{k-1} \phi_{t-l}^{(n)}(\beta^0),$$

$$\psi_{tijk}^{(n)}(\beta, \beta^0) = \frac{\partial^2 \phi_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j} \psi_{t-l,k-1}^{(n)}(\beta^0) = \frac{\partial^2 \phi_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j} \prod_{l=1}^{k-1} \phi_{t-l}^{(n)}(\beta^0),$$

and an analogous expression for order 3. The following is an application of Theorem 2' of AM.

Theorem 2'A

Consider a tdAR(1) process defined by (1.1) under the assumptions of Theorem 2' except that $H_{2'.1}$ is replaced by $H_{2'.1A}$. Let us suppose that there exist constants C, Ψ ($0 < \Psi < 1$), M_1, M_2, and M_3 such that the following inequalities hold:

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

$$\left|\left\{\frac{\partial^2\phi_t^{(n)}(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^0}\right| < M_2, \quad \left|\left\{\frac{\partial^3\phi_t^{(n)}(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_l}\right\}_{\beta=\beta^0}\right| < M_3.$$

then the results of Theorem 2' of Azrak and Mélard (2006) are still valid.

Proof

Let us show the first of the inequalities in $H_{2'.1}$ since the others are similar. Consider for $\nu = 1, ..., t-1$

$$\sum_{k=\nu}^{t-1}\{\psi_{tik}^{(n)}(\beta^0,\beta^0)\}^2 = \Big(\frac{\partial\phi_t^{(n)}(\beta)}{\partial\beta_i}\Big)^2_{\beta=\beta^0}\sum_{k=\nu}^{t-1}\{\psi_{t-l,k-1}^{(n)}(\beta^0)\}^2 \leq \frac{M_1^2C^2(\Psi^2)^{\nu-1}}{1-\Psi^2},$$

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

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 Azrak and Mélard (2006) 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.

3. Monte Carlo simulations and examples

The purpose of this section is to illustrate the procedure described in the previous section on further simulation results than in AM and then on real time series.

In AM, 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 have considered tdAR(2) models 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. Since we are only interested in autoregressive models, it doesn't seem necessary to compare the exact maximum likelihood and the approximate or conditional maximum likelihood methods. Numerical optimisation was used. Although the simulations in AM had not really stressed the method, we will

Figure 1: Plot of the data for one of the simulated tdAR(2) series for $n = 400$.

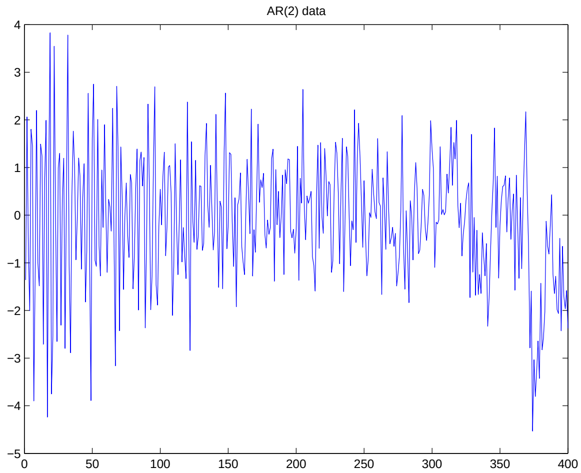


Table 1: Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and averages across simulations of estimated standard errors $\phi_1^{'}$, $\phi_1^{''}$, $\phi_2^{'}$, and $\phi_2^{''}$ for the tdAR(2) model described above, for $n = 400$; 999 replications (out of 1000).

Parameter true value	average	standard deviation	average of standard error
$\phi_1^{'} = 0.0$	0.007306	0.050587	0.043869
$\phi_1^{''} = 0.002551$	0.002422	0.000322	0.000333
$\phi_2^{'} = -0.2$	-0.193960	0.048853	0.043537
$\phi_2^{''} = 0.003571$	0.003421	0.000332	0.000325

treat a case where the coefficients $\phi_{t1}^{(n)}$ and $\phi_{t2}^{(n)}$ vary in an extreme way. The parameterisation used is

$$\phi_{tk}^{(n)}(\beta) = \phi_k^{'} + \frac{1}{n-1}(t - \frac{n+1}{2})\phi_k^{''}, \qquad k = 1, 2. \tag{3.1}$$

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 $1 - \phi_{t1}^{(n)} z - \phi_{t2}^{(n)} z^2$, 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. A plot of a sample series is shown in Figure 1 which illustrates that behavior. Table 1 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.

The fact that the method is stressed is demonstrated by Figure 2 which shows, when $n = 50$, the variations of the two coefficients, respectively $\phi_{t1}^{(n)}$ and $\phi_{t2}^{(n)}$ in function of time. It is therefore not surprising if the empirical results are not as bright as in AM. Note however, by comparison of the last two columns of Table 2, that the asymptotic standard errors are not badly estimated even if a larger proportion of fits have failed. We will see in Section 5 an example which is still more extreme.

Figure 2: Variations of $\phi_{t1}^{(n)}$ and $\phi_{t2}^{(n)}$ with respect to time t for $n = 50$.

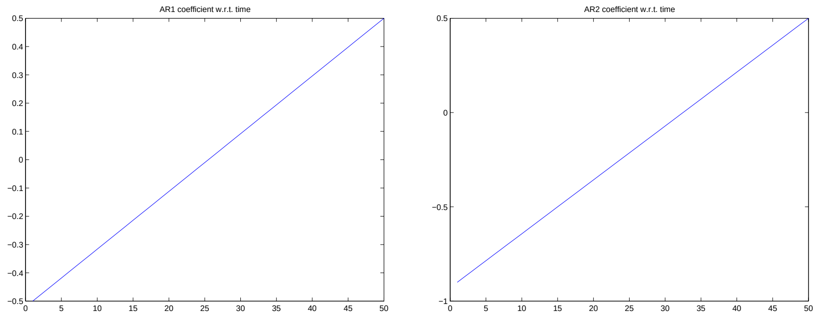


Table 2: Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and averages across simulations of estimated standard errors $\phi_1^{'}$, $\phi_1^{''}$, $\phi_2^{'}$, and $\phi_2^{''}$ for the tdAR(2) model described above, for $n = 50$; 934 replications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

Parameter true value	average	standard deviation	average of standard error
$\phi_1^{'} = 0.0$	0.01419	0.14260	0.13620
$\phi_1^{''} = 0.020408$	0.01640	0.00900	0.00957
$\phi_2^{'} = -0.2$	−0.19510	0.13972	0.12436
$\phi_2^{''} = 0.028571$	0.02355	0.00852	0.00749

In AM, we have already fitted ARMA models with time-dependent coefficients and a time-dependent innovation standard deviation to series from the Box *et al.* (1994). Here we have taken a dataset of indices, with 1985 as basis, for the monthly Belgian industrial production by the various branches of activity, 26 in all. The data cover the period from January 1985 to December 1994 but the last year is used to compute ex-post forecasts and the mean absolute percentage error (MAPE). We have applied Mélard and Pasteels (2000) procedure to fit autoregressive integrated or ARI(p, P) models to the series of 108 observations. The model can include ordinary and/or seasonal differences, transformations and interventions (additive or on the differenced series) which are not detailed here. We have retained the 20 series out of 26 for which pure autoregressive models were obtained. In the Box-Jenkins tradition, the autoregressive operator is a product of the ordinary polynomial of degree p and of the seasonal polynomial of degree $12P$. For example, for an ARI(2,1) model, the operator is written

$$(1-\phi_1 B-\phi_2 B^2)(1-\Phi_1 B^{12}) = (1-\phi_1 B-\phi_2 B^2-\Phi_1 B^{12}+\phi_1\Phi_1 B^{13}+\phi_2\Phi_1 B^{14}). \tag{3.2}$$

The fit is characterized by the value of the SBIC criterion. For using time-dependent ARI, or tdARI, models, we have then added slope parameters for each of the existing coefficients. This is not possible in multiplicative form so we have considered the developed form in the right hand side. We have often reduced the number of parameters by using, for example, the polynomial $(1-\phi_{t1}^{(n)}B-\phi_{t2}^{(n)}B^2-\phi_{t,12}^{(n)}B^{12}-\phi_{t,13}^{(n)}B^{13}+\phi_2^{'}\phi_{12}^{'}B^{14})$, where $\phi_{t1}^{(n)}$, and $\phi_{t2}^{(n)}$ and $\phi_{t,12}^{(n)}$ are like in (3.1) and

$$\phi_{t,13}^{(n)} = -\phi_1^{'}\phi_{12}^{'} + \frac{1}{n-1}(t-\frac{n+1}{2})\phi_{13}^{''},$$

with 7 parameters instead of the full form $= (1-\phi_{t1}^{(n)}B-\phi_{t2}^{(n)}B^2-\phi_{t,12}^{(n)}B^{12}-\phi_{t,13}^{(n)}B^{13}-\phi_{t,14}^{(n)}B^{14})$ that would involve 10 parameters in all. The models have therefore coefficients which are (nearly all) linear functions of time. Table 3 shows the main results, including those tdAR coefficients for which the test $\phi_k^{''}=0$ is rejected at the 5% level, and the corresponding t-statistic. Of course some nonsignificant parameters could have been omitted in order to reduce SBIC but this was not attempted. Nevertheless, even if tdARI models are not systematically better, they often produce better forecasts, and sometimes show a better fit or at least some statistically significant slope parameters at the 5 % level. Van Bellegem and von Sachs (2004) had already shown the usefulness of a time-dependent variance. Of course, an automatic selection procedure like the one exposed in Van Bellegem and Dahlhaus (2006) is possible.

Table 3: For each branch of the economy, we give the orders (p, P) of the model, SBIC and MAPE for the raw ARI model and for the tdARI model (results in bold are better), the values of the statistically significant slopes, where ARk denotes ϕ''_k, and the corresponding t-value for testing a parameter equal to 0.

Branch (Name)	orders (p,P)	ARI SBIC / *MAPE*	tdARI SBIC / *MAPE*	tdARI parameter	t-value
Food, beverages	(3,0)	**655**	669	none	
(ALIBOR)		*4.0*	***3.9***		
Other extraction of minerals	(2,0)	**904**	910	AR2	2.8
(AUEXTR)		*10.5*	***9.4***		
Wood-processing, furniture	(3,1)	**778**	792	AR12	2.3
(BOIME)		*6.4*	***5.9***		
Hosiery	(2,1)	702	**694**	AR2	4.7
(BONNE)		*9.1*	***5.5***	AR12	8.9
				AR13	-6.5
Commerce	(3,1)	**553**	556	AR12	3.6
(COMME)		***1.7***	*2.2*		
Construction	(3,0)	**827**	830	AR2	2.8
(CONST)		***8.0***	*23.6*		
Petrol derivatives	(1,0)	**888**	891	none	
(DERPE)		*5.2*	*5.2*		
Petrol distribution	(1,1)	**941**	951	AR12	−5.3
(DISPE)		*10.8*	***10.7***		
Metal processing	(2,0)	682	**681**	AR2	−2.1
(FABME)		***6.6***	*7.2*		
Manufacture of textiles	(2,0)	**739**	749	none	
(FILAT)		***5.5***	*6.4*		
Gas production/distribution	(3,1)	897	**891**	AR2	−2.1
(GAZ)		*2.1*	***1.9***	AR3	−4.3
Construction materials	(1,1)	809	**807**	AR13	−3.3
(MATCO)		***4.3***	*5.7*		
Non-metallic manufacturing	(2,1)	**826**	830	AR12	3.5
(NONFE)		***5.5***	*9.4*	AR13	2.0
Paper/paperboard industry	(3,1)	**746**	757	AR2	−2.1
(PAPCA)		*5.9*	***5.8***		
Iron and steel	(3,1)	**833**	837	AR1	3.8
(SIDER)		***5.9***	*9.2*	AR12	3.5
Manufacture of tobacco	(3,1)	**778**	791	AR3	2.9
(TABAC)		***10.9***	*13.2*	AR12	5.4
Aviation	(3,1)	**746**	753	AR12	4.5
(TRAER)		*8.8*	*8.8*		
Maritime transport	(2,1)	**732**	740	AR12	2.3
(TRMAR)		*3.6*	***2.6***	AR13	3.6
Land transport	(0,1)	854	**849**	AR12	−2.6
(TRTER)		***12.1***	*12.5*		
Manufacture of clothing	(2,0)	**757**	767	none	
(VETEM)		*26.7*	*26.7*		

4. Alternative assumptions under a mixing condition

In this section, we shall need that the processes satisfy a mixing condition. The definition we use, e.g. Bosq (1998), proposed by Kolmogorov and Rozanov (1960) in the context of stationary processes is the ρ-mixing condition.

Definition 2

Let $(w_t, t \in Z)$ be a process (not necessarily stationary) of random variables defined on a probability space (Ω, F, P). We say that the process is ρ-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

$$\rho(d) = \sup_{t\in\mathbb{Z}} \sup_{\substack{U \in \mathcal{L}^2(F_{-\infty}^t) \\ V \in \mathcal{L}^2(F_{t+d}^\infty)}} |\mathrm{corr}(U, V)|, \tag{4.1}$$

$F_{-\infty}^t$ is the σ-field spanned by $(w_s, s \leq t)$, and F_{t+d}^∞ is the σ-field spanned by $(w_s, s \geq t + d)$. Then $\rho(d)$ is called the ρ-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.

Lemma 1

Let $(w_t, t \in \mathbb{Z})$ be a process (not necessarily stationary) which satisfies the ρ-mixing condition. Let a random variable $U \in \mathcal{L}^2(F_{-\infty}^t)$ and a random variable $V \in \mathcal{L}^2(F_{t+d}^\infty)$; then,

$$\mathrm{cov}(U, V) \leq \rho(d)\{\mathrm{var}(U)\mathrm{var}(V)\}^{1/2}. \tag{4.2}$$

This is obvious taking into account (4.1), see Rio (2000).

Theorem 3A

Consider a pure autoregressive process under the assumptions of Theorem 2' except that $H_{2'.7}$ is replaced by $H_{2'.7A}$:
$H_{2'.7A}$ For $\beta = \beta^0$, let the process be ρ-mixing with mixing coefficient $\rho(d)$ bounded by an exponentially decreasing function, such that $|\rho(d)| < \rho^d$, with $0 < \rho < 1$.
Then the results of Theorem 2' of AM are still valid.

Proof

$H_{2'.7}$ is used to prove two assumptions, $H_{1'.3}$ and $H_{1'.5}$, of Theorem 1' of AM, but the former is more demanding. We have to show, see equation (A1.13) there, that

$$\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \mathrm{cov}_{\beta^0} \left(\frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} \{h_t^{(n)}(\beta)\}^{-1} \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j}, \frac{\partial e_{t+d}^{(n)}(\beta)}{\partial \beta_i} \{h_{t+d}^{(n)}(\beta)\}^{-1} \frac{\partial e_{t+d}^{(n)}(\beta)}{\partial \beta_j} \right). \tag{4.3}$$

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

$$\mathrm{E}_{\beta^0} \left(\frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} h_t^{(n)-1}(\beta) \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j} \right)^2$$

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 $\{h_t^{(n)}(\beta^0)\}^{-1}\{h_{t+d}^{(n)}(\beta^0)\}^{-1} H_{t,i,j,d}^{(n)}(\beta^0)$, where

$$H_{t,i,j,d}^{(n)}(\beta^0) = \mathrm{cov}_{\beta^0} \left(G_{t,i}^{(n)}(\beta) G_{t,j}^{(n)}(\beta), G_{t+d,i}^{(n)}(\beta) G_{t+d,j}^{(n)}(\beta) \right),$$

and $G_{ti}^{(n)}(\beta) = \partial e_t^{(n)}(\beta)/\partial \beta_i$. Given (2.5), $U = G_{t,i}^{(n)}(\beta^0) G_{t,j}^{(n)}(\beta^0) \in \mathcal{L}^2(F_{-\infty}^t)$ and, provided $d > p$, $V = G_{t+d,i}^{(n)}(\beta^0) G_{t+d,j}^{(n)}(\beta^0) \in \mathcal{L}^2(F_{t+d-p}^{\infty})$, 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 $\mathrm{E}_{\beta^0} \left(G_{t,i}^{(n)}(\beta) \right)^4 \leq m_1^2 (N_2 K^{1/2} + 3N_2)\sigma^4$, using $H_{2'.1}$, $H_{2'.3}$ and $H_{2'.5}$, uniformly in n, see (A1.9) of AM. Also $h_t^{(n)}(\beta^0)^{-1} h_{t+d}^{(n)}(\beta^0)^{-1} \leq m^{-2}$, using $H_{2'.3}$. By $H_{2'.7A}$ and using Lemma 1, $H_{t,i,j,d}^{(n)}(\beta^0)$ is bounded by

$$2\rho(d-p) \left\{ \mathrm{E}_{\beta^0} \left(G_{t,i}^{(n)}(\beta) G_{t,j}^{(n)}(\beta) \right)^2 \right\}^{1/2} \left\{ \mathrm{E}_{\beta^0} \left(G_{t+d,i}^{(n)}(\beta) G_{t+d,j}^{(n)}(\beta) \right)^2 \right\}^{1/2}.$$

Hence that expression is uniformly bounded, with respect to t. Since $H_{2'.7A}$ implies $\sum_{d=p+1}^{n-1} |\rho(d-p)| \leq \sum_{d=p+1}^{n-1} \rho^{d-p} < \infty$, hence (4.3) is $O(1/n)$. The argument is similar for checking $H_{1'.5}$ but the expression to consider is

$$\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \mathrm{cov}_{\beta^0} \left(K_t^{(n)i}(\beta) \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i}, K_{t+d}^{(n)j}(\beta) \frac{\partial e_{t+d}^{(n)}(\beta)}{\partial \beta_j} \right),$$

where

$$K_t^{(n)i}(\beta) = 4 \frac{E\{e_t^{(n)}(\beta)\}^3}{\sigma^4 (h_t^{(n)}(\beta))^3} \left\{ \frac{\partial h_t^{(n)}(\beta)}{\partial \beta_i} \right\}.$$

The proof continues like in Theorem 2 and 2' of AM, using a weak law of large numbers for a mixingale array (Andrews, 1988) and referring to Theorems 1 and 1' of AM, which make use of a central limit theorem for a martingale difference array (Hall and Heyde, 1980) modified with a Lyapounov condition.

Remark

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 φ-mixing was used but, as Bosq (1998) points out, for Gaussian stationary processes, φ-mixing implies m-dependence for some m so, the AR processes should behave like MA processes, leaving just white noise. Finally, we opted for ρ-mixing. There are results for stationary linear processes, Chanda (1974), and for ARMA processes, Pham and Tran (1985), but apparently none for the non-stationary processes considered here. In practice, even if the statement of Theorem 3A is more appealing, checking $H_{2'.7A}$ is more challenging than checking $H_{2'.7}$. For instance, in Example 3 of AM, we were able to check $H_{2'.7}$.

5. A comparison with 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, that means on a sequence of stationary processes, based on a stochastic integral representation

$$w_t^{(n)} = \int_{-\pi}^{\pi} exp(i\lambda t) A_t^{(n)}(\lambda) \mathrm{d}\xi(\lambda), \tag{5.1}$$

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 ϕ_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. Note that when ARMA processes with time-dependent coefficients are considered in AM, polynomials in terms of the lag operator have no special meaning, neither their roots.

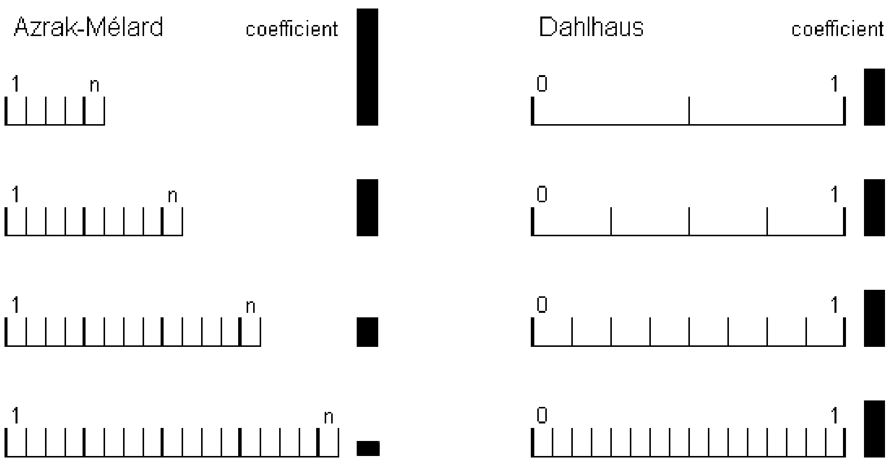
The estimation method is based either on a spectral approach or on a Whittle approximation of the Gaussian likelihood. Dahlhaus (1996a) also sketches an exact maximum likelihood estimation method like the one used here based on Mélard (1982) or Azrak and Mélard (1998).

As mentioned above, Dahlhaus approach of doing asymptotics relies on rescaling time t in $u = t/n$. That doesn't 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 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 (3.1), since $t/(n-1)$ is close to t/n for large n. Note however that Example 1 of AM is not in that class of processes. More generally, processes where the coefficients are a periodic function of t with a fixed period s, for example, 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.

Dependency with respect to u 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 Bibi and Francq (2003), accept discrete values of the coefficients with respect to time, without requiring a slow variation. They make instead assumptions of differentiability on dependency with respect to the parameters.

Another point of discussion is as follows. In order to handle economic and social data with an annual seasonality, Box *et al.* (1994) have proposed the so-called seasonal processes: 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. See (3.2) for example. Although series generated by these processes are not periodic, with suitably initial values, they can show a pseudo-periodic behavior with period s. The same objection stated above about cyclically time-dependent coefficients seems valid for not using such processes in the context of time rescaling. Let us consider such ARMA processes with time-dependent coefficients, for example an AR(12) defined by the equation $y_t = \phi_t^{(n)}(\beta)y_{t-12} + e_t$, 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.

Figure 3: Schematic presentation on how to interpret asymptotics in AM and Dahlhaus theories (see text for details).

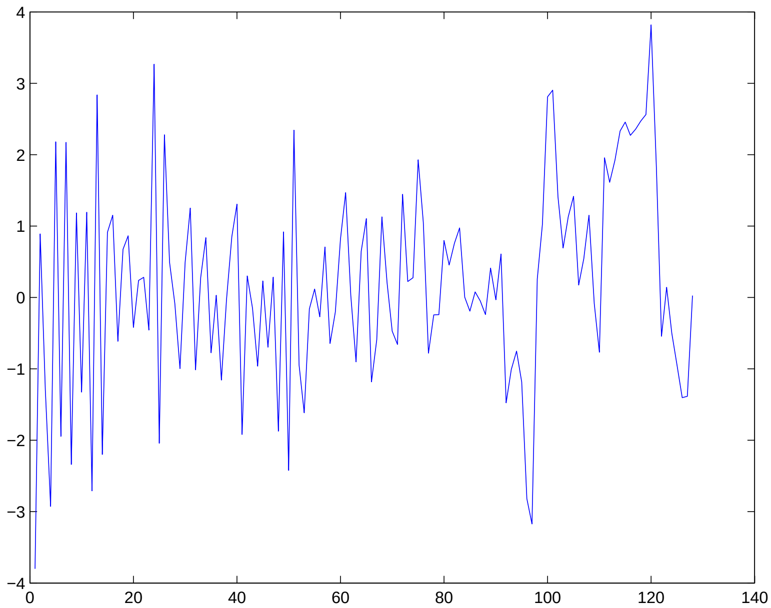


In the following example, we will consider a tdAR(1) process but with the innovation standard deviation being a periodic function of time. Let us first show a unique artificial series of length 128 generated by (1.1) with

$$\phi_t^{(n)} = \phi' + (t - \frac{n+1}{2})\phi'', \tag{5.2}$$

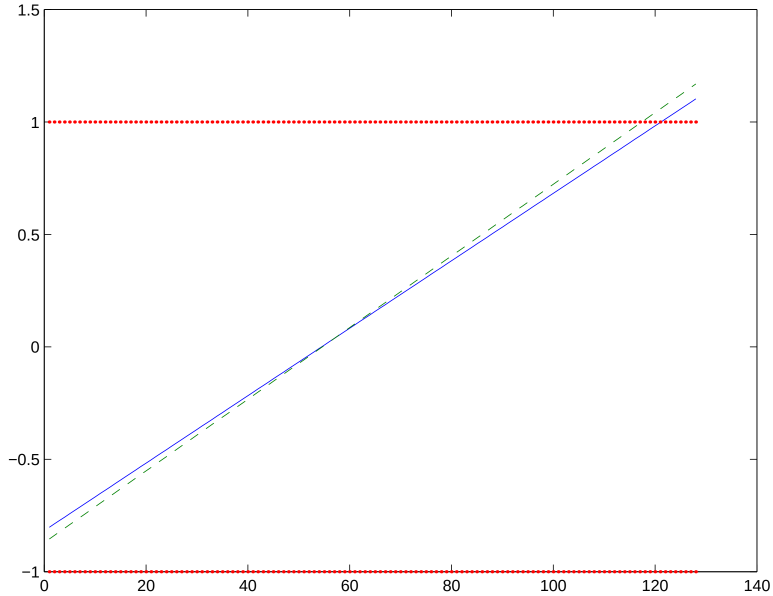
with $\phi' = 0.15$, $\phi'' = 0.015$ and the e_t are normally and independently distributed with mean 0 and variance h_t, where h_t is a periodic function of t with period 12, simulating a seasonal heteroscedasticity for monthly data. Furthermore h_t, which doesn't depend on n, assumes values $h = 0.5$ and $1/h = 2$, each during six consecutive time points. The series plotted in Figure 4 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 5). The parameters are estimated using the exact Gaussian maximum likelihood method which provides the following estimates (with the standard errors): $\hat{\phi}' = 0.157$ (± 0.062), $\hat{\phi}'' = 0.0159$ (± 0.0014), and $\hat{h} = 0.344$ (± 0.044), which are compatible with the true values. For $n = 128$, we provide the fit of $\phi_t^{(n)}$ and h_t, respectively, in Figure 5 and 6. Figure 7 and 8 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 theory of locally stationary processes since $\phi_t^{(n)} > 1$ for some t, and h_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 h_t is periodic with a fixed period.

Figure 4: Artificial series produced using the process defined by (1.1) and (5.2) (see text for details).



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.

Figure 5: True value of $\phi_t^{(n)}$ (which goes above 1!) (solid line) and its fit (discontinuous line).



The parameterisation is

$$\phi_t^{(n)}(\beta) = \phi^{'} + \left(t - \frac{n+1}{2}\right)\phi^{''} + \left(t - \frac{n+1}{2}\right)^2 \phi^{'''}, \qquad (5.3)$$

and $h_t^{1/2}$ is a periodic function which oscillates between the two values $h = 0.5$ and $1/h = 2$, 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 $h_t^{(n)}$. The latter capability is not used here but well an older implementation for intervention analysis (Mélard, 1981). Estimation are obtained by numerically maximising 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 4. Unfortunately, some estimates of the standard errors were unreliable so their averages were useless and replaced by medians. Note that 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 AM, 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 5, 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 994 series and the estimated standard error were always reliable so that their average across simulations are displayed.

Figure 6: True value of $h_t^{1/2}$ (solid line) and its fit (discontinuous line).

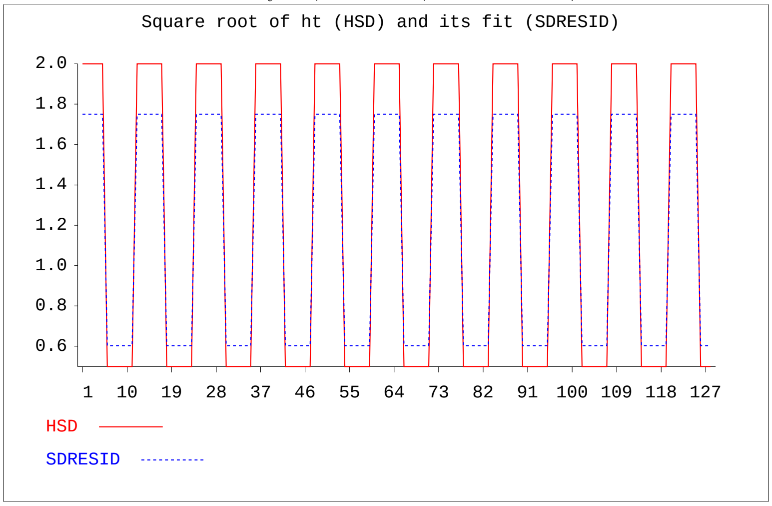


Table 4: Theoretical values of the parameters, averages and standard deviations of estimates across simulations, and medians across simulations of estimated standard errors $\phi^{'}$ (true value: 0.15), $\phi^{''}$ (true value: 0.015), and $\phi^{'''}$ (true value: 0), and h (true value 0.5) for the tdAR(1) model described above, for n = 128; 864 replications (out of 1000).

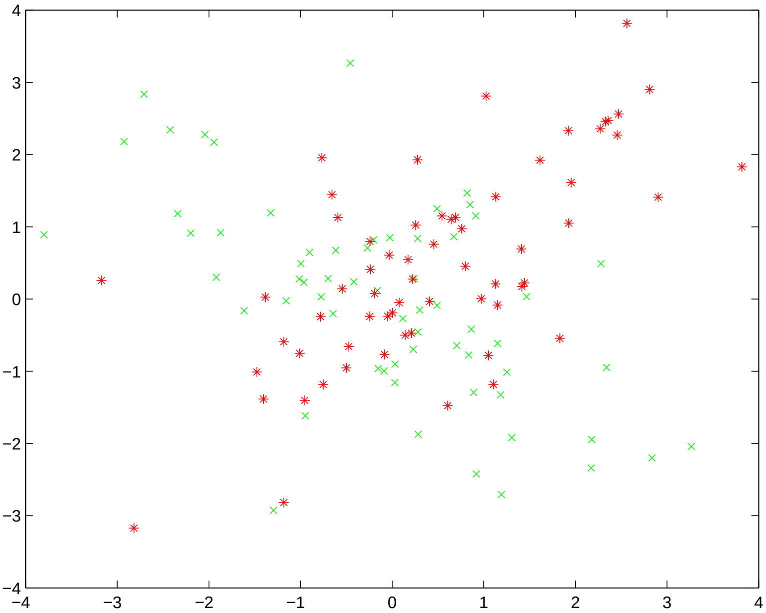
Parameter n true value	average	standard deviation	median of standard error
$\phi^{'} = 0.15$	0.23554	0.14611	0.10380
$\phi^{''} = 0.015$	0.01282	0.00222	0.00160
$\phi^{'''} = 0.0$	-0.00000	0.00005	0.00005
$h = 0.5$	0.54054	0.07857	0.08157

6. A comparison with the theory of cyclically time-dependent models

Here we will focus on BF, Bibi and Francq (2003), but part of the discussion is also appropriate for older approaches like Kwoun and Yajima (1986), Tjøstheim (1984), and Hamdoune (1995). BF have developed a general theory of estimation for linear models with time-dependent coefficients which is particularly aimed at the case of cyclically time-dependent coefficients. See also Francq and Gautier (2004a, b, c) and Gautier (2005).

The linear models include autoregressive but also moving average or ARMA models like AM. The coefficients can depend on t in a general way but not on n. 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-generalised least squares method. The basic assumptions 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

Figure 7: $w_t^{(128)}$ as a function of $w_{t-1}^{(128)}$ (crosses: $t \leq 64$, stars: $t > 64$)



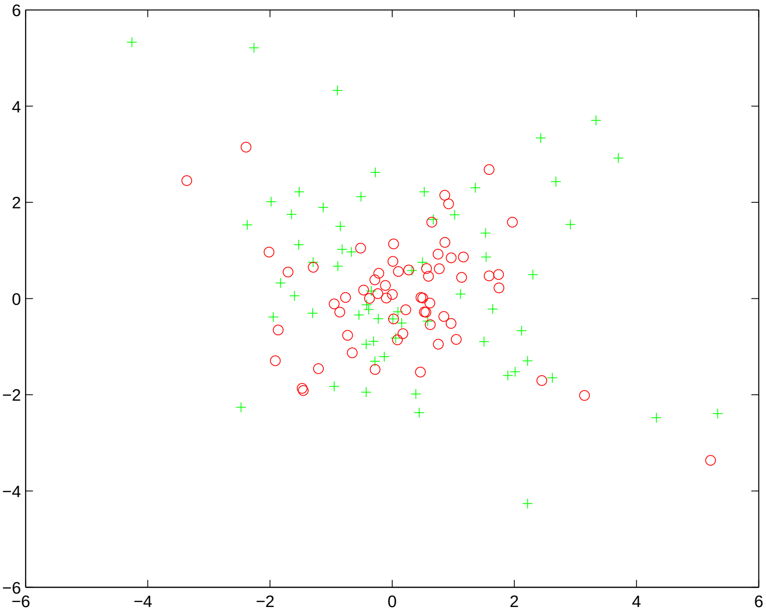
like in AM. The examples that are treated are either similar to Example 1 of AM, or consist on a conditionally heteroscedastic process with two regimes based on a sequence of Bernoulli random variables independent from the innovation process. It is difficult to see what models are excluded but Examples 2 to 5 of AM do not seem to be covered because their coefficients depend on t and n, not only on t.

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, not because of the periodicity in the innovation standard deviation, but well because the autoregressive coefficient is a function of t and n; not only on t. The simulations that were shown are also a sort of counterexample of the necessity of the assumptions made by

Table 5: 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 h (true value 0.5) for the tdAR(1) model described above, for $n = 128$; 864 replications (out of 1000).

Parameter n true value	average	standard deviation	average of standard error
$\phi^{'} = 0.15$	0.22023	0.12577	0.06683
$\phi^{''} = 0.015$	0.01305	0.00202	0.00146
$h = 0.5$	0.54290	0.07847	0.06984

Figure 8: $w_t^{(128)}$ as a function of $w_{t-1}^{(128)}$ (plusses: high scatter, when $h_t = 2$, stars: circles: small scatter, when $h_t = 0.5$.



BF. Nevertheless, it should be remarked that AM does not accept conditional heteroscedasticity at the present time so that the model considered by BF keeps a good position in that area.

7. Conclusions

This paper was motivated by suggestions to see if the results in AM simplify much in the case of autoregressive or even AR(1) processes, and by requests to compare more deeply the AM approach with others and push it in harder situations. We have shown that there are not many simplifications, perhaps due to the intrinsically complex nature of ARMA processes with time dependent coefficients. Nevertheless, we have been able to simplify one of the assumptions for AR(1) processes. We have taken the opportunity of this study on autoregressive processes with time dependent coefficients in order to develop an alternative approach based on a ρ-mixing condition instead of the strange assumption $H_{2'.7}$ made in AM. It was a strange assumption perhaps but at least could we check it 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 $H_{2'.7}$. The latter could be extended to MA and ARMA processes, which was not the case for the mixing condition. Although theoretical results for AR(2) processes could not be shown in closed form expressions, the simulations indicate that the method is robust when causality becomes questionable.

We have shown more stressing simulations than in AM and other examples on economic data that exhibit a nonstationary behavior. ARIMA models could have been possible for these simulations and examples but this paper has focused on autoregressive processes.

We have also compared the AM approach to others, more especially Dahlhaus theory of locally stationary processes and the BF approach aimed at cyclically time-dependent linear processes. Let us comment on this more deeply.

Like in Dahlhaus theory, a different process is considered 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 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 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. Nevertheless, BF can easily cope with some kind of conditionally heteroscedasticity, and assumes existence of lower moments. The example and simulations shown to distinguish AM from locally stationary processes is also illuminating the difference between AM and BF.

In some sense, AM can be seen as partly taking some features of both Dahlhaus 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 locally stationary processes. 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 which 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) economic time series and an innovation variance σ_t^2 which is a periodic function of time (this can be explained by seasonality like a winter/summer effect). To emphasize the difference with Dahlhaus' 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 σ_t^2 is piecewise constant, hence not compatible with differentiability at each time.

Appendix. Theorem 2' of Azrak and Mélard (2006)

Consider an autoregressive-moving average process and suppose that the functions $\phi_{tk}^{(n)}(\beta)$, $\theta_{tk}^{(n)}(\beta)$ and $h_t^{(n)}(\beta)$ are three times continuously differentiable with respect to β, in the open set B containing the true value β^0 of β, 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 $\forall t = 1, ..., n$ and uniformly with respect to n:

$$H_{2'.1} \quad \sum_{k=\nu}^{t-1} \{\psi_{tik}^{(n)}(\beta^0, \beta^0)\}^2 < N_1 \Phi^{\nu-1}, \quad \sum_{k=\nu}^{t-1} \{\psi_{tik}^{(n)}(\beta^0, \beta^0)\}^4 < N_2 \Phi^{\nu-1},$$

$$\sum_{k=\nu}^{t-1} \{\psi_{tijk}^{(n)}(\beta^0, \beta^0)\}^2 < N_3 \Phi^{\nu-1}, \quad \sum_{k=\nu}^{t-1} \{\psi_{tijk}^{(n)}(\beta^0, \beta^0)\}^4 < N_4 \Phi^{\nu-1},$$

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

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

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

$$H_{2'.3} \qquad 0 < m \leq h_t^{(n)}(\beta^0) \leq m_1;$$

$$H_{2'.4} \qquad E(\{w_t^{(n)}\}^4) \leq M;$$

$$H_{2'.5} \qquad (\sigma^{-8} E(e_t^8) \leq K.$$

Suppose furthermore that

$$H_{2'.6} \quad \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left[\sigma^{-2} E_{\beta^0} \left(\frac{\partial e_t^{(n)}(\beta)}{\partial \beta_i} \{h_t^{(n)}(\beta)\}^{-1} \frac{\partial e_t^{(n)}(\beta)}{\partial \beta_j} \right) \right.$$

$$\left. + \frac{1}{2} \left\{ \frac{\partial h_t^{(n)}(\beta)}{\partial \beta_i} \{h_t^{(n)}(\beta)\}^{-2} \frac{\partial h_t^{n)}(\beta)}{\partial \beta_j} \right\}_{\beta=\beta^0} \right] = V_{ij}(\beta^0),$$

$i, j = 1, ..., r$, where the matrix $V(\beta^0) = (V_{ij}(\beta^0))_{1\leq i,j\leq r}$ is a strictly definite positive matrix;

$$H_{2'.7} \quad \frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-1-d} \psi_{tik}^{(n)}(\beta^0)\psi_{tjk}^{(n)}(\beta^0)\psi_{t+d,i,k+d}^{(n)}(\beta^0)\psi_{t+d,j,k+d}^{(n)}(\beta^0)\{h_{t-k}^{(n)}(\beta^0)\}^2\kappa_{4,t-k}$$

$$= O(\frac{1}{n}), \qquad \text{and} \qquad \frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-d} \psi_{tik}^{(n)}(\beta^0)\psi_{t+d,j,k+d}^{(n)}(\beta^0)h_{t-k}^{(n)}(\beta^0) = O(\frac{1}{n}),$$

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;
- $n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V(\beta^0)^{-1}\, W(\beta^0)V(\beta^0)^{-1})$ where there exists a matrix $W(\beta^0)$ whose elements are defined by (2.9).

References

ANDREWS, D. W. K. (1988) Laws of large numbers for dependent non-identically distributed random variables, *Econom. Theory* **4**, 458-467.

AZRAK, R. and MELARD, G. (1998) The exact quasi-likelihood of time-dependent ARMA models, *J. Statist. Plann. Inference* **68**, 31-45.

AZRAK, R. and MELARD, G. (2006) Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients, *Statistical Inference for Stochastic Processes* **9**, 279-330.

BIBI, A. and FRANCQ, C. (2003) Consistent and asymptotically normal estimators for cyclically time-dependent linear models, *Ann. Inst. Statist. Math.* **55**, 41-68.

BOSQ, D. (1998) *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction*, Second Edition, Springer, New York.

BOX, G. E. P., JENKINS, G. M., and REINSEL, G. C. (1994) *Time Series Analysis, Forecasting and Control*, Third Edition, Prentice-Hall, New Jersey.

CHANDA, K. C. (1974) Strong mixing properties of linear stochastic processes. *J. Appl. Probability* **11**, 401408.

DAHLHAUS, R. (1996a) Maximum likelihood estimation and model selection for locally stationary processes, *J. Nonparametr. Statist.* **6**, 171-191.

DAHLHAUS, R. (1996b) Asymptotic statistical inference for nonstationary processes with evolutionary spectra, In *Athens Conference on Applied Probability an Time Series Analysis*, (P. M. Robinson and M. Rosenblatt, eds) **2**. Springer, New York, pp. 145-159.

DAHLHAUS, R. (1997) Fitting time series models to nonstationary processes, *Ann. Statist.* **25**, 1-37.

FRANCQ, C. and GAUTIER, A. (2004a) Estimation de modèles ARMA à changements de régime récurrents, *C. R. Acad. Sci. Paris, Ser. I* **339**, 55-58.

FRANCQ, C. and GAUTIER, A. (2004b) Large sample properties of parameter least squares estimates for time-varying ARMA models, *J. Time Ser. Anal.* **25**, 765-783.

FRANCQ, C. and GAUTIER, A. (2004c) Estimation of time-varying ARMA models with Markovian changes in regime, *Statist. Prob. Letters* **70**, 243-251.

GAUTIER, A. (2005), Influence asymptotique de la correction par la moyenne sur l'estimation d'un modèle AR périodique, *C. R. Acad. Sci. Paris, Ser. I* **340**, 315-318.

HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and its Application,* Academic Press, New York.

HAMDOUNE, S. (1995) Etude des problèmes d'estimation de certains modèles ARMA évolutifs, Thesis presented at Université Henri Poincaré, Nancy 1.

KOLMOGOROV, A. N. and ROZANOV, YU. A. (1960) On strong mixing conditions for stationary Gaussian processes, *Theory Probab. Appl.* **5**, 204-208.

KWOUN, G.H. and YAJIMA, Y. (1986) On an autoregressive model with time-dependent coefficients, *Ann. Inst. Statist. Math.* **38**, Part A, 297-309.

MELARD, G. (1981) On an alternative model for intervention analysis, in O.D. ANDERSON and M.R. PERRYMAN (eds), *Time Series Analysis*, North-Holland, Amsterdam, pp. 345-354.

MELARD, G. (1982) The likelihood function of a time-dependent ARMA model, in O.D. ANDERSON and M.R. PERRYMAN (eds), *Applied Time Series Analysis*, North-Holland, Amsterdam, pp. 229-239.

MELARD, G. and PASTEELS, J.-M. (2000) Automatic ARIMA modeling including interventions, using time series expert software", *International Journal of Forecasting* **16**, 497-508.

PHAM, D. and TRAN, T. (1985) Some mixing properties of time series models, *Stochastic Process. Appl.* **19**, 297-303.

RIO, E. (2000) *Théorie asymptotique des processus aléatoires faiblement dépendants*, Springer-Verlag, Berlin.

TJØSTHEIM, D. (1984) Estimation in linear time series models I: stationary series, Departement of Mathematics, University of Bergen 5000 Bergen, Norway and Departement of Statistics, University of North Carolina Chapel Hill, North

Carolina 27514.

VAN BELLEGEM, S. and DAHLHAUS, R. (2006) Semiparametric estimation by model selection for locally stationary processes, *J. Roy. Statist. Soc. Ser. B* **68**, 721-746.

VAN BELLEGEM, S. and von SACHS, R. (2004) Forecasting economic time series with unconditional time-varying variance, *International Journal of Forecasting* **20**, 611-627.

WEGMAN, E. J. (1974) Some results on non stationary first order autoregression, *Technometrics* **16**, 321-322.