<https://helda.helsinki.fi>

Optimal forecasting of noncausal autoregressive time series

Lanne, Markku

2012

Lanne , M , Luoto , J & Saikkonen , P 2012 , ' Optimal forecasting of noncausal autoregressive time series ' , International Journal of Forecasting , vol. 28 , no. 3 , pp. 623-631 . <https://doi.org/10.1016/j.ijforecast.2011.08.003>

<http://hdl.handle.net/10138/230228>

<https://doi.org/10.1016/j.ijforecast.2011.08.003>

cc_by_nc_nd

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Optimal Forecasting of Noncausal Autoregressive Time Series[†]

Markku Lanne*

University of Helsinki and HECER

Jani Luoto**

University of Helsinki and HECER

Pentti Saikkonen***

University of Helsinki and HECER

Abstract

In this paper, we propose a simulation-based method for computing point and density forecasts for univariate noncausal and non-Gaussian autoregressive processes. Numerical methods are needed to forecast such time series because the prediction problem is generally nonlinear and no analytic solution is therefore available. According to a limited simulation experiment, the use of a correct noncausal model can lead to substantial gains in forecast accuracy over the corresponding causal model. An empirical application to U.S. inflation demonstrates the importance of allowing for noncausality in improving point and density forecasts.

[†] We would like to thank the editor, associate editor and an anonymous referee for useful comments. Financial support from the Academy of Finland and the OP-Pohjola Group Research Foundation is gratefully acknowledged.

* Department of Political and Economic Studies, University of Helsinki, P.O.Box 17 (Arkadiankatu 7), FIN-00014 University of Helsinki, Finland, e-mail: markku.lanne@helsinki.fi

** Department of Political and Economic Studies, University of Helsinki, P.O.Box 17 (Arkadiankatu 7), FIN-00014 University of Helsinki, Finland, e-mail: jani.luoto@helsinki.fi

*** Department of Mathematics and Statistics, University of Helsinki, P.O.Box 68 (Gustaf Hållströmin katu 2b), FIN-00014 University of Helsinki, Finland, e-mail: pentti.saikkonen@helsinki.fi

1 Introduction

Univariate autoregressive (AR) models are commonly employed in analyzing economic time series. Typical fields of application include forecasting and the measurement of persistence. However, virtually all economic applications so far restrict themselves to causal autoregressive models where the current value of the variable of interest is forced to depend only on its past. In contrast, applications of noncausal AR models allowing for dependence on the future are almost nonexistent in econometrics. The major references in the relatively scant statistical literature on noncausal AR models include Breidt et al. (1991) and Rosenblatt (2000).¹ From the econometric perspective, noncausal AR models have recently been considered by Lanne and Saikkonen (2011), who proposed a new formulation of the model with attractive features from the viewpoints of statistical inference and economic interpretation. Their results suggest that expanding the set of univariate AR models in the noncausal direction may indeed be worthwhile in empirical economic research.

As pointed out above, the statistical literature on noncausal AR models is not voluminous and forecasting with these models has so far hardly been considered at all. To the best of our knowledge, the only exception is Rosenblatt (2000) whose Corollary 5.4.2 shows that in the non-Gaussian case the optimal (in mean square sense) one-step ahead predictor is generally nonlinear. However, no practically useful forecasting method seems to be available although forecasting is probably the most important application of univariate models. In addition, forecasts are needed in computing impulse responses on which measures of persistence in economic time series can be

¹Noncausal and potentially noninvertible autoregressive moving average models, as well as their special cases referred to as all-pass models, have also been studied in the statistical literature (see, inter alia, Lii and Rosenblatt (1996), Huang and Pawitan (2000), Breidt et al. (2001), and Andrews et al. (2006)).

based. Hence, being able to compute forecasts is crucial for noncausal AR models to be a useful tool in empirical economics. Moreover, devising techniques for forecasting in univariate models paves the way for the development of forecasting methods in corresponding multivariate models (see Lanne and Saikkonen (2009) for the noncausal vector autoregressive model), where forecasts are needed in conducting structural analysis.

In this paper, we propose a simulation-based method of forecasting with noncausal and non-Gaussian AR models. The Gaussian case will not be considered explicitly because then the noncausal AR model is indistinguishable from its causal counterpart and the conventional linear forecasting method is optimal. As already mentioned, in the non-Gaussian case the prediction problem is generally nonlinear which explains why numerical methods are needed to compute forecasts. Our forecasting method has some similarities to the recent method developed by Breidt and Hsu (2005) for non-Gaussian and potentially noninvertible moving average processes. In fact, our forecasts are formed by writing the noncausal AR model in a form in which the noncausal AR part is approximated by a long moving average containing future innovations. In practice this long moving average is recovered from the considered estimated noncausal AR model so that no moving average parameters are estimated directly. According to simulations the performance of the proposed method is good when the true model is noncausal.

The rest of the paper is organized as follows. The formulation of the noncausal AR model of Lanne and Saikkonen (2011) is presented and its maximum likelihood estimation and statistical inference are discussed in Section 2. Our forecast method is described in Section 3. To illustrate the properties of the forecast procedure and the gains in forecast accuracy over a causal model in the presence of noncausality, some Monte Carlo simulation results are reported in Section 4. An empirical application to U.S. inflation is provided in Section 5. Finally, Section 6 concludes.

2 Noncausal Autoregression

2.1 Model

In this section, we describe the formulation of the noncausal autoregressive model suggested by Lanne and Saikkonen (2011). As pointed out above, our formulation differs somewhat from that employed in the earlier literature. In particular, compared to Breidt et al. (1991), the autoregressive polynomial in our model explicitly involves both leads and lags. One advantage of this formulation is that statistical inference on autoregressive parameters is facilitated. Furthermore, the autoregressive parameters are orthogonal to the parameters in the distribution of the error term so that inference on these two sets of parameters is asymptotically independent.

Consider a stochastic process y_t ($t = 0, \pm 1, \pm 2, \dots$) generated by

$$\varphi(B^{-1})\phi(B)y_t = \epsilon_t, \quad (1)$$

where $\phi(B) = 1 - \phi_1 B - \dots - \phi_r B^r$, $\varphi(B^{-1}) = 1 - \varphi_1 B^{-1} - \dots - \varphi_s B^{-s}$, and ϵ_t is a sequence of independent, identically distributed (continuous) random variables with mean zero and variance σ^2 or, briefly, $\epsilon_t \sim i.i.d. (0, \sigma^2)$. Moreover, B is the usual backward shift operator, that is, $B^k y_t = y_{t-k}$ ($k = 0, \pm 1, \dots$), and the polynomials $\phi(z)$ and $\varphi(z)$ have their zeros outside the unit circle so that

$$\phi(z) \neq 0 \quad \text{for } |z| \leq 1 \quad \text{and} \quad \varphi(z) \neq 0 \quad \text{for } |z| \leq 1. \quad (2)$$

We use the abbreviation $\text{AR}(r, s)$ for the model defined by (1) and sometimes write $\text{AR}(r)$ for $\text{AR}(r, 0)$. If $\varphi_1 = \dots = \varphi_s = 0$, model (1) reduces to the conventional causal $\text{AR}(r)$ model with y_t depending on its past but not future values. The more interesting cases from the viewpoint of this paper arise, when this restriction does not hold. If $\phi_1 = \dots = \phi_r = 0$, we have the purely noncausal $\text{AR}(0, s)$ model with dependence on future values only. In the mixed $\text{AR}(r, s)$ case where neither restriction holds, y_t depends on its past as well as future values.

The conditions in (2) imply that y_t has the two-sided moving average representation

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad (3)$$

where ψ_j is the coefficient of z^j in the Laurent series expansion of $\varphi(z^{-1})^{-1} \phi(z)^{-1} \stackrel{\text{def}}{=} \psi(z)$. This expansion exists in some annulus $b < |z| < b^{-1}$ with $0 < b < 1$ and with $\psi_{|j|}$ converging to zero exponentially fast as $|j| \rightarrow \infty$. From (1) one also obtains the representation

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_r y_{t-r} + v_t, \quad (4)$$

where $v_t = \varphi(B^{-1})^{-1} \epsilon_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t+j}$ with β_j the coefficient of z^j in the power series expansion of $\varphi(B^{-1})^{-1}$. This representation will be used to obtain forecasts.

2.2 Estimation and Inference

A well-known feature of noncausal autoregressions is that a non-Gaussian error term is required to achieve identification. Thus, we assume that the error term ϵ_t is non-Gaussian and that its distribution has a (Lebesgue) density $f_{\sigma}(x; \lambda) = \sigma^{-1} f(\sigma^{-1}x; \lambda)$ which depends on the parameter vector λ ($d \times 1$) in addition to the scale parameter σ already introduced. The function $f(x; \lambda)$ is assumed to satisfy the regularity conditions stated in Andrews et al. (2006) and Lanne and Saikkonen (2011). These conditions imply that $f(x; \lambda)$ is twice continuously differentiable with respect to (x, λ) , non-Gaussian, and positive for all $x \in \mathbb{R}$ and all permissible values of λ .

Let $y_1, \dots, y_T$ be an observed time series generated by the noncausal autoregression (1). Define $u_t = \varphi(B^{-1}) y_t$ and $v_t = \phi(B) y_t$, and set $\mathbf{y} = (y_1, \dots, y_T)$ and $\mathbf{z} = (u_1, \dots, u_r, \epsilon_{r+1}, \dots, \epsilon_{T-s}, v_{T-s+1}, \dots, v_T)$. The unknown parameters of the model are collected in the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\boldsymbol{\phi}, \boldsymbol{\varphi}, \sigma, \lambda)$ where $\boldsymbol{\theta}_1 = (\boldsymbol{\phi}, \boldsymbol{\varphi})$ and $\boldsymbol{\theta}_2 = (\sigma, \lambda)$ with $\boldsymbol{\phi} = (\phi_1, \dots, \phi_r)$ and $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_s)$. As shown in Lanne and Saikkonen (2011), the vectors $\mathbf{y}$ and $\mathbf{z}$ are related by a linear transformation of the

form $\mathbf{z} = \mathbf{B}\mathbf{A}\mathbf{y}$ where the matrices $\mathbf{A}$ and $\mathbf{B}$ depend on the parameter vector $\boldsymbol{\theta}_1$. Moreover, the determinant of $\mathbf{B}$ is unity whereas the determinant of $\mathbf{A}$ is independent of the sample size T . We shall not provide explicit forms of these matrices because they will not be needed in our subsequent developments. As in Lanne and Saikkonen (2011) we can now conclude that the joint density function of the data vector $\mathbf{y} = (y_1, \dots, y_T)$ can be expressed as

$$p(\mathbf{y}; \boldsymbol{\theta}) = h_{\mathbf{u}}(\varphi(B^{-1})y_1, \dots, \varphi(B^{-1})y_r) \left(\prod_{t=r+1}^{T-s} f_{\sigma}(\varphi(B^{-1})\phi(B)y_t; \lambda) \right) \times h_{\mathbf{v}}(\phi(B)y_{T-s+1}, \dots, \phi(B)y_T) |\det(\mathbf{A})|, \quad (5)$$

where $h_{\mathbf{u}}$ and $h_{\mathbf{v}}$ signify the joint density functions of the random vectors $\mathbf{u} = (u_1, \dots, u_r)$ and $\mathbf{v} = (v_{T-s+1}, \dots, v_T)$, respectively.

Analogously to Breidt et al. (1991), Lanne and Saikkonen (2011) use the second factor on the right hand side of (5) to approximate the likelihood function. They show that the resulting (local) maximum likelihood (ML) estimator is asymptotically normally distributed and the covariance matrix of the limiting distribution is block diagonal with respect to the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Moreover, a consistent estimator of the limiting covariance matrix is obtained in the usual way from the standardized Hessian of the approximate log-likelihood function. Thus, standard errors of estimators and conventional Wald tests with asymptotic χ^2 -distribution under the null hypothesis can be constructed as usual and the same is true for likelihood ratio tests based on the approximate log-likelihood function.

So far, we have assumed that the model orders r and s are known. Procedures to specify these quantities are discussed by Lanne and Saikkonen (2011), who suggest selecting the orders that maximize the approximate likelihood function among the p th order models with $p = r + s$. The order p is first selected based on a Gaussian model capturing all autocorrelation in the series, and in specifying the non-Gaussian error distribution, the properties of the residuals of this models may be used. Upon

estimation, the adequacy of the selected model can be assessed by standard diagnostic checks. While other model selection procedures might be considered, in this paper, we employ this method that seems to work reasonably well according to the simulation results in Lanne and Saikkonen (2011).

3 Forecasting Method

3.1 Point Forecasts

As pointed out in the Introduction, the prediction problem in noncausal autoregressions has been studied by Rosenblatt (2000) whose Corollary 5.4.2 shows that in the non-Gaussian case the optimal (in mean square sense) one-step ahead predictor is generally nonlinear. In the Gaussian case the prediction problem is linear because the best linear predictor is always the best mean square predictor. As far as we know, no practical way to compute forecasts in the non-Gaussian noncausal case has been presented. One method is described below.

Let $E_T(\cdot)$ signify the conditional expectation operator given the observed data vector $\mathbf{y} = (y_1, \dots, y_T)$. From (4) it is seen that the optimal predictor of y_{T+h} ($h > 0$) based on the observed data satisfies

$$E_T(y_{T+h}) = \phi_1 E_T(y_{T+h-1}) + \dots + \phi_r E_T(y_{T+h-r}) + E_T(v_{T+h}).$$

Thus, if we can forecast the variable v_{T+h} , we can compute forecasts for the observed process recursively. To solve this problem we use the approximation $v_{T+h} \approx \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}$, where the integer M is supposed to be so large that the approximation error is negligible for all forecast horizons h of interest. To a close approximation we then have

$$E_T(y_{T+h}) \approx \phi_1 E_T(y_{T+h-1}) + \dots + \phi_r E_T(y_{T+h-r}) + E_T\left(\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}\right). \quad (6)$$

To be able to compute the last conditional expectation on the right hand side of (6) we derive the conditional density of $\epsilon^+ = (\epsilon_{T+1}, \dots, \epsilon_{T+M})$ given the data vector $\mathbf{y}$. If $\mathbf{z}$, $\mathbf{A}$ and $\mathbf{B}$ are as in Section 2.2, the matrix of the linear transformation $(\mathbf{y}, \epsilon^+) \rightarrow (\mathbf{z}, \epsilon^+)$ is $\text{diag}(\mathbf{B}\mathbf{A}, I_M)$. The Jacobian of this transformation is $\det(\mathbf{A})$ so that the joint density function of $(\mathbf{y}, \epsilon^+)$ is

$$p(\mathbf{y}, \epsilon^+; \boldsymbol{\theta}) = h_{\mathbf{u}}(\mathbf{u}(\boldsymbol{\varphi})) \left(\prod_{t=r+1}^{T-s} f_{\sigma}(\epsilon_t(\boldsymbol{\theta}_1); \lambda) \right) h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\boldsymbol{\phi}), \epsilon^+) |\det(\mathbf{A})|, \quad (7)$$

where we have simplified the notation by writing $\mathbf{u}(\boldsymbol{\varphi}) = (u_1(\boldsymbol{\varphi}), \dots, u_r(\boldsymbol{\varphi})) = (\varphi(B^{-1})y_1, \dots, \varphi(B^{-1})y_r)$, $\mathbf{v}(\boldsymbol{\phi}) = (v_{T-s+1}(\boldsymbol{\phi}), \dots, v_T(\boldsymbol{\phi})) = (\phi(B)y_{T-s+1}, \dots, \phi(B)y_T)$ and $\epsilon_t(\boldsymbol{\theta}_1) = \varphi(B^{-1})\phi(B)y_t$, $t = r+1, \dots, T-s$. Using (5) and (7) we can now write the conditional density function of ϵ^+ given $\mathbf{y}$ as

$$p(\epsilon^+ | \mathbf{y}; \boldsymbol{\theta}) = \frac{h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\boldsymbol{\phi}), \epsilon^+)}{h_{\mathbf{v}}(\mathbf{v}(\boldsymbol{\phi}))} = \frac{h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\boldsymbol{\phi}), \epsilon^+)}{\int h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\boldsymbol{\phi}), \epsilon^+) d\epsilon^+}.$$

The last conditional expectation on the right hand side of (6) can thus be expressed as

$$E_T \left(\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} \right) = \frac{1}{h_{\mathbf{v}}(\mathbf{v}(\boldsymbol{\phi}))} \int \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\boldsymbol{\phi}), \epsilon^+) d\epsilon^+. \quad (8)$$

Next we have to find a feasible way to handle the density functions $h_{\mathbf{v}}(\mathbf{v}(\boldsymbol{\phi}))$ and $h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\boldsymbol{\phi}), \epsilon^+)$. To this end, consider the linear transformation $(\mathbf{v}, \epsilon^+) \rightarrow (\epsilon_{T-s+1}, \dots, \epsilon_T, \epsilon^+)$. As $v_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t+j}$ its inverse transformation satisfies the approximate relation

$$\begin{bmatrix} 1 & \beta_1 & \cdots & \cdots & \cdots & \cdots & \beta_{M+s-1} \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & \beta_1 & \cdots & \cdots & \beta_M \\ \vdots & & \ddots & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{T-s+1} \\ \vdots \\ \epsilon_T \\ \epsilon_{T+1} \\ \vdots \\ \epsilon_{T+M} \end{bmatrix} \approx \begin{bmatrix} v_{T-s+1} \\ \vdots \\ v_T \\ \epsilon_{T+1} \\ \vdots \\ \epsilon_{T+M} \end{bmatrix}.$$

We write this briefly as

$$\mathbf{C}\mathbf{e} \approx \mathbf{w}.$$

Using the approximate inverse transformation $\mathbf{e} \approx \mathbf{C}^{-1}\mathbf{w}$ and ignoring the approximation error it is seen that, for $1 \leq j \leq s$, the j th component of the vector $\mathbf{e}$ depends linearly on $v_{T-s+j}, \dots, v_T$ and $\epsilon_{T+1}, \dots, \epsilon_{T+M}$. We therefore write

$$\mathbf{e} = (e_{T-s+1}(\mathbf{v}_{T-s+1}, \boldsymbol{\epsilon}^+), \dots, e_T(\mathbf{v}_T, \boldsymbol{\epsilon}^+), \epsilon_{T+1}, \dots, \epsilon_{T+M}),$$

where $\mathbf{v}_{T-s+j} = (v_{T-s+j}, \dots, v_T)$ and, if $\mathbf{D} = [d_{ij}] = \mathbf{C}^{-1}$,

$$\mathbf{e} \approx \begin{bmatrix} 1 & d_{12} & \cdots & \cdots & \cdots & \cdots & d_{1,M+s} \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & 1 & d_{s,s+1} & \cdots & \cdots & d_{s,M+s} \\ \vdots & & \ddots & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{T-s+1} \\ \vdots \\ v_T \\ \epsilon_{T+1} \\ \vdots \\ \epsilon_{T+M} \end{bmatrix}. \quad (9)$$

Thus, $e_{T-s+j}(\mathbf{v}_{T-s+j}, \boldsymbol{\epsilon}^+) = \sum_{k=j}^s d_{jk} v_{T-s+k} + \sum_{k=s+1}^{M+s} d_{jk} \epsilon_{T-s+k}$ ($1 \leq j \leq s$), where $d_{jj} = 1$ and, as the determinant of the matrix $\mathbf{D}$ is unity, a close approximation for the density function of $(\mathbf{v}, \boldsymbol{\epsilon}^+) = \mathbf{w}$ is given by

$$h_{\mathbf{v}, \boldsymbol{\epsilon}^+}(\mathbf{v}, \boldsymbol{\epsilon}^+) \approx \prod_{j=1}^s f_{\sigma}(e_{T-s+j}(\mathbf{v}_{T-s+j}, \boldsymbol{\epsilon}^+); \lambda) \cdot \prod_{t=T+1}^{T+M} f_{\sigma}(\epsilon_t; \lambda). \quad (10)$$

An approximation for the density function of $\mathbf{v}$ is obtained from this by integrating over $\boldsymbol{\epsilon}^+$.

To compute the value of the conditional expectation (8) we have to compute values of the density functions $h_{\mathbf{v}}(\mathbf{v})$ and $h_{\mathbf{v}, \boldsymbol{\epsilon}^+}(\mathbf{v}, \boldsymbol{\epsilon}^+)$ at the point $\mathbf{v} = \mathbf{v}(\phi)$ which depends on the observations (and the value of the parameter ϕ). Consider first the former.

Using (10) we get the approximation

$$\begin{aligned} h_{\mathbf{v}}(\mathbf{v}(\phi)) &= \int h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\phi), \epsilon^+) d\epsilon^+ \\ &\approx \int \prod_{j=1}^s f_{\sigma}(e_{T-s+j}(\mathbf{v}_{T-s+j}, \epsilon^+); \lambda) \cdot \prod_{t=T+1}^{T+M} f_{\sigma}(\epsilon_t; \lambda) d\epsilon^+. \end{aligned}$$

The last expression can be interpreted as the expectation of the first product therein with respect to the distribution of ϵ^+ . Using simulation, this expectation can therefore be approximated as

$$h_{\mathbf{v}}(\mathbf{v}(\phi)) \approx N^{-1} \sum_{i=1}^N \left(\prod_{j=1}^s f_{\sigma}(e_{T-s+j}(\mathbf{v}_{T-s+j}(\phi), \epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}); \lambda) \right),$$

where $\{\epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}\}$, $i = 1, \dots, N$, are mutually independent simulated realizations from the distribution of ϵ^+ . As $N \rightarrow \infty$, the right hand side of this approximation converges almost surely and provides an approximation for $h_{\mathbf{v}}(\mathbf{v}(\phi))$ that can be made arbitrarily accurate by choosing the integer M large enough.

In a similar way, we can use (10) and obtain an approximation for

$$\begin{aligned} &\int \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\phi), \epsilon^+) d\epsilon^+ \\ &\approx \int \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} \prod_{k=1}^s f_{\sigma}(e_{T-s+k}(\mathbf{v}_{T-s+k}, \epsilon^+); \lambda) \cdot \prod_{t=T+1}^{T+M} f_{\sigma}(\epsilon_t; \lambda) d\epsilon^+. \end{aligned}$$

Specifically, we have

$$\begin{aligned} &\int \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} h_{\mathbf{v}, \epsilon^+}(\mathbf{v}(\phi), \epsilon^+) d\epsilon^+ \\ &\approx N^{-1} \sum_{i=1}^N \left(\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} \prod_{k=1}^s f_{\sigma}(e_{T-s+k}(\mathbf{v}_{T-s+k}(\phi), \epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}); \lambda) \right). \end{aligned}$$

From the preceding discussion we can now conclude that the conditional expecta-

tion (8) can be approximated as

$$E_T \left(\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} \right) \approx \frac{N^{-1} \sum_{i=1}^N \left(\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j} \prod_{k=1}^s f_{\sigma} \left(e_{T-s+k}(\mathbf{v}_{T-s+k}(\phi), \epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}); \lambda \right) \right)}{N^{-1} \sum_{i=1}^N \left(\prod_{k=1}^s f_{\sigma} \left(e_{T-s+k}(\mathbf{v}_{T-s+k}(\phi), \epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}); \lambda \right) \right)}. \quad (11)$$

Using this approximation in (6), we can compute approximate forecasts recursively. The accuracy of the approximation depends on the choice of the integers M and N . For a good approximation these integers should be large enough but, on the other hand, the larger they are, the heavier is the computational burden. As $v_{T+h} \approx \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}$ the integer M should be so large that the coefficients β_j are practically zero for $j > M$. This in turn depends on the roots of the polynomial $\varphi(z)$. The closer the roots of this polynomial are to the unit circle the larger value of M should be used. The simulation results in Section 4 suggest that even relatively small values of M and N (50 and 10 000, respectively) are sufficient for a reasonable approximation that cannot be much improved upon by further increases.

3.2 Density and Interval Forecasts

The forecast method developed in the preceding section can be extended to compute interval forecasts. First note that the arguments used to obtain the forecast for $\sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}$ in (11) can readily be modified to obtain a forecast for any function of ϵ^+ . Specifically, the forecast of $q(\epsilon^+)$, say, is given by

$$E_T(q(\epsilon^+)) \approx \frac{N^{-1} \sum_{i=1}^N \left(q(\epsilon^+) \prod_{j=1}^s f_{\sigma} \left(e_{T-s+j}(\mathbf{v}_{T-s+j}(\phi), \epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}); \lambda \right) \right)}{N^{-1} \sum_{i=1}^N \left(\prod_{j=1}^s f_{\sigma} \left(e_{T-s+j}(\mathbf{v}_{T-s+j}(\phi), \epsilon_{T+1}^{(i)}, \dots, \epsilon_{T+M}^{(i)}); \lambda \right) \right)}. \quad (12)$$

For instance, letting $\mathbf{1}(\cdot)$ stand for the indicator function and choosing $q(\epsilon^+) = \mathbf{1}\left(\sum_{j=0}^{M-1} \beta_j \epsilon_{T+1+j} \leq x\right)$ in (12) yields a forecast for the conditional cumulative distribution function of $v_{T+1} \approx \sum_{j=0}^{M-1} \beta_j \epsilon_{T+1+j}$ at the point x . Choosing a grid $x_1, \dots, x_K$ with K large enough, one can obtain a forecast for the whole conditional cumulative distribution function of v_{T+1} . As $y_{T+1} = v_{T+1} + \sum_{j=1}^r \phi_j y_{T+1-j}$ (see (4)) a forecast for the conditional cumulative distribution function of y_{T+1} is obtained from this by treating $\sum_{j=1}^r \phi_j y_{T+1-j}$ as a constant. Using appropriate quantiles from the lower and upper tail of this forecast, an interval forecast for y_{T+1} can further be constructed.

Obtaining interval forecasts for y_{T+h} with $h > 1$ is slightly more complicated. Define the $r \times 1$ vector $\mathbf{y}_t = (y_t, \dots, y_{t-r+1})$ and write (4) in the companion form

$$\mathbf{y}_t = \Phi \mathbf{y}_{t-1} + \iota v_t,$$

where $\iota = (1, 0, \dots, 0)$ ($r \times 1$) and

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \cdots & \phi_r \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \quad (r \times r)$$

Using repetitive substitution and the approximation $v_{T+h} \approx \sum_{j=0}^{M-h} \beta_j \epsilon_{T+h+j}$, one now obtains

$$\mathbf{y}_{T+h} = \Phi^h \mathbf{y}_T + \sum_{i=0}^{h-1} \Phi^i \iota v_{T+h-i} \approx \Phi^h \mathbf{y}_T + \sum_{i=0}^{h-1} \Phi^i \iota \sum_{j=0}^{M-h+i} \beta_j \epsilon_{T+h-i+j}.$$

Thus, we need to find a forecast for the conditional cumulative distribution function of $\sum_{i=0}^{h-1} \iota' \Phi^i \iota \sum_{j=0}^{M-h+i} \beta_j \epsilon_{T+h-i+j}$. This is obtained from (12) with

$$q(\epsilon^+) = \mathbf{1}\left(\sum_{i=0}^{h-1} \iota' \Phi^i \iota \sum_{j=0}^{M-h+i} \beta_j \epsilon_{T+h-i+j} \leq x_k\right), \quad k = 1, \dots, K.$$

After this, a forecast for the conditional cumulative distribution function of y_{t+h} and interval forecasts for y_{t+h} can be constructed as in the case $h = 1$.

4 Simulation Study

In order to illustrate the gains in forecast accuracy of correctly allowing for non-causality and the properties of the proposed forecasting method, we provide a small Monte Carlo simulation experiment. Following Clements and Smith (1999) along with a number of other simulation studies on the forecasting performance of non-linear models, we use the model estimated with actual data as the data generating process (DGP). In particular, we consider $AR(r, s)$ models for the demeaned seasonally adjusted annualized quarterly U.S. inflation series based on the GDP implicit price deflator series extracted from the FRED database of the Federal Reserve Bank of St. Louis for the period from 1955:1 to 1989:4. In Section 5, we will provide evidence on the forecasting performance of various AR models for this series.

Using the model selection procedure suggested by Lanne and Saikkonen (2011), the $AR(1,4)$ model with t -distributed errors is selected.² The Gaussian $AR(5)$ model turns out sufficient in capturing the autocorrelation in the inflation series, and among the fifth-order models the $AR(1,4)$ model maximizes the approximate log-likelihood function. The p-values of the Wald test against the sixth-order $AR(2,4)$ and $AR(1,5)$ models equal 0.385 and 0.313, respectively, indicating adequacy of the $AR(1,4)$ specification. Judged by the Q-Q plot of the residuals (not shown), the t -distribution assumption seems reasonable, and this conclusion is also backed up by the precisely estimated relatively low value of the degree-of-freedom parameter, 3.99 (with standard error 1.40). The residuals are not autocorrelated but some remaining conditional

²Student's t distribution has been successfully employed in modeling inflation dynamics by Lanne and Saikkonen (2011) and Lanne and Luoto (2010). While other distributional assumptions could be entertained, the diagnostic checks below suggest that the t distribution fits quit well.

heteroskedasticity is detected. The estimation results are presented in Table 1.

We simulate 10 000 realizations of length $T+8$ from the estimated DGP. Using the first T observations in each realization, we estimate a causal AR(5,0) and a noncausal AR(1,4) model. Then we compute the point forecasts 1, 2, 4 and 8 periods ahead and finally compute the mean-square forecast error (MSFE) for each horizon over all the realizations. We consider two sample sizes ($T = 100, 200$) and three choices of the number of simulated realizations in the forecasting procedure ($N = 1000, 10\,000, 100\,000$). Three values of the truncation parameter M , 25, 50 and 100, are considered.

Table 2 presents the MSFEs of the AR(1,4) model 1, 2, 4 and 8 periods ahead. The accuracy of forecasts seems to increase with the sample size T . As far as the number of replications N is concerned, there is always clear improvement when N increases from 1 000 to 10 000, but virtually no improvement or even a slight deterioration related to the increase of N from 10 000 to 100 000. With $N = 10\,000$, the differences in the MSFEs based on the different values of M are minor. In conclusion, $N = 10\,000$ seems sufficient, and at least in this experiment, the value of M is not critical.

The relative MSFEs, i.e., the MSFEs of the AR(1,4) model divided by those of the AR(5,0) model, are reported in Table 3. The fact that all entries are below unity indicates the superiority of the (true) noncausal specification at all horizons considered. According to the test of Diebold and Mariano (1995) and West (1996), all differences are also significant at least at the 10% level, in most cases at the 1% level.³ In line with the results on the MSFEs in Table 2, the superiority of the noncausal model tends to improve as N increases from 1 000 to 10 000, but hardly

³We ignore the effect of estimation error as asymptotically irrelevant. As pointed out by West (1996), with fixed estimation period, asymptotic irrelevance requires asymptotic block diagonality between the estimators of the autoregressive parameters and the mean squared forecast errors. As it is not clear whether such asymptotic block diagonality holds in noncausal autoregressions, the test results should be interpreted with caution.

at all as the number of replications is further increased to 100 000. The increase in the sample size favors the noncausal model when forecasting one or two periods ahead, whereas at longer forecast horizons, its relative performance tends to slightly deteriorate with an increase in the sample size. The larger number of observations presumably facilitates more accurate approximation of the correct noncausal model by the misspecified causal AR model at longer forecast horizons.

5 Empirical Application

In this section, we illustrate the use of our forecasting method by an application to quarterly U.S. inflation. As already mentioned in Section 4, we compute inflation based on the seasonally adjusted GDP implicit price deflator series. Specifically, we consider recursive pseudo out-of-sample forecasts for the period between the first quarter of 1990 and the second quarter of 2008. The model is reestimated at each step, with the estimation sample always starting from the first quarter of 1955.⁴ Following the standard practice in the literature on inflation forecasting (for a survey of the recent literature, see Stock and Watson (2008)), we consider point forecasts of average inflation instead of the inflation prevailing at a particular quarter in the future.

Because the results in Section 4 suggest that a fifth-order AR model would be sufficient, we compare the forecasts of all six fifth-order models with t-distributed errors. In addition, we consider the Gaussian causal AR(5) model. In computing the forecasts, we set the number of replications N at 50 000 and the truncation parameter M at 100.⁵

⁴We also experimented with rolling and fixed schemes, but the differences turned out to be minor. The detailed results are not reported, but they are available upon request.

⁵To check the robustness with respect to the choice of M , we computed forecasts also with M set a 25 and 50. For the AR(1,4) model the results are virtually independent of the choice of M ,

The MSFEs are reported in Table 4. Comparison of the recursive out-of-sample MSFEs suggests four findings. First, the noncausal AR models outperform the causal models in that the best model is always one of the noncausal specifications. Second, with the exception of the two-quarter forecast horizon, the AR(1,4) model selected in the estimation period turns out to be the most accurate. At the two-quarter horizon, it is slightly dominated by the AR(2,3) specification. Third, at longer forecast horizons, the differences between the best noncausal model and the causal AR(5,0) model are also statistically significant according to the test of Diebold and Mariano (1995) and West (1996). In testing, we assume that the estimation error does not matter because of the recursive scheme and the fact that the estimation period is relatively long compared to the forecast period (see, West (1996)). Finally, AR(2,3) and AR(1,4) models always outperform the the Gaussian AR(5) model, and also the differences between the best noncausal model and the causal Gaussian AR(5) model are significant at the 10% level, except at the one-quarter horizon.

As discussed in Section 3, the forecasting method can easily be modified to produce density and interval forecasts. As an example, Figure 1 depicts the predictive cumulative distribution function of inflation for the last quarter of 1989 based on the AR(1,4) model estimated on data up to the preceding quarter. The observed value equals 2.63%, and the point forecast is 2.91%. Any interval forecast can be read off the distribution function; for instance, the 90% interval forecast comprises values between 0.94% and 4.84%. Hence, the forecast interval includes the observed inflation rate.

A more complete idea of the performance of the noncausal AR(1,4) model in density forecasting is given by the box-and-whisker plots of recursive one-step density forecasts in Figure 2. The bottom and top of the box are the 25% and 75% points, the interior line is the median, the bottom whisker is the 5% and the top whisker is the 95%.

and also in the case of the AR(0,5) model the changes are negligible.

the 95% point. In almost all cases, the observed value is well within the interquartile range.

6 Conclusion

To the best of our knowledge, the method proposed in this paper is the first attempt to obtain a practical forecasting procedure for noncausal autoregressions. Apparently forecasting has not been of much interest in the previous statistical literature on noncausal models with applications mostly confined to natural sciences and engineering. In many of these applications, it may actually not be reasonable to think of the employed model as a time series model but rather as a one-dimensional random field in which the direction of “time” is irrelevant and prediction is not of interest. However, the ability to compute forecasts is necessary for these models to be useful in economics and finance. Here we have only considered the univariate noncausal autoregression, but it should be possible to extend the method to multivariate models, such as the noncausal vector autoregression of Lanne and Saikonen (2009). In a multivariate setting, structural analysis, including impulse responses and forecast error decompositions, is based on forecasts, which emphasizes their importance from the econometric point of view.

The results of our simulation experiment and empirical application to U.S. inflation are encouraging, but more work is needed to evaluate the performance of the proposed method in different situations. Although our simulation experiment suggested that the computational burden of obtaining accurate forecasts is not very heavy, this may not hold generally, especially in a multivariate model. Moreover, only practical experience on forecasting different kinds of economic time series will reveal the true benefits of noncausal autoregressions in forecasting. To this end, Lanne et al. (2011) have recently employed the methods introduced in this article to compare

the forecasts of causal and noncausal autoregressive models of 170 monthly and 18 quarterly U.S. macroeconomic and financial time series. According to their results, the noncausal models consistently outperform the causal models in terms of the mean square and mean absolute forecast errors. While also their study only considers linear models, an interesting future research topic is the performance of noncausal AR models relative to a richer set of alternative models, including a number of commonly used nonlinear models. As far as U.S. inflation is concerned, Lanne and Luoto (2010) provide evidence in favor of forecasts of noncausal AR models compared to a number of Phillips curve forecasts. Apart from the illustration in Section 5, there is no practical experience on density forecasting with noncausal AR models, but the performance of density forecasts vis-à-vis relevant alternatives could also be tested. Corradi and Swanson (2006) provide a survey of tests of the accuracy of density forecasts that should be applicable also in this context.⁶

References

- Andrews, B. R.A. Davis, and F.J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* 97, 1638–1659.
- Breidt, F.J., R.A. Davis, K.S. Lii, and M. Rosenblatt (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis* 36, 175–198.
- Breidt, F.J., R.A. Davis, and A.A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *The Annals of Statistics* 29, 919–946.

⁶We would like to thank an anonymous referee for pointing this out.

Breidt, F.J. and N.-J. Hsu (2005). Best mean square prediction for moving averages. *Statistica Sinica*, 15, 427-446.

Clements, M.P., and J. Smith (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics* 14, 123-141.

Corradi, V., and N.R. Swanson (2006). Predictive density evaluation. In G. Elliott, C.W.J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*. Elsevier.

Diebold, F.X., T. Gunther, and A. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863-883.

Diebold, F.X., and R.S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

Fisher, J.D.M., C.T. Liu, and R. Zhou (2002). When can we forecast inflation? *Economic Perspectives* 1Q/2002, Federal Reserve Bank of Chicago, 30-42.

Huang, J. and Y. Pawitan (2000). Quasi-likelihood estimation of noninvertible moving average processes. *Scandinavian Journal of Statistics* 27, 689-710.

Lanne, M., and J. Luoto (2010). Has U.S. inflation really become harder to forecast? HECER Discussion Paper No. 307.

Lanne, M., H. Nyberg, and E. Saarinen (2011). Forecasting U.S. macroeconomic and financial time series with noncausal and causal AR models: A comparison. HECER Discussion Paper No. 319.

Lanne, M., and P. Saikkonen (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics* (forthcoming).

Lanne, M., and P. Saikkonen (2009). Noncausal vector autoregression. Bank of Finland Discussion Paper 18/2009.

Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for non-Gaussian nonminimum phase ARMA sequences. *Statistica Sinica* 6, 1-22.

Rosenblatt, M. (2000). Gaussian and non-Gaussian linear time series and random fields. Springer-Verlag, New York.

Stock, J.H., and M.W. Watson (2008). Phillips curve inflation Forecasts. In J. Fuhrer, Y. Kodrzycki, J. Little, and G. Olivei (eds.), *Understanding Inflation and the Implications for Monetary Policy*. Cambridge: MIT Press.

West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

Figure 1: The conditional cumulative distribution function of U.S. inflation in the first quarter of 1990 predicted by the AR(1,4) model.

Figure 2: One-step recursive out-of-sample density forecasts of U.S. inflation from the AR(1,4) model. The boxes represent the inter-quartile range of the forecasts and the inner line represents the median. The tails represent the 5th and 95th percentiles. The dots represent the observed inflation

Table 1: Estimation results of the AR(1,4) model for the demeaned U.S. inflation.

Parameter	Estimate	Standard error
ϕ_1	0.569	0.085
φ_1	0.002	0.108
φ_2	0.291	0.074
φ_3	0.226	0.078
φ_4	0.282	0.080
σ	1.171	0.159
λ	3.986	1.398
Log-likelihood	-201.477	

Table 2: Mean-square forecast errors of the AR(1,4) model estimated with data generated from the model in Table 1.

Horizon	T					
	100			200		
	N			N		
	1 000	10 000	100 000	1 000	10 000	100 000
$M = 25$						
1	1.335	1.316	1.318	1.250	1.238	1.256
2	1.696	1.679	1.682	1.587	1.576	1.579
4	2.472	2.459	2.462	2.355	2.338	2.338
8	3.607	3.599	3.601	3.348	3.318	3.319
$M = 50$						
1	1.325	1.323	1.312	1.252	1.253	1.252
2	1.684	1.690	1.679	1.591	1.581	1.580
4	2.477	2.460	2.464	2.356	2.337	2.347
8	3.653	3.630	3.631	3.415	3.347	3.356
$M = 100$						
1	1.320	1.314	1.315	1.250	1.254	1.256
2	1.690	1.681	1.682	1.606	1.580	1.586
4	2.479	2.463	2.464	2.372	2.345	2.343
8	3.643	3.638	3.632	3.368	3.356	3.367

The entries are based on 10 000 realizations. T is the sample size, and N is the number of replications in the forecasting procedure. The truncation parameter M is set at 50.

Table 3: Relative mean-square forecast errors of the AR(1,4) model compared to the AR(5,0) model estimated with data generated from the model in Table 1.

Horizon	T					
	100			200		
	N			N		
	1 000	10 000	100 000	1 000	10 000	100 000
$M = 25$						
1	0.952***	0.938***	0.940***	0.939**	0.929***	0.943*
2	0.921***	0.911***	0.913***	0.908***	0.902***	0.910***
4	0.903***	0.898***	0.899***	0.925***	0.919***	0.918***
8	0.891***	0.889***	0.889***	0.937***	0.929***	0.929***
$M = 50$						
1	0.944***	0.943***	0.935***	0.940***	0.941**	0.940***
2	0.914***	0.917***	0.911***	0.910***	0.905***	0.904***
4	0.904***	0.898***	0.900***	0.925***	0.918***	0.922***
8	0.902***	0.896***	0.897***	0.956	0.937***	0.939***
$M = 100$						
1	0.941***	0.937***	0.938***	0.939***	0.941**	0.943**
2	0.917***	0.913***	0.913***	0.919***	0.904***	0.908***
4	0.905***	0.899***	0.900***	0.932***	0.921***	0.920***
8	0.900***	0.898***	0.897***	0.943***	0.939***	0.942***

The entries are based on 10 000 realizations. T is the sample size, and N is the number of replications in the forecasting procedure. The truncation parameter M is set at 50. ***, ** and * indicate rejection in the Diebold-Mariano test at the 1%, 5% and 10% level, respectively.

Table 4: Mean-square forecast errors of different AR models for the U.S. inflation.

Model	Horizon			
	1	2	4	8
(5,0)- N	0.516	0.409	0.381	0.518
(5,0)	0.524	0.414	0.377	0.487
(4,1)	0.544	0.438*	0.419**	0.597***
(3,2)	0.562*	0.456**	0.443**	0.648***
(2,3)	0.515	0.342	0.319*	0.475***
(1,4)	0.486	0.347	0.281**	0.377***
(0,5)	0.533	0.417	0.393	0.525

The entries are the MSFEs of the $AR(r,s)$ models at different forecast horizons. The row entitled (5,0)- N corresponds to the causal Gaussian $AR(5)$ model. *, **, and *** denote significance at the 10%, 5% and 1% level of significance in the Diebold-Mariano test against the $AR(5,0)$ model, respectively.