Optimal Forecasting of Noncausal Autoregressive Time Series

Lanne, Markku and Luoto, Jani and Saikkonen, Pentti

University of Helsinki, HECER

February 2010

Helsinki Center of Economic Research

Discussion Papers

Optimal Forecasting of Noncausal Autoregressive Time Series

Markku Lanne
University of Helsinki and HECER

Jani Luoto
University of Helsinki and HECER

and

Pentti Saikkonen
University of Helsinki and HECER

Discussion Paper No. 286
February 2010

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014
University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 286

Optimal Forecasting of Noncausal Autoregressive Time Series*

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.

JEL Classification: C22, C53, C63

Keywords: Noncausal autoregression, density forecast, inflation

Markku Lanne

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: markku.lanne@helsinki.fi

Jani Luoto

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: jani.luoto@helsinki.fi

Pentti Saikkonen

Department of Mathematics and Statistics
University of Helsinki
P.O. Box 68 (Gustaf Hällströmin katu 2b)
FI-00014 University of Helsinki
FINLAND

e-mail: pentti.saikkonen@helsinki.fi

* Financial support from the Academy of Finland and OP-Pohjola Group Research Foundation is gratefully acknowledged.

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 (2008), 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

[1] Noncausal and potentially noninvertible autoregressive moving average models, as well as their 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 in noncausal.

The rest of the paper is organized as follows. The formulation of the noncausal AR model of Lanne and Saikkonen (2008) 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 (2008). 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 $(t = 0, \pm 1, \pm 2, ...)$ generated by

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

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

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

We use the abbreviation AR(r, s) for the model defined by (1) and sometimes write AR(r) for AR$(r, 0)$. If $\varphi_1 = \cdots = \varphi_s = 0$, model (1) reduces to the conventional causal 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 $\phi_1 = \cdots = \phi_r = 0$, we have the purely noncausal AR$(0, s)$ model with dependence on future values only. In the mixed 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

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

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

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

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 (2008). 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 $y_1, ..., y_T$ 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 $\boldsymbol{y} = (y_1, ..., y_T)$ and $\boldsymbol{z} = (u_1, ..., u_r, \epsilon_{r+1}, ..., \epsilon_{T-s}, v_{T-s+1}, ..., v_T)$. 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 $\boldsymbol{\phi} = (\phi_1, ..., \phi_r)$ and $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_s)$. As shown in Lanne and Saikkonen (2008), the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ are related by a linear transformation of the

form $\boldsymbol{z} = \boldsymbol{BAy}$ where the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ depend on the parameter vector $\boldsymbol{\theta}_1$. Moreover, the determinant of $\boldsymbol{B}$ is unity whereas the determinant of $\boldsymbol{A}$ 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 (2008) we can now conclude that the joint density function of the data vector $\boldsymbol{y} = (y_1, ..., y_T)$ can be expressed as

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

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

Analogosly to Breidt et al. (1991), Lanne and Saikkonen (2008) 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 θ_1 and θ_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.

In this section we have assumed that the model orders r and s are known. Procedures to specify these quantities are discussed in the next section.

2.3 Model Selection

Because noncausal AR processes are not identified by Gaussian likelihood, the first step in modeling a potentially noncausal time series is to search for signs of nonnormality. To this end, Lanne and Saikkonen (2008) suggest estimating a Gaussian AR(p) model that adequately captures the autocorrelation in the series and checking its residuals for nonnormality. For economic and financial time series, the residuals are often leptokurtic, indicating that Student's t-distribution might be a suitable error distribution.

Provided nonnormality is detected, the next step is to select among the alternative AR(r, s) specifications. As the AR(p) model has been found to adequately capture the autocorrelation in the series, it seems reasonable to restrict oneself to models with $r + s = p$. Following Breidt et al. (1991), Lanne and Saikkonen (2008) suggest selecting among these the model that produces the greatest value of the approximate likelihood function. Finally, the adequacy of the selected specification is checked diagnostically and the model is augmented if needed. In addition to examining the fit of the t-distribution, Lanne and Saikkonen (2008) checked the residuals for remaining autocorrelation and conditional heteroskedasticity. Moreover, they tested the significance of an additional lead and lag.

The purpose of fitting a Gaussian AR model in the first step is only to help determine the correct lag length and checking for nonnormality. Sometimes it may not be possible to come up with a satisfactory Gaussian AR model, in which case an adequate model might still be found among different non-Gaussian AR(r, s) specifications. In any case, one possibility is to start out with a non-Gaussian distribution, say the t-distribution, and select the model by employing information criteria.

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 $\boldsymbol{y} = (y_1, ..., y_T)$. From (4) it is seen that the optimal predictor of y_{t+h} $(h > 0)$ based on the observed data satisfies

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

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

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

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

that the joint density function of $(\mathbf{y}, \boldsymbol{\epsilon}^+)$ is

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

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

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

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

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

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

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

We write this briefly as

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

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

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

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

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

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

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

An approximation for the density function of $\boldsymbol{v}$ 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 $h_{\boldsymbol{v}}(\mathbf{v})$ and $h_{\boldsymbol{v},\boldsymbol{\epsilon}^+}(\boldsymbol{v}, \boldsymbol{\epsilon}^+)$ at the point $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{\phi})$ which depends on the observations (and the value of the parameter $\boldsymbol{\phi}$). Consider first the former. Using (10) we get the approximation

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

The last expression can be interpreted as the expectation of the first product therein with respect to the distribution of $\boldsymbol{\epsilon}^+$. Using simulation, this expectation can therefore be approximated as

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

where $\epsilon_{T+1}^{(i)}, ..., \epsilon_{T+M}^{(i)}$, $i = 1, ..., N$, are mutually independent simulated realizations from the distribution of $\boldsymbol{\epsilon}^+$. As $N \to \infty$, the right hand side of this approximation converges almost surely and provides an approximation for $h_{\boldsymbol{v}}\left(\boldsymbol{v}\left(\boldsymbol{\phi}\right)\right)$ 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

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

Specifically, we have

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

From the preceding discussion we can now conclude that the conditional expectation (8) can be approximated as

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

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. As in Breidt and Hsu (2005) an importance sampling scheme could be used to improve the approximation or make it accurate for a smaller value of N.

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 $\boldsymbol{\epsilon}^+$. Specifically, the forecast of $q(\boldsymbol{\epsilon}^+)$, say, is given by

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

For instance, letting $\mathbf{1}(\cdot)$ stand for the indicator variable and choosing $q(\boldsymbol{\epsilon}^+) = \mathbf{1}\left(\sum_{j=0}^{M-1} \beta_j \epsilon_{T+1+j} \leq x\right)$ 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 $x_1, ..., x_K$ 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 $\boldsymbol{y}_t = (y_t, ..., y_{t-r+1})$ and write (4) in the companion form

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

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

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

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

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

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

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

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 noncausality 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 nonlinear 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 1960:1 to 2008:2. In Section 5, we will provide evidence on the forecasting performance of various AR models for a number of subsamples of this series.

Using the model selection procedure outlined in Section 2 above, the AR(1,4) model with t-distributed errors is selected. 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, 3.25. The residuals are not autocorrelated but some remaining conditional 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). Throughout the truncation parameter M is set at 50. We also considered two other values, 25 and 100 (results not reported), but differences between these different choices of M turned out negligible.

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 both the sample size T and and the number of replications N, with the latter effect being more minor such that virtually no improvement is seen when N increases from 10 000 to 100 000. 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), all differences are also significant at least at the 5% level. In general, the superiority of the noncausal model tends to improve as N increases, but the differences are small, especially between 10 000 and 100 000 replications. The increase in the sample size favors the noncausal model when forecasting one period ahead, whereas at greater forecast horizons, its relative performance remains virtually unchanged and at the eight-period horizon even deteriorates. All in all, the choice of N does not seem to be critical as long as it is not too small, and the relative accuracy of the noncausal over causal model is not much affected by the sample size at least at short forecast horizons, whereas the forecast accuracy of the noncausal model seems to clearly improve with an increase in the sample size.

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. We consider forecasts for three different sample periods, from 1972:1 until 2008:2 and two subsample periods (1972:1–1990:1 and 1990:2–2008:2) of equal length. 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.

Forecasts 1, 2, 4 and 8 quarters ahead are computed as recursive pseudo out-of-sample forecasts based on reestimated models at each step with the estimation sample always starting from the first quarter of 1960. Several alternative forecast series are compared. First, forecasts from causal and purely noncausal fifth-order AR models as well as the AR(1,4) model selected for the entire sample in Section 4 are considered. Second, we produce forecast series based on reselecting the model at each step. Two alternative model selection procedures are employed, the Akaike information criterion (AIC) among the causal AR$(r, 0)$ or noncausal AR(r, s) models with the maximum values of r and s set at five and the procedure based on maximizing the log-likelihood function outlined in Section 2 above. In the latter case the sum of r and s is set at the AIC estimate of the Gaussian AR(p) model with the maximum value of p equal to five. In computing the forecasts, we set the number of replications N at 100 000 and the truncation parameter M at 50.

The MSFEs are reported in Table 4. Comparison of the recursive out-of-sample MSFEs suggests four findings. Firstly, the noncausal AR models generally outperform the causal models, especially at the longer forecast horizons. In the full sample, the purely noncausal AR model represents strong improvement in forecasting accuracy over the alternative models at the horizons from two to eight quarters, while the causal models perform relatively well in one-step ahead forecasting. Secondly, although the AR(r_{AIC}, s_{AIC}) model forecasts relatively poorly in the first subsample period, it does particularly well during the low inflation subsample period of 1990:1 through 2008:2,

producing the most accurate inflation forecast at all horizons. One reason for this may be the fact that, according to our experience, accurate estimation of the parameters of the mixed models requires a reasonably large number of observations. This is especially true when the orders of the autoregressive polynomials are high. In our sample, high-order AR(r, s) models are typically selected by the AIC, and since the sample starts from the first quarter of 1960, the sample used in the first reestimation consists of only 48 observations. Thirdly, among the models, the AR(r, s) model performs well in all out-of-sample periods and at all forecast horizons. However, there is no period or horizon in which it is able to produce the most accurate point forecasts on average. Finally, the out-of-sample forecasts indicate that with the exception of the latter subsample period, the Gaussian autoregressive model, AR(p_{AIC}), is not systematically outperformed by the AR(r_{AIC},0) model based on the t-distribution, suggesting that the distributional assumptions are not particularly critical in forecasting with causal AR models.

Often it is the direction of change rather than the magnitude that is of interest. Therefore, in addition to the MSFEs we also report the percentages of correct change-of-direction forecasts of the different models and model selection procedures in Table 5. At least one of the noncausal models always outperforms the causal models in all sample periods and at all forecast horizons, with a tie only in the first subsample period at the one-quarter horizon. Furthermore, all models predict the direction correctly over 50 percent of the time in all samples and at all horizons with only one exeption. Fisher, Liu and Zhou (2002) obtained similar results for U.S. consumer price inflation using multivariate (Phillips curve) models.

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% point. In almost all cases, the observed value is well within the interquartile range, with nearly all exceptions in the 1970's. For comparison, Figure 3 contains the corresponding plot for the causal AR(5,0) model, shose perfomance in density forecasting seems inferior. There are far more quarters when the observation is not even included in the 90% forecast interval, especially in the 1970's, and in other cases, the observation often lies on the boudary of the interquartile range or outside it. Hence, allowing for noncausality seems to improve the density forecasts of autoregressive models.

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 straightforward 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 reponses 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 inflation are encouraging, but more work is needed to evaluate the performance of the proposed method in different situations. In particular, as pointed out in Section 3, an importance sampling scheme could be used to improve accuracy or shorten the computation time. 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.

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.

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.

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 P. Saikkonen (2008). Modeling expectations with noncausal autoregressions. HECER Discussion Paper No. 212.

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. NBER Working Paper No. W14322.

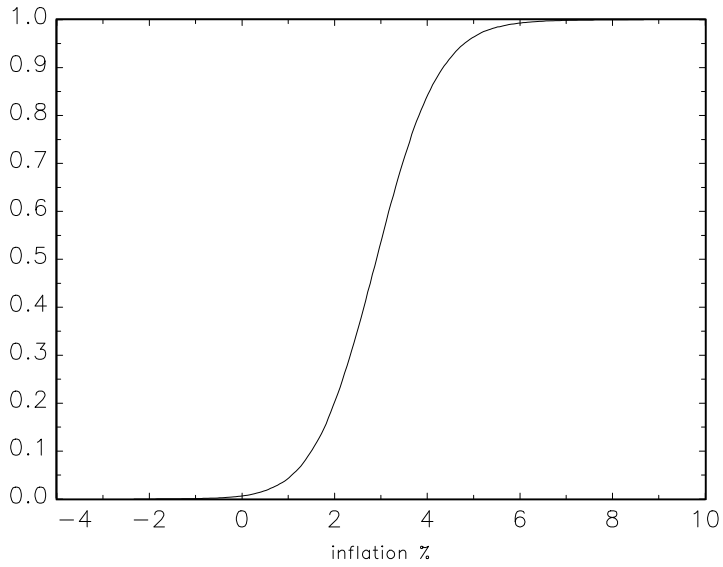

Figure 1: The conditional cumulative distribution function of U.S. inflation in the last quarter of 1989 predicted by the AR(1,4) model.

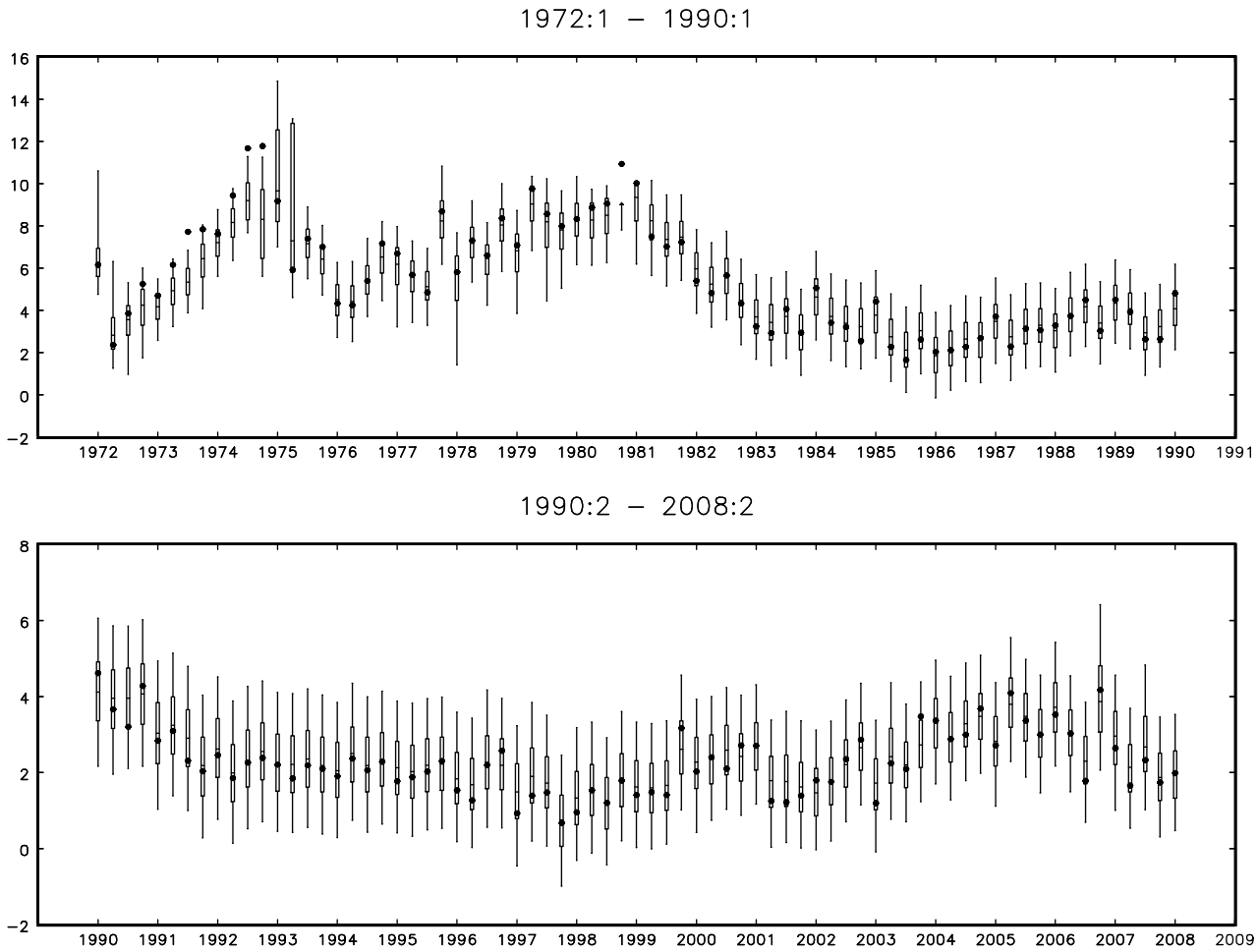

Figure 2: One-step recursive out-of-sample density forecasts of U.S. inflation from the AR(1,4) model. The boxes represent the inter-quartile range of the forecasts and the inner line representes the median. The tails represent the 5th and 95th percentiles. The dots represent the observed inflation.

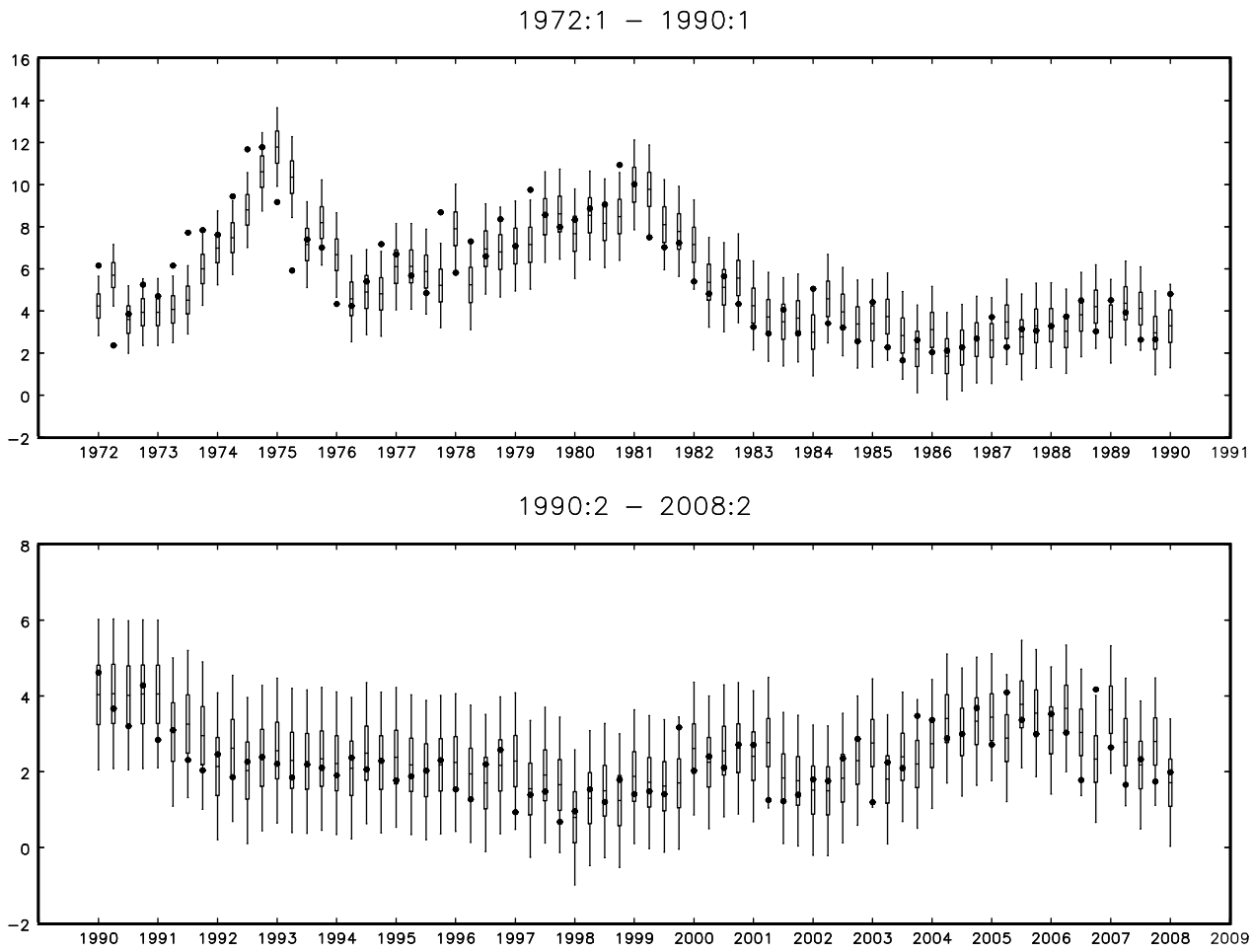

Figure 3: One-step recursive out-of-sample density forecasts of U.S. inflation from the AR(5,0) model. The boxes represent the inter-quartile range of the forecasts and the inner line representes the median. The tails represent the 5th and 95th percentiles. The dots represent the observed inflation.

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

Parameter	Estimate	Standard error
ϕ_1	0.672	0.065
φ_1	–0.166	0.080
φ_2	0.116	0.073
φ_3	0.304	0.054
φ_4	0.363	0.061
σ	1.164	0.207
λ	3.253	0.980
Log-likelihood	–261.181	

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

	T					
	100			200		
	N			N		
Horizon	1 000	10 000	100 000	1 000	10 000	100 000
1	1.358	1.325	1.321	1.284	1.261	1.253
2	1.553	1.522	1.522	1.494	1.449	1.454
4	2.367	2.343	2.355	2.245	2.222	2.211
8	4.143	4.107	4.104	3.889	3.858	3.871

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.

	T					
	100			200		
	N			N		
Horizon	1 000	10 000	100 000	1 000	10 000	100 000
1	0.934**	0.912**	0.909**	0.920**	0.904**	0.898**
2	0.870**	0.853**	0.853**	0.873*	0.847*	0.850**
4	0.857**	0.848**	0.853*	0.865**	0.856**	0.852**
8	0.887**	0.879*	0.879**	0.915**	0.902**	0.905*

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% and 5% level, respectively.

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

	Model						
Horizon	(5,0)	(0,5)	(1,4)	$(r_{AIC}, 0)$	(p_{AIC})	(r_{AIC}, s_{AIC})	(r, s)
	1972:1–2008:2						
1	1.369	1.406	1.707	1.370	1.382	1.648	1.470
2	1.343	1.245	1.727	1.275	1.276	1.594	1.286
4	1.604	1.322	1.827	1.520	1.516	1.634	1.401
8	2.585	1.896	2.289	2.575	2.580	2.561	2.023
	1972:1–1989:4						
1	2.264	2.333	2.953	2.265	2.282	2.831	2.478
2	2.318	2.117	3.119	2.180	2.178	2.853	2.231
4	2.836	2.269	3.324	2.666	2.651	2.949	2.467
8	4.629	3.254	4.074	4.609	4.596	4.636	3.536
	1990:1–2008:2						
1	0.475	0.479	0.461	0.475	0.483	0.464	0.461
2	0.380	0.383	0.358	0.380	0.384	0.356	0.358
4	0.381	0.379	0.351	0.381	0.389	0.333	0.350
8	0.561	0.548	0.522	0.561	0.583	0.508	0.521

The columns entitled $(p_{AIC}, 0)$, $(r_{AIC}, 0)$ and (r_{AIC}, s_{AIC}) contain the MSFEs for AR models selected by the Akaike information criterion. The first model assumes Gaussian errors. The column entitled (r, s) contains the AR models selected by maximizing the log-likelihood function among all fifth-order models.

Table 5: Percentages of correctely forecast change of direction of different AR models for the U.S. inflation.

	Model						
Horizon	(5,0)	(0,5)	(1,4)	$(r_{AIC}, 0)$	(p_{AIC})	(r_{AIC}, s_{AIC})	(r, s)
	1972:1–2008:2						
1	62.3%	63.0%	60.3%	60.3%	59.6%	56.8%	58.9%
2	62.8%	65.5%	64.8%	63.4%	62.8%	65.5%	66.9%
4	58.0%	61.5%	58.7%	59.4%	58.7%	62.9%	62.9%
8	56.1%	58.3%	56.8%	54.7%	55.4%	57.6%	59.0%
	1972:1–1989:4						
1	61.6%	61.6%	57.5%	57.5%	56.2%	56.1%	54.8%
2	58.3%	59.7%	54.2%	59.7%	58.3%	58.3%	58.3%
4	59.2%	63.4%	54.9%	62.0%	60.6%	62.0%	63.4%
8	59.4%	62.3%	58.0%	56.5%	58.0%	65.2%	62.3%
	1990:1–2008:2						
1	63.0%	64.4%	63.0%	63.0%	63.0%	57.5%	63.0%
2	68.1%	72.2%	75.0%	68.1%	68.1%	72.2%	75.0%
4	57.7%	60.6%	62.0%	57.7%	57.7%	64.8%	62.0%
8	53.6%	55.1%	55.1%	53.6%	53.6%	49.3%	55.1%

See notes to Table 4.